When More Candidates Hurt: A Proposal–Evidence Reliability Frontier for Causal Selection

Anonymous authors
Paper under double-blind review

Abstract

Candidate generation can scale faster than causal evaluation. Under a fixed evidence pool, enlarging a candidate prefix changes both the best attainable action and the statistical burden of comparing candidates with uneven causal support. We characterize this joint movement through a *proposal–evidence reliability frontier*. Proposal-Certified Selection (PCS) uses proposal-side predictions to order candidates and lower endpoints of simultaneous causal intervals to select among them. Its all-prefix certificate balances comparator miss against certification width across all covered prefixes. In a Gaussian benchmark, a lower bound yields a $\sqrt{\log K/n}$ comparison requirement, showing that logarithmic multiplicity persists beyond the plug-in selector. A proposal/support-aware allocation uniquely minimizes the stated expected certificate-width objective; realized decision performance is evaluated separately. Controlled diagnostics reproduce the predicted selection-error and multiplicity patterns. In Upworthy replays, the paired PCS gain over tuned fixed depth rises from 0.042 to 0.949 basis points as evidence increases from 0.5% to 5%. In a 26.9M-impression recommendation-log study, the randomized-audit click-through-rate (CTR) gain over inverse-propensity-score (IPS) plug-in is 26.70 basis points (95% cluster interval $[10.99, 42.42]$). The ten-issue persuasion study remains unresolved, and uniform allocation nearly matches adaptive allocation in recommendation. Together, these results separate proposal coverage from statistically usable decision improvement.

1 Introduction

A newsroom can propose headlines faster than it can A/B test them; a language model can generate interventions faster than a randomized study can establish their effects; and a recommender can rank a catalog despite sparse randomized exposure (Matias et al., 2021; 2024; Hackenburg et al., 2025; Saito et al., 2021). The decision is not simply which candidate looks best, but how much proposal breadth the available causal evidence can reliably distinguish.

The two resources play different roles. A predictor, generator, or historical ranker can prioritize many actions, whereas randomized or adjusted outcomes supply evidence about their causal values. When evidence is limited or exposure is weak, those causal estimates remain noisy. Proposal quality thus governs which actions enter consideration, while causal support governs how reliably they can be compared.

This creates a basic tension. Adding an action cannot worsen the best attainable value in a nested proposal sequence, yet it can worsen the value actually selected: more comparisons create opportunities for optimistic errors, and later candidates may be more weakly supported. Thus the same evidence pool must support both a moving comparator and a growing comparison family. Figure 1 visualizes the resulting gap between *attainable coverage* and *certifiable decision quality*.

In this work, we study this tradeoff through *Proposal-Certified Selection (PCS)*. Proposal-side information orders the candidates, while one simultaneous family of causal intervals evaluates them. Across growing prefixes, PCS selects by lower confidence endpoints. This induces the *proposal–evidence reliability frontier*: a prefix-indexed lower envelope that trades comparator miss against candidate-specific certification width. The frontier separates three questions that are otherwise easy to

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

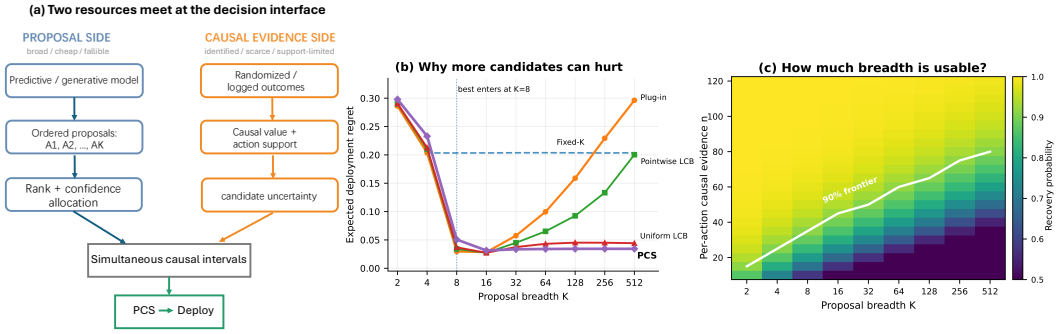

Figure 1: **Proposal breadth is limited by available causal evidence.** Cheap proposal generation creates many candidate actions, but causal evidence for distinguishing them is scarce. Increasing breadth amplifies multiplicity for plug-in selection, while simultaneous lower-confidence selection yields a proposal–evidence frontier for reliable recovery.

conflate: simultaneous certification across growing prefixes, the intrinsic statistical cost of comparing a larger family, and how pre-outcome proposal/support information can allocate confidence across candidates.

Our contributions are threefold. **(i)** We characterize causal selection over growing proposal prefixes through an all-prefix decision certificate and a Gaussian lower bound with $\sqrt{\log K/n}$ comparison-rate dependence. **(ii)** We derive the unique proposal/support-aware allocation that minimizes the stated stochastic certificate-width objective and separate this structural optimum from the directly deployable rank-weighted rule. **(iii)** Controlled diagnostics and retrospective real-data studies stress-test the predicted roles of multiplicity, proposal quality, causal support, and information budget, including regimes in which empirical differences remain unresolved. The causal statements are conditional on identification and interval validity; the real-data analyses are replays or randomized audits.

2 Related Work

We organize the closest literature around three lines: **(1)** proposal models and Best-of-N selection, **(2)** causal evaluation and pessimistic policy selection, and **(3)** multiplicity control and best-arm identification. PCS lies at their interface: proposal-side information determines which actions enter the ordered family, while causal evidence determines how reliably that family can be compared.

(1) Proposal models and Best-of-N. Predictive and generative systems make broad search inexpensive, while reward-overoptimization and Best-of-N analyses show that proxy-selected quality need not improve monotonically with search breadth (Gao et al., 2023; Coste et al., 2024; Jinnai et al., 2025; Huang et al., 2025; Yu et al., 2026; Sriraman & Block, 2026). We retain the distinction between proposal coverage and selection quality, but the protected quantity is an action's causal value rather than the proposal score. Candidate-specific experimental or logged support therefore determines the precision with which additional proposals can be compared. **(2) Causal evaluation and pessimistic policy selection.** Doubly robust estimation and policy learning provide causal value estimates and decision rules (Dudík et al., 2011; Swaminathan & Joachims, 2015; Chernozhukov et al., 2018; Athey & Wager, 2021), while high-confidence and pessimistic methods protect selection over uncertain policy families (Thomas et al., 2015; Yang et al., 2022; Wang et al., 2024). Large-action and support-aware methods further address uneven evidence (Saito & Joachims, 2022; Saito et al., 2023; Chapagain, 2026; Shekhar & Howard, 2026). These methods motivate the interval engine and fixed-family comparators used here. PCS focuses on the additional joint movement created when the attainable comparator, action-specific support, and covered comparison family all change with the proposal prefix; uniform simultaneous LCB remains a central comparator. **(3) Multiplicity, allocation, and active identification.** Weighted simultaneous coverage provides the error-spending primitive (Weinstein & Ramdas, 2020; Tian & Ramdas, 2021; Dobriban et al., 2015), whereas best-arm identification allocates future samples to resolve an unknown winner (Garivier & Kaufmann, 2016; Yang et al., 2025; de Heide, 2026). PCS keeps the evidence pool fixed and optionally redistributes *confidence*

across candidates rather than observations. The resulting optimization concerns confidence allocation under a fixed evidence pool. PCS contributes the growing-prefix causal decision object in which the attainable comparator, candidate-specific support, and covered comparison family move jointly with proposal breadth; Appendix C gives a detailed comparison with adjacent formulations.

3 Causal Decisions over Growing Proposal Prefixes

Notation. For a context x, $\mathcal{A}(x)$ denotes the admissible action library, $A_{1:K_{\max}}$ the ordered proposal stream, and $S_k = \{A_1, \dots, A_k\}$ its first-k prefix. We write $\nu(a, x)$ for causal value and $V^\star(x)$ for the global benchmark. For candidate A_j, $\hat{\nu}_j$, $[L_j, U_j]$, W_j, and π_j denote its value estimate, simultaneous interval, interval width, and confidence-budget weight. Finally, $\mathcal{G}$ collects proposal/design information fixed before held-out outcome noise. Appendix A gives the complete notation and assumptions.

More proposals can improve the best attainable action but also enlarge the set over which noisy causal estimates compete. Decision error therefore separates into proposal miss and selection error; Figure 2 illustrates this decomposition. Throughout, proposal breadth varies against a fixed causal-evidence pool: PCS reallocates confidence but does not collect additional outcomes, distinguishing the setting from adaptive best-arm identification (Garivier & Kaufmann, 2016; Yang et al., 2025).

3.1 Decision Target and Ordered Candidates

Let $X \in \mathcal{X}$ be a deployment context and $\mathcal{A}(X)$ a finite or countable library of distinct admissible actions. A proposer returns

$$A_{1:K_{\max}} \sim Q_\theta(\cdot \mid X), \qquad S_k(X) = \{A_1, \dots, A_k\}. \tag{1}$$

For a supported action a, its potential outcome is $Y(a)$ and its causal value is

$$\nu(a, x) = \mathbb{E}[Y(a) \mid X = x] \tag{2}$$

Fix context x and write $\nu_j := \nu(A_j, x)$; the prefix guarantee is not uniform over contexts. Let $V^\star(x) := \sup_{a \in \mathcal{A}(x)} \nu(a, x)$ and $j_k^\star \in \arg\max_{j \le k} \nu(A_j, x)$. The supremum need not be attained. The proposal gap is

$$\Delta_{\text{prop}}(k, x) = V^\star(x) - \nu(A_{j_k^\star}, x), \tag{3}$$

It measures the loss from not proposing a better action, even if selection inside the prefix were perfect. A plug-in selector chooses $\hat{J}_k \in \arg\max_{j \le k} \hat{\nu}_j$ and incurs

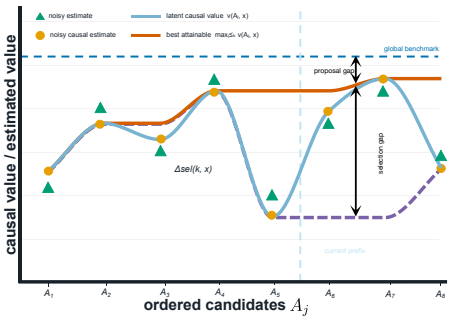

Figure 2: **Proposal and selection errors under a growing candidate prefix.** The orange envelope is the best causal value attainable within each prefix. At the displayed prefix, the gap to the global benchmark is the proposal error, while noisy causal evaluation can induce an additional selection error. The figure is schematic.

$$R_k(x) = V^\star(x) - \nu(A_{\hat{J}_k}, x) = \Delta_{\text{prop}}(k, x) + \underbrace{\nu(A_{j_k^\star}, x) - \nu(A_{\hat{J}_k}, x)}_{\Delta_{\text{sel}}(k, x)}. \tag{4}$$

Only the first term must decrease with breadth. The second depends on which noisy comparison wins.

3.2 Proposal Information and Causal Evidence

The order may use generation randomness, retrieval, pretrained scores, or disjoint historical outcomes. Let $\mathcal{G}$ contain proposal randomness, fitted objects, the realized sequence, and pre-outcome design metadata, excluding held-out interval fluctuations. Proposal scores need not be calibrated or causally interpretable.

The causal evaluator plays a different role from the proposer. It does not determine which candidates enter the prefix; it assigns causal values and uncertainty to candidates that have already been proposed. We use the augmented inverse-propensity score below as one concrete evaluation engine (Dudík

et al., 2011; Chernozhukov et al., 2018), while the PCS selection theory itself only requires valid candidate-wise intervals.

Causal observations are $O_i = (X_i, A_i, Y_i)$, with assignment law $\mu(\cdot \mid X_i)$. Write $m_a(x) = \mathbb{E}[Y \mid A = a, X = x]$ and $\mu_a(x) = \mathbb{P}(A = a \mid X = x)$. With true nuisance functions, the augmented inverse-propensity score is

$$\psi_a(O) = m_a(X) + \frac{\mathbf{1}\{A = a\}}{\mu_a(X)}\{Y - m_a(X)\} \tag{5}$$

Under identification, $\mathbb{E}[\psi_a \mid X = x] = \nu(a, x)$, but its variance remains

$$\mathrm{Var}(\psi_a \mid X = x) = \frac{\mathrm{Var}(Y \mid A = a, X = x)}{\mu_a(x)}. \tag{6}$$

Thus an identifiable action can still be difficult to compare. With out-of-fold nuisance estimates, an observation-level score and its aggregated causal value estimate are, respectively,

$$\hat{\psi}_{a,i} = \hat{m}_a(X_i) + \frac{\mathbf{1}\{A_i = a\}}{\hat{\mu}_a(X_i)}\{Y_i - \hat{m}_a(X_i)\}, \qquad \hat{\nu}_j(x) = \sum_i w_{ij}(x)\hat{\psi}_{A_j,i}, \tag{7}$$

Here $w_{ij}(x) \geq 0$ sum to one. $\hat{\psi}_{a,i}$ is an observation-level score; $\hat{\nu}_j$ averages/localizes such scores. The standard doubly robust property addresses nuisance misspecification under its usual conditions (Dudík et al., 2011; Chernozhukov et al., 2018); sampling uncertainty in the aggregated value source is handled separately in Appendix H.

3.3 Information Conditions and Selection Risk

Consistency and no interference connect observed and potential outcomes; conditional exchangeability $Y(a) \perp A \mid X$ and overlap $\mu(a \mid x) > 0$ identify evaluated actions under the usual causal and off-policy conditions (Dudík et al., 2011; Athey & Wager, 2021). Randomization supplies exchangeability by design, whereas observational use requires credible adjustment or sensitivity intervals. Sample separation is a sufficient route to freezing $\mathcal{G}$; cross-fitting supplies out-of-fold nuisance predictions (Chernozhukov et al., 2018), while interval validity is supplied by the interval engine. Finite conditional variances justify the variance comparison; the finite-sample radius later requires a stronger conditional tail bound. Unsupported actions remain outside the core theorem (Appendices A–B).

These conditions govern identification and interval construction rather than the growing-prefix selection argument itself. Once valid conditional candidate-wise intervals are available, the PCS result below is agnostic to the particular causal estimator used to construct them. Even without nuisance error or proposal miss, noisy selection can deteriorate. Let Φ denote the standard normal cumulative distribution function in the following construction.

Theorem 1 (Plug-in regret can increase with every proposal)**.** *Fix $\Delta, \sigma > 0$. Candidate 1 has value 0 and is estimated without error. Every $j \geq 2$ has value $-\Delta$ and estimate $-\Delta + \varepsilon_j$, where $\varepsilon_j \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$. For the plug-in selector over the first k candidates,*

$$\mathbb{E}[R_k] = \Delta\{1 - \Phi(\Delta/\sigma)^{k-1}\}, \tag{8}$$

which increases strictly with k and converges to Δ.

The best action is already present: additional candidates add chances for noise to defeat it, not better attainable values. The target, proposal information, and causal source are now specified. What remains is a rule whose decision certificate survives selection across prefixes.

4 From Causal Intervals to the Reliability Frontier

An individually accurate estimate need not protect the winner of many comparisons. Uncertainty must remain valid after the proposal/design stage and across the inspected family. This requirement produces the frontier; its residual comparison cost then motivates allocation and an executable selector. We first derive the all-prefix certificate in Theorem 2 and Corollary 3, then characterize the comparison cost and confidence-allocation problem in Theorems 4 and 7, before giving the outcome-separated deployment rule in Section 4.3.

4.1 Simultaneous Certification and Prefix Regret

Fix $\delta \in (0,1)$. We use a weighted error-spending construction motivated by weighted multiple inference and online familywise control (Dobriban et al., 2015; Tian & Ramdas, 2021). Candidate A_j receives an interval $[L_j, U_j]$ and pre-outcome confidence weight π_j. Here π_j is not an action-selection or sampling probability; it is candidate j's share of the family-level confidence budget δ. Allocating a larger share permits a smaller stochastic radius in the interval construction below, while the total normalized budget remains at most one. The weights satisfy

$$\mathbb{P}(\nu(A_j, x) \notin [L_j, U_j] \mid \mathcal{G}) \le \delta\pi_j, \qquad \pi_j = \pi_j(\mathcal{G}) \ge 0, \quad \sum_j \pi_j \le 1 \text{ a.s.} \tag{9}$$

The weights may depend on $\mathcal{G}$, but not on the held-out fluctuations they cover. Candidate estimates may be dependent. One simultaneous coverage event follows by summing the conditional failure probabilities.

For example, let $n_{\mathrm{eff},j} = 1/\sum_i w_{ij}^2$. If the centered weighted estimate is conditionally $\sigma_j^2/n_{\mathrm{eff},j}$-sub-Gaussian and $B_j(x) \ge 0$ bounds its conditional bias, a valid radius for $\pi_j > 0$ is

$$r_j(x) = B_j(x) + \sigma_j(x)\sqrt{\frac{2\log\{2/(\delta\pi_j)\}}{n_{\mathrm{eff},j}(x)}}, \qquad [L_j, U_j] = [\hat{\nu}_j(x) - r_j(x), \hat{\nu}_j(x) + r_j(x)]. \tag{10}$$

The bias envelope B_j allows persistent estimation error to enter the interval explicitly. Validity is a property of the interval engine: cross-fitting or a reported standard error alone does not establish the conditional coverage used below. Other interval constructions can be substituted when they provide the required coverage interpretation.

Motivated by high-confidence and pessimistic policy selection (Thomas et al., 2015; Yang et al., 2022; Wang et al., 2024), PCS chooses $\hat{J}_k \in \arg\max_{j \le k} L_j$ and deploys $A_{\hat{J}_k}$. Define $\Delta_j(x) = V^\star(x) - \nu(A_j, x)$ and $W_j(x) = U_j - L_j$. The first measures how much value is lost by using candidate j as the comparator; the second measures how much causal uncertainty is required to certify it. Their decision-relevant combination is

$$\mathfrak{F}_k(x) := \inf_{j \le k}\{\Delta_j(x) + W_j(x)\}. \tag{11}$$

Figure 3 shows why proposal rank alone does not determine the useful comparator: smaller proposal miss can be offset by wider causal uncertainty. The next theorem shows that this candidate-wise tradeoff indeed controls PCS regret simultaneously across all covered prefixes.

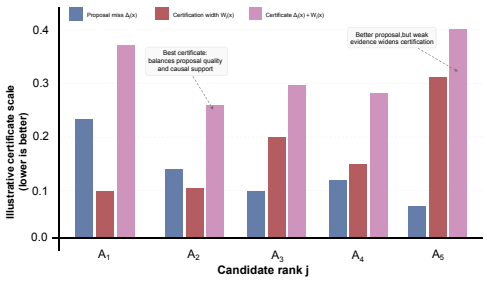

Figure 3: **Proposal quality and causal reliability jointly determine the certificate.** For candidate A_j, the relevant scale is $\Delta_j(x) + W_j(x)$. Better proposals need not yield tighter certificates when causal evidence is weak; the frontier takes the minimum over candidates in the prefix. Values are schematic; lower is better.

Theorem 2 (Proposal–evidence reliability frontier)**.** *Fix $\delta \in (0,1)$. Let $\mathcal{J}$ be a finite or countable realized proposal stream with weights satisfying Equation (9), and define $\hat{J}_k \in \arg\max_{j \le k} L_j$. With probability at least $1-\delta$, simultaneously for every prefix k and comparator $j \le k$,*

$$V^\star(x) - \nu(A_{\hat{J}_k}, x) \le \underbrace{V^\star(x) - \nu(A_j, x)}_{\Delta_j(x)} + \underbrace{U_j - L_j}_{W_j(x)} \tag{12}$$

Consequently, on the same event,

$$R_{\mathrm{PCS},k}(x) \le \mathfrak{F}_k(x) = \inf_{j \le k}\{\Delta_j(x) + W_j(x)\} \le \inf_{\ell \le k}\{\Delta_{\mathrm{prop}}(\ell, x) + W_{j^\star_\ell}(x)\}. \tag{13}$$

The statement also holds at any finite data-dependent stopping prefix covered by the original interval family.

On common coverage, $\nu_{\hat{J}_k} \ge L_{\hat{J}_k} \ge L_j$ and $\nu_j \le U_j$. Thus PCS is compared with *every* candidate, not just the largest estimate. An early, well-supported comparator can remain useful after uncertain candidates arrive (proof in Appendix F).

For fixed covered intervals, $\mathfrak{F}_k$ is nonincreasing in k because its infimum expands with the prefix. Realized regret is a distinct quantity and can move differently. Because the frontier contains unknown comparator gaps, it should be interpreted as a simultaneous decision certificate rather than an estimated risk curve or stopping objective.

Corollary 3 (Operational frontier under support)**.** *For $\delta \in (0,1)$, if Equation (10) is valid, then with probability at least $1-\delta$, simultaneously for every k,*

$$R_{\mathrm{PCS},k}(x) \le \inf_{j \le k} \left[\underbrace{\Delta_j(x)}_{\textit{proposal miss}} + \underbrace{2B_j(x) + 2\sigma_j(x)\sqrt{\frac{2\log\{2/(\delta\pi_j)\}}{n_{\mathrm{eff},j}(x)}}}_{\textit{causal reliability cost}} \right]. \tag{14}$$

The certificate separates three reliability costs: bias, support-limited precision, and multiplicity. Holding the other ingredients fixed, more effective evidence contracts the stochastic-width component. We next ask whether the logarithmic comparison term is specific to plug-in selection.

4.2 Comparison Limits and Width-Optimal Allocation

The plug-in example only diagnoses one selector. To determine whether multiplicity is intrinsic, we use a Gaussian comparison experiment and a Fano-type lower-bound argument, where uncertainty concerns the location of the best action. This experiment is an analytic comparison benchmark, not a deployment assumption: the PCS certificate itself requires valid candidate-wise intervals, not Gaussian noise, equal variances, or equal per-action support. Its μ_j denotes an arm mean, not the logging propensity $\mu(a \mid x)$.

Theorem 4 (A Gaussian lower bound for the multiplicity cost)**.** *Consider $K \ge 2$ actions. Under hypothesis $v \in [K]$, action v has mean Δ and all others mean 0; all Kn observations are independent, with n observations from each action distributed as $\mathcal{N}(\mu_j, \sigma^2)$. If, for $\delta \in (0,1)$, a selector identifies the best action with error probability at most δ uniformly over v, then*

$$\frac{n\Delta^2}{\sigma^2} \ge \big[(1-\delta)\log K - \log 2\big]_+. \tag{15}$$

In this Gaussian benchmark, reliable selection over an unknown best action requires a gap of order $\sigma\sqrt{\log K/n}$.

Here n counts observations *per action*, totaling Kn, rather than a fixed total budget. Corollaries 5–6 (Appendix P) place an expected-regret lower bound and a high-probability upper bound at the same $\sqrt{\log K/n}$ scale, while keeping their distinct risk criteria explicit. Figure 4 examines the corresponding diagnostics.

Within this Gaussian benchmark, the logarithmic comparison cost persists across selectors, but the family-level confidence budget can still be spent nonuniformly. Let q_j describe a proposal-side prior on the useful comparator rank, and let a_j be a design-side stochastic scale. Intuitively, q_j records how plausible it is, before held-out outcomes are seen, that rank j supplies a useful comparator; a_j summarizes how difficult that candidate is to estimate from the available design support, with larger a_j indicating a wider stochastic scale. We therefore ask a narrower optimization question: which allocation minimizes the expected stochastic width attached to that comparator?

Theorem 7 (Reliability-adaptive width-optimal allocation)**.** *Fix $\delta \in (0,1)$ and a finite family. Conditional on $\mathcal{G}$, let $q_j > 0$ satisfy $\sum_{j=1}^K q_j = 1$ and let $a_j > 0$ be fixed proposal/design-side quantities that do not depend on the held-out outcome fluctuation used by the final intervals. With $c = 2/\delta$, consider*

$$\mathcal{C}(\pi) = \sum_{j=1}^{K} q_j a_j \sqrt{2\log\frac{c}{\pi_j}}, \qquad \pi_j > 0, \quad \sum_j \pi_j = 1. \tag{16}$$

$\mathcal{C}$ is strictly convex and has a unique minimizer $\pi^\star$. It is characterized by

$$\frac{q_j a_j}{\pi_j^\star \sqrt{2\log(c/\pi_j^\star)}} = \lambda^\star, \qquad j = 1, \dots, K, \tag{17}$$

for a common $\lambda^\star > 0$. Uniform allocation is optimal if and only if $q_j a_j$ is constant across j. If J is a comparator rank drawn from the proposal-side prior q and candidate j has stochastic half-width exactly $a_j\sqrt{2\log(c/\pi_j)}$, then $2\mathcal{C}(\pi)$ is the expected stochastic full-width term; hence $\pi^\star$ minimizes that component of the decision certificate.

Theorem 7 is a structural characterization of the width-optimal allocation, not the default deployment rule. Because $q_j a_j > 0$ and $c = 2/\delta > 2$, $\mathcal{C}$ is strictly convex and diverges at the simplex boundary, so its minimizer is unique and interior; Appendix G gives the proof and bisection construction. Using this adaptive allocation requires q_j and a_j to be fixed from eligible proposal/design information before the held-out interval noise. Otherwise, Section 4.3 uses the prespecified rank-power allocation. Accordingly, width optimality under a supplied comparator prior need not imply downstream decision-value improvement when that prior is misspecified.

4.3 Outcome-Separated Selection and Deployment

For deployment, we use a prespecified rank-power allocation fixed before held-out outcome noise is observed,

$$\pi_j^{K,p} = \frac{j^{-p}}{H_{K,p}}, \qquad H_{K,p} = \sum_{r=1}^{K} r^{-p}, \quad p > 0, \tag{18}$$

where K is the fixed evaluator horizon, not a new normalization at each stopping prefix. An unbounded proposal stream can instead use

$$\pi_j^{\infty} = \frac{1}{j(j+1)}, \qquad \sum_{j\geq 1} \pi_j^{\infty} = 1, \tag{19}$$

A finite data-dependent stopping prefix is covered within this originally protected family and its fixed error budget. Outcome-adaptive proposal regeneration or repeated reallocation changes the information structure and requires a separate analysis.

Algorithm 1 Proposal-Certified Selection (PCS)

Input: Context x, proposal data $\mathcal{D}_{\text{prop}}$, causal-evaluation data, confidence level δ, horizon K.

1: Fit and freeze the proposer; obtain supported ordered candidates $A_1, \ldots, A_K$ and proposal-side information $\mathcal{G}$.
2: Fix simultaneous confidence weights $\pi_1, \ldots, \pi_K$ using only pre-outcome information.
3: Estimate candidate values $\hat{\nu}_j(x)$ using out-of-fold nuisances or known randomization, and construct $[L_j, U_j]$ at level $\delta\pi_j$.
4: For prefix S_k, select

$$\hat{J}_k \in \arg\max_{j\leq k} L_j.$$

5: Freeze and return the selected action $\hat{A}_k = A_{\hat{J}_k}$ before test/audit outcomes are observed.

Algorithm 1 implements the certificate. Validation choices may enter $\mathcal{G}$ when they are separated from the interval noise covered by the final confidence family. The displayed argmax guarantee covers candidate selection; fallback or abstention can be added with separate coverage. Unsupported actions require an explicit transport model, such as Equation (23), or abstention because representation similarity alone does not identify causal value (Appendices I–J).

5 Experiments: Which Breadth Is Statistically Usable?

5.1 Experimental Setup

Each experiment targets a specific implication of the theory. The Gaussian diagnostics probe the selection and $\sqrt{\log K/n}$ comparison mechanisms in Theorems 1 and 4; evidence-budget and support-thinning experiments stress the support-aware certificate; and the OBD audit examines the additional confidence-allocation refinement. In the real-data studies, we ask whether broader comparison becomes more useful as causal evidence strengthens and whether simultaneous selection remains stable as candidate-specific support deteriorates. Weak-signal or small-cluster regimes may therefore remain unresolved. In OBD, adaptive PCS uses the allocation solver of Theorem 7 with empirical production-side scales;

Table 1: **Key experimental settings.**

Quantity	Setting
Plug-in diagnostic	$\Delta = \sigma = 1$
Recovery diagnostic	$n = 50$ per action
Recovery K	$\{2, 8, 32, 128, 512\}$
Upworthy evidence (%)	$\{0.5, 1, 2, 5\}$
Upworthy PCS	$K_{\max} = 6$; $(\delta, p) = (.1, 2)$
Persuasion evidence (%)	$\{25, 50, 100\}$
Persuasion PCS	$(\delta, p) = (.2, 2)$
OBD audit	54 blocks; 18 clusters
Support skew	$\{1{:}1, 2{:}1, 4{:}1, 8{:}1\}$

because the IPS intervals are selection scores rather than verified finite-sample instances of Equation (9), the audit probes this allocation refinement empirically rather than instantiating the full PCS guarantee.

• **Controlled diagnostics:** Two Gaussian experiments isolate selection and multiplicity under known truth. The first instantiates Theorem 1 as the proposal prefix grows; the second varies candidate-family size at fixed per-action evidence to probe the $\sqrt{\log K/n}$ comparison scale in Theorem 4. We additionally stress candidate-specific support through outcome-independent late-rank thinning while leaving randomized audit outcomes fixed.

• **Retrospective randomized reanalyses:** Across Upworthy, persuasion, and OBD, proposal ordering and selector tuning are fixed before the held-out or randomized outcomes used for evaluation are opened. Upworthy uses a historical headline-only ranker (Matias et al., 2021; 2024); persuasion separates held-out participants from selection evidence and uses other issues for ordering and tuning (Hackenburg et al., 2025); OBD uses historical BTS logs for ordering, current BTS logs for selection, and time-matched Uniform Random outcomes for audit (Saito et al., 2021). The OBD IPS intervals are selection scores rather than a finite-sample verification of Equation (9).

• **Controls and ablations:** Matched controls include fixed-depth and plug-in selectors, pointwise and simultaneous LCBs, and validation-tuned uncertainty penalties; OBD additionally includes rank-1 and adaptive PCS. All selectors share the same order, evidence partition, and holdout within each replay. Further proposal-order, support, interval-scale, and allocation stresses are reported in Appendices K–L, with artifact scope in Appendix O.

5.2 Main Results and Cross-Setting Robustness

With no proposal miss, the exact plug-in risk at $\Delta = \sigma = 1$ grows from 0.159 at $K = 2$ to essentially 1 by $K = 64$. The Gaussian recovery diagnostic at $n = 50$ requires gaps of about 0.26 and 0.63 at $K = 2$ and 512 (Figure 4). These are selection/multiplicity checks, not real-world effect sizes.

As causal uncertainty decreases, broader comparison becomes more usable. Upworthy's paired PCS–fixed-depth gain grows from 0.042 to 0.949 bp with evidence (Table 2). At the point-estimate level, rank+variance is lower-regret at 0.5% and 2%, while PCS is lower-regret at 1% and 5% (Figure 5). The transition is therefore reliability-dependent rather than uniformly method-dependent.

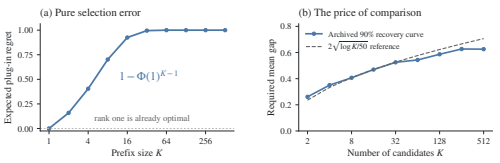

Figure 4: **Broader comparison can amplify error even without proposal miss.** (a) Exact evaluation of Theorem 1 at $\Delta = \sigma = 1$. (b) Reported 90%-recovery gaps across candidate-family sizes, shown with the $2\sqrt{\log K/50}$ reference. The diagnostic exhibits the logarithmic growth predicted by the Gaussian comparison benchmark; lines connect reported points only.

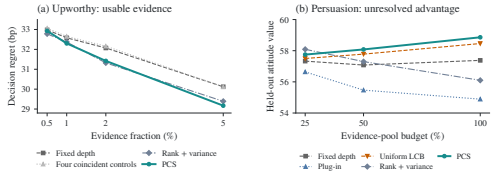

Figure 5: **More evidence changes the value of comparison without implying a uniform selector advantage.** Curves show the reported means at each evidence budget. In Upworthy, plug-in, pointwise LCB, uniform LCB, and the CV variance penalty coincide at the shown precision and share one gray curve. Persuasion shows five central controls; all seven appear in Appendix L.4. Lines connect reported budgets only; paired uncertainty is reported in Table 2.

Paired uncertainty separates statistically resolved gains from unresolved real-data trends (Table 2).

Persuasion marks a statistical boundary. At the point-estimate level, PCS is highest at 50% and 100% evidence (58.08 and 58.86 versus uniform LCB's 57.78 and 58.46), while rank+variance is highest at 25%. All three issue-bootstrap intervals versus fixed depth include zero; at full evidence, the gain is 1.48 with interval $[-0.56, 3.68]$. Ten issue-level contexts therefore leave the comparative advantage unresolved.

OBD separates certification from allocation. PCS gains 26.70 bp over IPS plug-in; uniform LCB and adaptive PCS gain 28.41 and 28.42 bp. Their approximately 0.008 bp difference does not establish an allocation advantage. Six-day clustering widens the PCS interval to $[1.78, 51.63]$ (Appendix N.5).

5.3 Ablation and Sensitivity Analyses

Candidate-specific causal support provides the most direct stress test of the reliability mechanism.

Outcome-independent thinning of late-rank selection evidence creates increasing support imbalance while leaving the randomized audit half unchanged (Table 3). Plug-in value drops sharply under this stress, whereas simultaneous lower-confidence selection remains substantially more stable. PCS has the highest point estimate in all four controlled regimes. This is a support intervention on an RCT-derived shell, not evidence about naturally occurring confounding. Proposal quality and certification are distinct. Randomizing only the proposal order raises Upworthy PCS regret by 2.95, 2.25, 1.50, and 1.21 bp; learned order raises persuasion value by 1.04, 1.13, and 1.40. Uniform LCB keeps simultaneous coverage while removing nonuniform spending. In the proposal-signal stress, adaptive allocation is lower when proposal signal is absent and higher by about 0.012–0.017 at full signal. This separates the width objective in Theorem 7 from downstream decision value when the comparator prior is misspecified.

Table 2: **Real-data paired uncertainty audit.** Positive values favor the listed selector; intervals use the study-specific paired resampling procedures.

Study / evidence	Selector	Paired gain	95% interval
Upworthy / 0.5%	PCS	0.042 bp	$[-0.050,\ 0.137]$
Upworthy / 1%	PCS	0.283 bp	$[0.127,\ 0.442]$
Upworthy / 2%	PCS	0.651 bp	$[0.424,\ 0.889]$
Upworthy / 5%	PCS	0.949 bp	$[0.671,\ 1.230]$
Persuasion / 25%	PCS	0.43	$[-1.67,\ 2.44]$
Persuasion / 50%	PCS	1.00	$[-1.29,\ 3.33]$
Persuasion / 100%	PCS	1.48	$[-0.56,\ 3.68]$
OBD / rolling	PCS	26.70 bp	$[10.99,\ 42.42]$
OBD / rolling	Uniform LCB	28.41 bp	$[13.19,\ 43.64]$
OBD / rolling	Adaptive PCS	28.42 bp	$[12.52,\ 44.32]$

Sensitivity sweeps over (δ, p) preserve the evidence-budget dependence: persuasion gains are 1.22–1.57 at full evidence but range from -0.28 to $+0.54$ at the smallest budget. OBD interval-scale perturbations remain positive but serve as calibration sensitivities rather than validity checks. Model-family, frozen-label, and measured-confounding stresses in Appendices M–N probe supported-action perturbations; domain-shift and backbone generalization are separate questions.

5.4 Efficiency and Error Analysis

PCS scans K lower endpoints in $O(K)$ time. The selector-only CPU benchmark is 1.7–2.2 μs for $K = 32$–512; the adaptive solver takes 47–107 ms. Appendix M.8 reports memory and timing details, separate from proposal fitting and causal estimation. Appendix M.9 gives qualitative OBD cases showing both sides of pessimism: conservative selection can reject an optimistic noisy estimate, but can also miss a useful uncertain candidate. These cases are illustrations rather than frequency estimates.

Table 3: **Controlled causal-support stress.** Held-out value under outcome-independent late-rank thinning; the audit half is fixed.

	Early:late support ratio			
Method	1:1	2:1	4:1	8:1
Tuned fixed-K	57.43	56.93	57.09	57.69
Plug-in	55.29	53.14	54.32	54.18
Uniform LCB	58.75	57.73	57.81	57.59
PCS	**59.07**	**58.09**	**57.98**	**57.83**

6 Discussion, Limitations, and Conclusion

The main lesson is not that more candidates are harmful, but that candidate generation and causal resolution scale differently. A proposal system can make an increasingly good action attainable long before the available evidence can reliably distinguish it from its alternatives. The relevant notion of scale is therefore not the number of candidates that can be generated, but the amount of proposal breadth that the available causal evidence makes statistically usable. Proposal capacity expands the opportunity set; evidential capacity determines how much of that expansion can be converted into a reliable decision. The proposal–evidence reliability frontier is the boundary between these two regimes. Beyond it, additional generation increases what is available without necessarily increasing what can be safely selected. A scalable decision system should therefore match the growth of its proposal space with the growth of the evidence required to resolve it. The guarantees require valid candidate-wise causal intervals for supported actions and proposal/design information fixed before the held-out outcome noise; unsupported actions or outcome-adaptive proposal regeneration require additional structure beyond the present analysis. The real-data studies are retrospective reanalyses or randomized audits, and the width optimality in Theorem 7 does not by itself imply downstream decision-value improvement in every deployment regime.

AI USE STATEMENT

Generative AI assisted the manuscript with structural editing, interpretation checks, scientific-figure code, a separately labeled reference implementation, and consistency audits. It was not used to manufacture empirical observations or replace the authors' theorem statements. The authors remain responsible for the text, citations, code, disclosure, and empirical provenance.

ETHICS STATEMENT

The studies reuse existing public experimental and logging data. The persuasion setting concerns political attitudes; retrospective analysis is not an endorsement of manipulative deployment. Any operational use requires appropriate consent, privacy safeguards, and domain-specific review. The statistical certificate does not establish fairness, non-harm, or safety outside its stated value and identification assumptions.

REPRODUCIBILITY STATEMENT

Algorithm 1 and Appendices A–J document the information contract and proofs; Appendices K–N give the experimental protocols and reported results. The current artifact package supports formula checks, panel-level reproduction from reported quantities, and separately labeled reference-code tests. End-to-end raw-data reproduction additionally requires the experimental programs, aggregate result files, provenance metadata, and third-party raw archives listed in Appendix O.

REFERENCES

Susan Athey and Stefan Wager. Policy learning with observational data. *Econometrica*, 89(1): 133–161, 2021.

Nilson Chapagain. CASP: Support-aware offline policy selection for two-stage recommender systems. *arXiv preprint arXiv:2604.23022*, 2026.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.

Thomas Coste, Usman Anwar, Robert Kirk, and David Krueger. Reward model ensembles help mitigate overoptimization. In *International Conference on Learning Representations*, 2024.

Rianne de Heide. Where does the union bound go? best-arm identification and strong FWER control. *arXiv preprint arXiv:2608.19903*, 2026.

Edgar Dobriban, Kristen Fortney, Stuart K. Kim, and Art B. Owen. Optimal multiple testing under a gaussian prior on the effect sizes. *Biometrika*, 102(4):753–766, 2015.

Miroslav Dudík, John Langford, and Lihong Li. Doubly robust policy evaluation and learning. In *Proceedings of the 28th International Conference on Machine Learning*, pp. 1097–1104, 2011.

Germano Gabbianelli, Gergely Neu, and Matteo Papini. Importance-weighted offline learning done right. In *Proceedings of the 35th International Conference on Algorithmic Learning Theory*, volume 237, pp. 614–634, 2024.

Leo Gao, John Schulman, and Jacob Hilton. Scaling laws for reward model overoptimization. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pp. 10835–10866, 2023.

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

Kobi Hackenburg, Ben M. Tappin, Paul Röttger, Scott A. Hale, Jonathan Bright, and Helen Margetts. Scaling language model size yields diminishing returns for single-message political persuasion. *Proceedings of the National Academy of Sciences*, 122(10):e2413443122, 2025.

Konstantin Hess, Dennis Frauen, Valentyn Melnychuk, and Stefan Feuerriegel. Efficient and sharp off-policy learning under unobserved confounding. In *International Conference on Learning Representations*, 2026.

Audrey Huang, Adam Block, Qinghua Liu, Nan Jiang, Akshay Krishnamurthy, and Dylan J Foster. Is best-of-N the best of them? Coverage, scaling, and optimality in inference-time alignment. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267, pp. 25075–25126, 2025.

Yuu Jinnai, Tetsuro Morimura, Kaito Ariu, and Kenshi Abe. Regularized best-of-N sampling with minimum Bayes risk objective for language model alignment. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 9321–9347, 2025.

Zhewei Kang, Xuandong Zhao, and Dawn Song. Scalable best-of-N selection for large language models via self-certainty. In *Advances in Neural Information Processing Systems*, volume 38, pp. 19720–19745, 2025.

J. Nathan Matias, Kevin Munger, Marianne Aubin Le Quere, and Charles Ebersole. The upworthy research archive, a time series of 32,487 experiments in U.S. media. *Scientific Data*, 8(1):195, 2021.

J. Nathan Matias, Kevin Munger, Marianne Aubin Le Quere, and Charles Ebersole. Author correction: The upworthy research archive, a time series of 32,487 experiments in U.S. media. *Scientific Data*, 11(1):765, 2024.

Alizée Pace, Hugo Yèche, Bernhard Schölkopf, Gunnar Rätsch, and Guy Tennenholtz. Delphic offline reinforcement learning under nonidentifiable hidden confounding. In *The Twelfth International Conference on Learning Representations*, 2024.

Yuta Saito and Thorsten Joachims. Off-policy evaluation for large action spaces via embeddings. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162, pp. 19089–19122, 2022.

Yuta Saito, Shunsuke Aihara, Megumi Matsutani, and Yusuke Narita. Open bandit dataset and pipeline: Towards realistic and reproducible off-policy evaluation. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, volume 1, 2021.

Yuta Saito, Qingyang Ren, and Thorsten Joachims. Off-policy evaluation for large action spaces via conjunct effect modeling. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pp. 29734–29759, 2023.

Prashant Shekhar and Caroline Howard. Support-aware offline policy selection for advertising marketplaces. *arXiv preprint arXiv:2605.21736*, 2026.

Ved Sriraman and Adam Block. Revisiting the (sub)optimality of best-of-N for inference-time alignment. In *Proceedings of Thirty Ninth Conference on Learning Theory*, volume 336, pp. 5980–6028, 2026.

Adith Swaminathan and Thorsten Joachims. Counterfactual risk minimization: Learning from logged bandit feedback. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37, pp. 814–823, 2015.

Philip Thomas, Georgios Theocharous, and Mohammad Ghavamzadeh. High confidence policy improvement. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37, pp. 2380–2388, 2015.

Jinjin Tian and Aaditya Ramdas. Online control of the familywise error rate. *Statistical Methods in Medical Research*, 30(4):976–993, 2021.

Lequn Wang, Akshay Krishnamurthy, and Alex Slivkins. Oracle-efficient pessimism: Offline policy optimization in contextual bandits. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238, pp. 766–774, 2024.

Asaf Weinstein and Aaditya Ramdas. Online control of the false coverage rate and false sign rate. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pp. 10193–10202, 2020.

Le Yang, Vincent Y. F. Tan, and Wang Chi Cheung. Best arm identification with possibly biased offline data. In *Proceedings of the Forty-first Conference on Uncertainty in Artificial Intelligence*, volume 286, pp. 4715–4730, 2025.

Mengjiao Yang, Bo Dai, Ofir Nachum, George Tucker, and Dale Schuurmans. Offline policy selection under uncertainty. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, volume 151, pp. 4376–4396, 2022.

Zhuohao Yu, Steven Wu, and Adam Block. From curiosity to caution: Mitigating reward hacking for best-of-N with pessimism. In *International Conference on Learning Representations*, 2026.

Bohan Zhang, Yixin Wang, and Paramveer S. Dhillon. Policy learning with a natural language action space: A causal approach. *arXiv preprint arXiv:2502.17538*, 2025.

Appendix

CONTENTS

A Notation, Information Structure, and Assumptions **15**
A.1 Notation at a glance . . . 15
A.2 Information structure . . . 15
A.3 Identification conditions . . . 16
A.4 Decision objects . . . 16

B Assumption Audit and Relation to Prior Work **16**

C Expanded Related Work and Novelty Boundary **17**

D Proof of the Plug-in Failure Result **18**

E Multiplicity Lower-Bound Proofs **19**

F Proof of the Prefix-Uniform Proposal Oracle Bound **20**

G Matching-Rate and Allocation Proofs **21**

H Causal Estimation and Interval Construction **23**
H.1 Identification by the augmented inverse-propensity score . . . 23
H.2 Orthogonality remainder . . . 23
H.3 Localized value and effective sample size . . . 23
H.4 Variance and overlap . . . 24
H.5 Alternative interval engines . . . 24

I Extension to Unsupported or Newly Generated Text **24**

J Full Algorithmic Specification **25**

K Upworthy Data and Experimental Protocol **25**
K.1 Dataset construction . . . 25
K.2 Historical proposal ranker . . . 26
K.3 Selectors and intervals . . . 26
K.4 Metrics and uncertainty . . . 26
K.5 Reproducibility controls . . . 27

L Complete Baseline Audit and Experimental Results **27**
L.1 Persuasion replay construction . . . 27

L.2 Baseline definitions and validation budgets 27
L.3 Full Upworthy strong-baseline table 27
L.4 Full persuasion strong-baseline table 27

M Full Sensitivity, Robustness, Efficiency, and Error Analyses **28**
M.1 Controlled logging-support stress on an RCT-derived shell 28
M.2 Small-cluster inference audit for persuasion 28
M.3 Why several Upworthy uncertainty baselines coincide 29
M.4 Hyperparameter sensitivity 29
M.5 Proposal-order ablation 29
M.6 Action-source-family robustness 29
M.7 Controlled evidence-noise stress 30
M.8 Selector-only efficiency 30
M.9 Qualitative cases and failure taxonomy 30

N Theory-Facing and External-Evidence Audits **31**
N.1 Multiplicity-scaling diagnostic 31
N.2 Proposal-prior/support differential stress 31
N.3 Measured-confounding reweighting audit 31
N.4 Released GPT-label proposer 31
N.5 Natural production logging with randomized audit 32

O Full Experimental Reproducibility Map **32**

P Supporting Equal-Support Statements **33**

Q Visual Provenance and Artifact Boundaries **34**

APPENDIX OVERVIEW AND PROOF MAP

The appendix collects notation, assumptions, proofs, extended literature comparisons, implementation details, and additional empirical diagnostics. Appendix A summarizes the notation and information structure; Appendix B states the identification and support conditions; and Appendix C positions the decision object relative to adjacent problem formulations. The proof chain is direct: Theorem 1 is proved in Appendix D; Theorem 4 and Corollary 5 in Appendix E; Theorem 2 and Corollary 3 in Appendix F; and Corollary 6 and Theorem 7 in Appendix G. The remaining appendices provide the interval engine, unsupported-action extension, complete algorithm, dataset protocols, baseline tuning, robustness analyses, external audits, and reproducibility information.

A NOTATION, INFORMATION STRUCTURE, AND ASSUMPTIONS

This section collects the conditioning structure, support assumptions, and decision objects used throughout Secs. 3–4, so that the notation and information access are explicit before the proofs.

A.1 NOTATION AT A GLANCE

Symbol	Meaning
X	deployment context; $\mathcal{A}(X)$ is the admissible action library.
A_j, S_k	jth proposed action and the first-k candidate prefix $\{A_1, \ldots, A_k\}$.
Q_θ, $\mathcal{G}$	proposal/ranking mechanism and the proposal/design sigma-field fixed before held-out outcome noise.
$Y(a)$, $\nu(a, x)$	potential outcome under action a and its conditional causal value.
$\mu(a \mid x)$	logging/assignment probability; it controls both overlap and precision.
$[L_j, U_j]$, W_j	valid interval for candidate j and its full width $W_j = U_j - L_j$.
π_j	simultaneous confidence-budget weight for candidate/rank j.
q_j, a_j	proposal-side comparator prior and pre-outcome support/scale used in adaptive allocation.
$\Delta_{\text{prop}}(k, x)$	gap between the global benchmark and the best true value in prefix k.
$R_{\text{PCS},k}(x)$	deployment regret of PCS at prefix k.

Table 4: Core notation. Prediction/proposal quantities are frozen before the held-out outcome fluctuations used by the final interval engine.

A.2 INFORMATION STRUCTURE

The proposal/prediction model is allowed to be pretrained or fitted on a disjoint historical training sample and may depend on prompts, decoding randomness, retrieval, reward-model scores, or an internal world model. It is therefore a broad source of cheap predictive guidance rather than a causal estimator. We collect all of that information in

$$\mathcal{G} = \sigma\Big(\mathcal{D}_{\text{prop}}, Q_\theta, X, A_1, A_2, \ldots, \Pi, \mathcal{M}_{\text{design}}, \widehat{\mathcal{M}}_{\text{nuis}}\Big),$$

where $\Pi = (\pi_1, \pi_2, \ldots)$, $\mathcal{M}_{\text{design}}$ denotes design-only support metadata, and $\widehat{\mathcal{M}}_{\text{nuis}}$ denotes nuisance fits frozen before the held-out outcome fluctuations used by the final interval engine. The interval statements are conditional on $\mathcal{G}$. A sufficient implementation is a three-way split: one fold fits the proposer, one fits nuisance functions, and one constructs action-value intervals. Cross-fitting can recycle observations while preserving out-of-fold predictions. The conditional formulation matters because an unconditional interval for a pre-specified action need not remain valid after the same data have been searched to create and rank the candidates.

The strict-support theory assumes a finite or countable admissible library $\mathcal{A}(x)$. This includes a common deployment pattern in which a language model retrieves, edits, or ranks interventions from an approved catalog. It also covers generated strings that are deduplicated and mapped to pre-registered executable actions before evaluation. The extension to a genuinely new string is discussed in Appendix I.

A.3 IDENTIFICATION CONDITIONS

For each $a \in \mathcal{A}(x)$, let $Y(a)$ be well defined. The causal value in Equation (2) is identified under the following conditions.

Assumption 1 (Consistency and no interference)**.** *If $A = a$, then $Y = Y(a)$. One unit's action does not change another unit's outcome in the one-step analysis.*

Assumption 2 (Conditional exchangeability)**.** *For every supported a, $Y(a) \perp A \mid X$.*

Assumption 3 (Evaluation support for proposed actions)**.** *For every candidate returned with positive proposal probability, $\mu(a \mid x) > 0$ on the context region used by the evaluator. For finite-sample guarantees we additionally require a lower bound sufficient to control the interval width.*

Randomized assignment satisfies exchangeability by design. With observational feedback, X must include all measured common causes used to justify the assumption. If hidden confounding remains plausible, one can replace point intervals by sensitivity intervals, as in recent pessimistic and sharp off-policy analyses (Pace et al., 2024; Hess et al., 2026). The selection theorem continues to hold for any valid interval, but the interval may be wide enough to force abstention.

A.4 DECISION OBJECTS

For a realized proposal sequence and context x, assume $V^\star(x) := \sup_{a \in \mathcal{A}(x)} \nu(a, x) < \infty$ and use:

$$\begin{aligned} \nu_j &:= \nu(A_j, x), & j_k^\star &\in \arg\max_{j \le k} \nu_j, \\ \Delta_j(x) &:= V^\star(x) - \nu_j, & \Delta_{\mathrm{prop}}(k, x) &:= \Delta_{j_k^\star}(x), \\ W_j(x) &:= U_j - L_j, & R_{\mathrm{PCS},k}(x) &:= V^\star(x) - \nu(A_{\hat{j}_k}, x). \end{aligned}$$

The proposal-selection coverage statements condition on the realized proposal/design sigma-field $\mathcal{G}$ and then integrate over it for unconditional probability statements. The Gaussian plug-in and Fano constructions are stated directly under their own experiment laws; they do not require this conditioning device.

These definitions are exactly the objects consumed by the method and theory in the main text; subsequent appendices add proofs and diagnostics rather than a different estimand.

B ASSUMPTION AUDIT AND RELATION TO PRIOR WORK

This section expands the identification contract in Section 3 and records the role of each assumption, what can be checked or stress-tested, and which additional structure is needed when an assumption is not credible. The scope remains the strict-support setting of the main theorem.

The strict-support core uses consistency, conditional exchangeability, and overlap. These are not paper-specific assumptions: they are standard identification conditions underlying doubly robust evaluation and logged-policy learning (Dudík et al., 2011; Swaminathan & Joachims, 2015; Chernozhukov et al., 2018). Likewise, summable confidence spending is an inherited multiple-testing/online-coverage primitive (Weinstein & Ramdas, 2020; Tian & Ramdas, 2021). Offline policy selection already studies uncertainty-aware ranking of a fixed policy set (Yang et al., 2022), and high-confidence policy improvement certifies batch policy proposals (Thomas et al., 2015). The present object differs because an ordered prefix simultaneously changes the attainable comparator and the covered comparison family. This is also distinct from BAI, which allocates samples to identify an arm; here evidence collection is fixed before the proposal-indexed decision (Garivier & Kaufmann, 2016; Yang et al., 2025).

The assumptions play distinct roles. Consistency links the observed response to the potential outcome under the assigned action. Exchangeability identifies action values from the observed assignment mechanism; randomized trials make this condition design-based, whereas observational deployments require measured-confounder adjustment or sensitivity analysis. Overlap is both an identification and precision condition: as $\mu(a \mid x)$ approaches zero, inverse-propensity variation grows and the certificate widens. Independence between proposal construction and interval evaluation is a sufficient route to the conditional coverage statement in Equation (9). Cross-fitting can provide out-of-fold

Problem formulation	Proposal-side ordering	Causal decision target	Prefix-dependent comparator	Candidate-specific support	All-prefix simultaneity	Confidence vs. sample allocation
Best-of-N / proxy selection (Gao et al., 2023)	✓	△	✓	–	–	–
Pessimistic fixed-family selection (Yang et al., 2022; Wang et al., 2024)	–	✓	–	△	–	–
Large-action OPE (Saito & Joachims, 2022; Saito et al., 2023)	–	✓	–	✓	–	–
Weighted simultaneous inference (Dobriban et al., 2015; Tian & Ramdas, 2021)	–	△	–	△	△	confidence
Best-arm identification (Garivier & Kaufmann, 2016; Yang et al., 2025)	–	△	–	✓	–	sample
PCS (this work)	✓	✓	✓	✓	✓	confidence

Table 5: **Object-level comparison with adjacent formulations.** ✓ denotes a central element of the representative formulation, △ denotes a variant-dependent or partial element, and – denotes that the dimension is not central to that formulation. The dimensions describe problem structure rather than methodological superiority; individual variants can differ from the representative citations.

nuisance predictions when nuisance models use the same overall sample, but does not by itself establish conditional simultaneous coverage.

The main theorem remains on strict support. Equation (23) defines a separate transport extension, where the Lipschitz constant is domain substantive and generally not identified from text embeddings alone. Calibration, sensitivity analysis, partial identification, or abstention is therefore required before unsupported free text enters the selector.

The audit therefore supports the scope stated in the main text: the all-prefix decision guarantee is only as credible as the action-value intervals it consumes, and unsupported or hidden-confounded actions require additional identifying structure.

C Expanded Related Work and Novelty Boundary

This section positions PCS relative to adjacent problem formulations using the same information structure as the main text. The distinguishing object is a *growing ordered candidate family* under a fixed causal-evidence pool: proposal-side information determines the available prefix, while causal evidence determines the precision and simultaneous comparison cost within that prefix.

The table makes the boundary explicit without treating inherited ingredients as new. PCS combines proposal ordering with a causal decision target, candidate-specific support, a prefix-dependent comparator, and one family-level confidence budget; the theoretical contribution concerns the consequences of that joint structure.

Prediction/proposal versus causal evaluation. Reward models, critics, search procedures, and Best-of-N methods use model-side scores to choose among generated outputs (Gao et al., 2023; Coste et al., 2024; Jinnai et al., 2025; Kang et al., 2025; Huang et al., 2025; Yu et al., 2026; Sriraman & Block, 2026). They motivate our use of proposal-side information, but their score is generally the object being optimized. In our information model, the predictive score is only allowed to order candidates or set pre-outcome confidence weights; the target remains a causal value learned from randomized or adjusted logged outcomes. This distinction is what prevents a strong predictive score from being silently treated as a causal estimate.

Fixed policy families and pessimistic selection. High-confidence policy improvement and offline policy selection already show how uncertainty can protect decisions over a fixed set (Thomas et al., 2015; Yang et al., 2022). Pessimistic and importance-weighted offline learning similarly encode uncertainty in policy objectives (Wang et al., 2024; Gabbianelli et al., 2024). The growing-prefix setting changes the comparator as well as the covered family: the best available action can improve with k at the same time that multiplicity and action-specific uncertainty worsen. The oracle bound in Theorem 2 is designed for that joint movement.

Large action spaces and support. Large-action OPE uses structure such as embeddings or conjunct effects to share statistical strength across sparse actions (Saito & Joachims, 2022; Saito et al., 2023); natural-language policy learning treats text as the action representation (Zhang et al., 2025). Those directions can supply a stronger interval engine or proposer. PCS instead treats the interval engine as modular and asks how valid action-wise uncertainty should be converted into a post-selection

decision across prefixes. The strict-support theorem intentionally does not identify genuinely unseen text without additional transport assumptions.

Multiple comparisons and active best-arm identification. Weighted Bonferroni and online familywise-error control justify nonuniform confidence spending (Dobriban et al., 2015; Weinstein & Ramdas, 2020; Tian & Ramdas, 2021). Best-arm identification allocates additional samples based on observed evidence (Garivier & Kaufmann, 2016; Yang et al., 2025; de Heide, 2026). Our allocation problem is different: q_j and a_j are fixed from proposal/design information before held-out outcome fluctuations, and the evidence pool is not adaptively enlarged. Theorem 7 is therefore an optimality statement only for the stated confidence-width objective, not a global optimality claim over all experimental designs or selectors.

Novelty boundary. Across these adjacent literatures, the distinctive contribution is the growing-prefix causal decision object and the theory induced by its joint movement: the proposal-gap/selection-cost decomposition, the all-prefix certificate, the Gaussian comparison-rate lower bound, and the associated confidence-allocation problem. The causal identification and simultaneous-coverage ingredients remain modular inputs to this decision layer.

D Proof of the Plug-in Failure Result

This section verifies Theorem 1, the counterexample used in Section 3 to show that nested proposal sets need not improve deployment. The construction isolates selection error; it does not rely on causal identification assumptions.

Exact restatement of Theorem 1.

> Fix $\Delta, \sigma > 0$. Candidate 1 has value 0 and is estimated without error. Every $j \geq 2$ has value $-\Delta$ and estimate $-\Delta + \varepsilon_j$, where $\varepsilon_j \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$. For the plug-in selector over the first k candidates,
> $$\mathbb{E}[R_k] = \Delta\{1 - \Phi(\Delta/\sigma)^{k-1}\},$$
> which increases strictly with k and converges to Δ.

Proof of Theorem 1. Candidate one has estimated value zero. A suboptimal candidate $j \geq 2$ defeats it exactly when

$$-\Delta + \varepsilon_j > 0 \quad \Longleftrightarrow \quad \varepsilon_j > \Delta.$$

Because the Gaussian errors are independent and ties have probability zero, the plug-in selector chooses the optimal candidate if and only if every suboptimal estimate is at most zero. Therefore

$$\mathbb{P}(\hat{J}_k = 1) = \prod_{j=2}^{k} \mathbb{P}(\varepsilon_j \leq \Delta) = \Phi(\Delta/\sigma)^{k-1}.$$

The regret is zero when candidate one is selected and Δ otherwise, yielding

$$\mathbb{E}[R_k] = \Delta\mathbb{P}(\hat{J}_k \neq 1) = \Delta\{1 - \Phi(\Delta/\sigma)^{k-1}\}.$$

Since $0 < \Phi(\Delta/\sigma) < 1$, this expression is strictly increasing in k and converges to Δ. □

Heteroskedastic extension. If candidate j has independent Gaussian noise with standard deviation σ_j, then

$$\mathbb{E}[R_k] = \Delta \left\{ 1 - \prod_{j=2}^{k} \Phi(\Delta/\sigma_j) \right\}.$$

The regret remains nondecreasing, and every added candidate with finite $\sigma_j > 0$ increases it strictly. Low logging propensity raises the variance in Equation (6), supplying a causal route to large σ_j.

Unequal gaps. For values $\nu_j = -\Delta_j$ and exact estimation of candidate one, the probability of selecting a suboptimal action is

$$1 - \prod_{j=2}^{k} \Phi(\Delta_j/\sigma_j),$$

although expected regret additionally depends on which candidate attains the maximum. The simple equal-gap construction is sufficient to establish that nested proposals do not imply monotone decision value.

Thus Theorem 1 isolates a pure selection mechanism: proposal growth can worsen deployment even when the true best action is already available at rank one.

E Multiplicity Lower-Bound Proofs

This section proves Theorem 4 and Corollary 5. Its role is to show that the logarithmic comparison cost is not an artifact of PCS, under the Gaussian model stated in the main text.

Exact restatement of Theorem 4.

> Consider $K \geq 2$ actions. Under hypothesis $v \in [K]$, action v has mean Δ and all others mean 0; all Kn observations are independent, with n observations from each action distributed as $\mathcal{N}(\mu_j, \sigma^2)$. If, for $\delta \in (0,1)$, a selector identifies the best action with error probability at most δ uniformly over v, then
>
> $$\frac{n\Delta^2}{\sigma^2} \geq \big[(1-\delta)\log K - \log 2\big]_+.$$
>
> In this Gaussian benchmark, reliable selection over an unknown best action requires a gap of order $\sigma\sqrt{\log K/n}$.

Exact restatement of Corollary 5.

> For $K \geq 16$, consider the same Gaussian observation model and the K instances with one best arm of gap
>
> $$\Delta_K = \frac{\sigma}{2}\sqrt{\frac{\log K}{n}}.$$
>
> Then every selector obeys
>
> $$\inf_{\widehat{J}} \sup_{v\in[K]} \mathbb{E}_v[\mu_v - \mu_{\widehat{J}}] \geq \frac{\sigma}{4}\sqrt{\frac{\log K}{n}}.$$

Proof map. The multiplicity lower bound compares K Gaussian experiments that differ only in the location of the best arm and applies Fano's inequality. The minimax simple-regret corollary then chooses a hard gap at the Fano scale.

Proof of Theorem 4. Let V be uniform on $[K]$. Under P_v, arm v has mean Δ and every other arm has mean zero, with n independent samples per arm and variance σ^2. For $v \neq w$,

$$D_{\mathrm{KL}}(P_v \| P_w) = \frac{n\Delta^2}{2\sigma^2} + \frac{n\Delta^2}{2\sigma^2} = \frac{n\Delta^2}{\sigma^2}.$$

The mutual information between V and the complete observation is therefore at most $n\Delta^2/\sigma^2$ by the standard pairwise-KL bound. Fano's inequality gives

$$\Pr(\widehat{V} \neq V) \geq 1 - \frac{n\Delta^2/\sigma^2 + \log 2}{\log K}.$$

Uniform error at most δ implies average error at most δ; rearranging proves Equation (15). The positive part covers small K or weak target confidence for which Fano's numerical lower bound is vacuous. □

Proof of Corollary 5. Set $\Delta = \Delta_K = (\sigma/2)\sqrt{\log K/n}$. The pairwise KL divergence in the preceding proof becomes $(1/4)\log K$. With V uniform on $[K]$, Fano therefore gives

$$\Pr(\widehat{J} \neq V) \geq 1 - \frac{\frac{1}{4}\log K + \log 2}{\log K}.$$

For $K \geq 16$, $\log 2/\log K \leq 1/4$, so the average error probability is at least $1/2$. Every error incurs simple regret exactly Δ_K in this hard family. Hence

$$\frac{1}{K}\sum_{v=1}^{K}\mathbb{E}_v[\mu_v - \mu_{\widehat{J}}] \geq \frac{\Delta_K}{2} = \frac{\sigma}{4}\sqrt{\frac{\log K}{n}}.$$

The supremum over v is at least this average, and the statement holds for every selector; taking the infimum over selectors proves Equation (25). □

Together, these bounds justify the main-text claim that a logarithmic multiplicity price is unavoidable in the stated Gaussian comparison problem; they do not assert constant-matched minimax optimality for every risk criterion.

F Proof of the Prefix-Uniform Proposal Oracle Bound

This section proves Theorem 2 and Corollary 3. It verifies that one simultaneous interval event is sufficient to control every covered prefix and comparator; no additional independence assumption is introduced inside the proof.

Exact restatement of Theorem 2.

> Fix $\delta \in (0,1)$. Let $\mathcal{J}$ be a finite or countable realized proposal stream with weights satisfying Equation (9), and define $\hat{J}_k \in \arg\max_{j\leq k} L_j$. With probability at least $1-\delta$, simultaneously for every prefix k and comparator $j \leq k$,
>
> $$V^\star(x) - \nu(A_{\hat{J}_k}, x) \leq \underbrace{V^\star(x) - \nu(A_j, x)}_{\Delta_j(x)} + \underbrace{U_j - L_j}_{W_j(x)}.$$
>
> Consequently, on the same event,
>
> $$R_{\mathrm{PCS},k}(x) \leq \inf_{j\leq k}\{\Delta_j(x) + W_j(x)\} \leq \inf_{\ell\leq k}\{\Delta_{\mathrm{prop}}(\ell, x) + W_{j^\star_\ell}(x)\}.$$
>
> The statement also holds at any finite data-dependent stopping prefix covered by the original interval family.

Exact restatement of Corollary 3.

> For $\delta \in (0,1)$, if Equation (10) is valid, then with probability at least $1-\delta$, simultaneously for every k,
>
> $$R_{\mathrm{PCS},k}(x) \leq \inf_{j\leq k}\left[\Delta_j(x) + 2B_j(x) + 2\sigma_j(x)\sqrt{\frac{2\log\{2/(\delta\pi_j)\}}{n_{\mathrm{eff},j}(x)}}\right].$$

Proof map. The argument has two steps. First, the proposal-side allocation and candidate-wise conditional validity imply one simultaneous coverage event for the entire finite or countable candidate family. Second, on that event, the maximized lower endpoint at *every* prefix is compared to any candidate inside that prefix. No independence among candidate estimates is used.

Proof of Theorem 2. Let $\mathcal{J}$ be the realized finite or countable proposal index set and define

$$\mathcal{E} = \bigcap_{j\in\mathcal{J}}\{\nu_j \in [L_j, U_j]\}.$$

Conditioning on $\mathcal{G}$ and using Equation (9),

$$\mathbb{P}(\mathcal{E}^c \mid \mathcal{G}) \le \sum_{j \in \mathcal{J}} \mathbb{P}(\nu_j \notin [L_j, U_j] \mid \mathcal{G}) \le \delta \sum_{j \in \mathcal{J}} \pi_j \le \delta.$$

Hence $\mathbb{P}(\mathcal{E}) \ge 1 - \delta$ after integrating over $\mathcal{G}$. On $\mathcal{E}$, fix an arbitrary prefix k and comparator $j \le k$. By definition of $\hat{J}_k$,

$$\nu_{\hat{J}_k} \ge L_{\hat{J}_k} \ge L_j, \qquad \nu_j \le U_j,$$

so $\nu_j - \nu_{\hat{J}_k} \le U_j - L_j = W_j$. Adding and subtracting ν_j yields

$$V^\star(x) - \nu_{\hat{J}_k} \le \Delta_j(x) + W_j(x).$$

Because $\mathcal{E}$ does not depend on the chosen k or j, the inequality holds simultaneously for all prefixes and all comparators inside each prefix. Taking the infimum over $j \le k$ proves the first inequality in Equation (13). Choosing $j = j^\star_\ell$ for any $\ell \le k$ and then minimizing over ℓ proves the second. Uniformity over k immediately permits evaluation at any finite data-dependent stopping prefix τ within the covered family. □

Proof of Corollary 3. Under Equation (10), $W_j = 2r_j$. Substitute the radius into Equation (12); because the theorem is simultaneous over (k, j), taking the infimum over $j \le k$ preserves the same event for every prefix. □

Finite-horizon and anytime allocations. The primary experiments use exactly the finite-horizon family in Equation (18). For every $K < \infty$ and every $p > 0$, $\pi_j^{K,p} > 0$ and $\sum_{j=1}^{K} \pi_j^{K,p} = 1$. Thus the sensitivity grid includes $p = 1$ without pretending that a harmonic sequence defines an infinite-stream allocation. The anytime special case in Equation (19) instead satisfies

$$\sum_{j=1}^{\infty} \frac{1}{j(j+1)} = \sum_{j=1}^{\infty} \left(\frac{1}{j} - \frac{1}{j+1} \right) = 1.$$

It may therefore cover a countable stream without a known horizon. This distinction is both theoretical and implementation-level: the released primary experiment scripts call the same canonical finite-horizon allocation helper used to generate the reported results.

Relation to alpha-spending. Summable allocations inherit the alpha-spending/online FWER validity mechanism (Weinstein & Ramdas, 2020; Tian & Ramdas, 2021). The proposal oracle theorem consumes that simultaneous event at the decision level: at every prefix it controls the selected causal action relative to the best comparator in that prefix, while the support-aware corollary exposes how candidate-specific causal precision enters regret.

Why a lower endpoint is enough for selection but a full interval appears in the bound. The algorithm uses only L_j. The proof compares the chosen lower endpoint with a comparator's true value, so it also needs upper control $\nu_j \le U_j$. For symmetric intervals, the penalty is $2r_j$; for asymmetric exact intervals, the sharper object is the actual width $U_j - L_j$.

This completes the all-prefix verification used in the main text: the result consumes simultaneous interval validity and does not secretly add a stronger sampling or independence assumption.

G Matching-Rate and Allocation Proofs

This section verifies the equal-support scaling comparison in Corollary 6 and the width-optimal allocation in Theorem 7. The optimality statement remains restricted to the displayed confidence-width criterion.

Exact restatement of Corollary 6.

> Fix $\delta \in (0,1)$. Suppose the best action is among a finite family of K supported candidates, each candidate has n independent Gaussian observations with variance σ^2, $B_j = 0$, and the family budget is uniform, $\pi_j = 1/K$. Then, with probability at least $1-\delta$,
>
> $$R_{\mathrm{PCS},K} \le 2\sigma\sqrt{\frac{2\log(2K/\delta)}{n}}.$$
>
> For fixed δ and $K \ge 16$, Equation (**??**) share the same $\sqrt{\log K/n}$ dependence. This compares a high-probability PCS upper bound with an expected-regret minimax lower bound, so the claim is matching K, n scaling rather than constant-matched minimax optimality under one common risk criterion.

Exact restatement of Theorem 7.

> Fix $\delta \in (0,1)$ and a finite family. Conditional on $\mathcal{G}$, let $q_j > 0$ satisfy $\sum_{j=1}^K q_j = 1$ and let $a_j > 0$ be fixed proposal/design-side quantities that do not depend on the held-out outcome fluctuation used by the final intervals. With $c = 2/\delta$, consider
>
> $$\mathcal{C}(\pi) = \sum_{j=1}^K q_j a_j \sqrt{2\log\frac{c}{\pi_j}}, \qquad \pi_j > 0, \quad \sum_j \pi_j = 1.$$
>
> $\mathcal{C}$ is strictly convex and has a unique minimizer $\pi^\star$. It is characterized by
>
> $$\frac{q_j a_j}{\pi_j^\star\sqrt{2\log(c/\pi_j^\star)}} = \lambda^\star, \qquad j = 1,\dots,K,$$
>
> for a common $\lambda^\star > 0$. Uniform allocation is optimal if and only if $q_j a_j$ is constant across j. If J is a comparator rank drawn from the proposal-side prior q and candidate j has stochastic half-width exactly $a_j\sqrt{2\log(c/\pi_j)}$, then $2\mathcal{C}(\pi)$ is the expected stochastic full-width term; hence $\pi^\star$ minimizes that component of the decision certificate.

Proof map. The matching-rate corollary specializes the support-aware oracle bound to equal support and uniform family budget, then compares its K, n dependence with the minimax lower bound. The allocation theorem is a finite-dimensional convex program: strict convexity gives uniqueness, and the KKT equations reduce computation to a scalar multiplier.

Proof of Corollary 6. Choose the true best candidate as the comparator in Corollary 3. Then $\Delta_j = 0$, $B_j = 0$, $n_{\mathrm{eff},j} = n$, $\sigma_j = \sigma$, and $\pi_j = 1/K$, which gives Equation (26). Comparing it with Corollary 5 yields the common $\sqrt{\log K/n}$ dependence for fixed confidence. □

Proof of Theorem 7. Write $f(\pi) = \sqrt{2\log(c/\pi)}$ for $0 < \pi \le 1$. Since $c = 2/\delta > 2$,

$$f''(\pi) = \frac{2\log(c/\pi) - 1}{\pi^2\{2\log(c/\pi)\}^{3/2}} > 0.$$

Thus Equation (16) is strictly convex on the simplex interior and diverges as any $\pi_j \downarrow 0$, so it has a unique interior minimizer. The Lagrangian first-order condition is

$$-\frac{q_j a_j}{\pi_j\sqrt{2\log(c/\pi_j)}} + \lambda = 0,$$

which is Equation (17). Conversely, strict convexity makes the KKT point the unique global optimum. At a uniform allocation the denominator is common across j, so the KKT equations hold exactly when $q_j a_j$ is constant. Finally, under the theorem's exact half-width condition, the stochastic comparator width at rank j is $2a_j f(\pi_j)$; averaging over $J \sim q$ gives $2\mathcal{C}(\pi)$. In Corollary 3, the specialization is $a_j = \sigma_j(x)/\sqrt{n_{\mathrm{eff},j}(x)}$ when that scale is $\mathcal{G}$-measurable. □

Computation. The map $h(\pi) = \pi\sqrt{2\log(c/\pi)}$ is increasing on $(0, 1]$ because $2\log(c/\pi) > 1$. For a fixed multiplier λ, each KKT equation has a unique solution $\pi_j = h^{-1}(q_j a_j/\lambda)$; a scalar bisection on λ enforces $\sum_j \pi_j = 1$. A reference implementation can follow this construction directly; the theoretical result depends on the KKT characterization rather than on a particular software implementation.

The two results therefore close the main theory chain: the equal-support rate matches the lower-bound dependence in K, n, while the adaptive result is optimal only for the stated width objective conditional on proposal/design quantities.

H Causal Estimation and Interval Construction

This section expands the modular interval engine in Section 4. It shows how standard doubly robust scores, cross-fitting, support, and bias envelopes can produce the candidate-wise intervals consumed by the selection theorem.

H.1 Identification by the augmented inverse-propensity score

Write $m_a(x) = \mathbb{E}[Y \mid A = a, X = x]$ and $\mu_a(x) = \mathbb{P}(A = a \mid X = x)$. Under the assumptions in Appendix A,

$$\nu(a, x) = m_a(x).$$

For the score in Equation (5), conditioning on $X = x$ gives

$$\begin{aligned}\mathbb{E}[\psi_a(O) \mid X = x] &= m_a(x) + \frac{1}{\mu_a(x)}\mathbb{E}[\mathbf{1}\{A = a\}\{Y - m_a(x)\} \mid X = x] \\ &= m_a(x) + \mathbb{E}[Y - m_a(x) \mid A = a, X = x] \\ &= m_a(x) = \nu(a, x).\end{aligned}$$

If the outcome model is correct, the residual term has mean zero even with a misspecified propensity. If the propensity is correct, the residual term corrects a misspecified outcome model. This is the usual double-robust property (Dudík et al., 2011; Chernozhukov et al., 2018).

H.2 Orthogonality remainder

For fitted nuisances, suppress the arguments (a, x) and condition on X. Then

$$\begin{aligned}\mathbb{E}[\hat{\psi}_a \mid X] &= \hat{m}_a + \frac{\mu_a}{\hat{\mu}_a}(m_a - \hat{m}_a) \\ &= m_a + \left(1 - \frac{\mu_a}{\hat{\mu}_a}\right)(\hat{m}_a - m_a).\end{aligned}$$

Subtracting m_a gives the explicit orthogonality remainder

$$\mathbb{E}[\hat{\psi}_a - \psi_a \mid X] = \left(1 - \frac{\mu_a}{\hat{\mu}_a}\right)(\hat{m}_a - m_a). \tag{20}$$

Integrating and applying Cauchy–Schwarz,

$$\left|\mathbb{E}[\hat{\psi}_a - \psi_a]\right| \le \left\|1 - \frac{\mu_a}{\hat{\mu}_a}\right\|_2 \|\hat{m}_a - m_a\|_2 . \tag{21}$$

If overlap is bounded below and both nuisance errors are $o_p(n^{-1/4})$, their product is $o_p(n^{-1/2})$. Cross-fitting avoids empirical-process conditions that would otherwise couple these errors to the evaluation residuals (Chernozhukov et al., 2018).

H.3 Localized value and effective sample size

For a target context x, let $w_{ij}(x) \ge 0$ sum to one. The weights may depend on context distance, an action representation, or a learned partition, but must be fixed conditional on $\mathcal{G}$ for the concentration

statement below. Define

$$\hat{\nu}_j(x) = \sum_i w_{ij}(x)\hat{\psi}_{A_j,i}, \qquad n_{\text{eff},j}(x) = \left\{\sum_i w_{ij}(x)^2\right\}^{-1}. \tag{22}$$

For the displayed finite-sample radius, assume the weighted centered sum obeys the conditional MGF bound of a $\sigma_j^2 \sum_i w_{ij}^2$-sub-Gaussian variable. A sufficient special case is conditional independence of the centered evaluation terms together with individual σ_j^2-sub-Gaussian tails; a joint conditional sub-Gaussian assumption also suffices. Under this condition,

$$\mathbb{P}\left(|\hat{\nu}_j - \mathbb{E}[\hat{\nu}_j \mid \mathcal{G}]| > \sigma_j\sqrt{\frac{2\log(2/\alpha)}{n_{\text{eff},j}}}\,\middle|\,\mathcal{G}\right) \le \alpha.$$

If a deterministic envelope $B_j(x)$ satisfies $|\mathbb{E}[\hat{\nu}_j(x) \mid \mathcal{G}] - \nu_j| \le B_j(x)$, the triangle inequality gives Equation (10) with $\alpha = \delta\pi_j$. For localized or continuous contexts, $B_j(x)$ also absorbs any smoothing or localization bias. This is an assumption on the interval engine, not a consequence of cross-fitting alone.

H.4 Variance and overlap

With the true outcome regression, the random part of Equation (5) is

$$\frac{\mathbf{1}\{A=a\}}{\mu_a(x)}\{Y - m_a(x)\}.$$

Its conditional second moment is

$$\mathbb{E}\left[\frac{\mathbf{1}\{A=a\}}{\mu_a(x)^2}\{Y-m_a(x)\}^2\,\middle|\,X=x\right] = \frac{\mathrm{Var}(Y \mid A=a, X=x)}{\mu_a(x)},$$

which proves Equation (6). This is the direct mathematical link between proposal–logging mismatch and the width in Equation (14).

H.5 Alternative interval engines

The selection proof requires only simultaneous coverage. Depending on the data, $[L_j, U_j]$ may come from exact randomization inference, empirical Bernstein bounds, influence-function standard errors, a multiplier bootstrap, conformalized residuals, or a sensitivity region for hidden confounding. A black-box ensemble standard deviation is not automatically a valid interval; it must be calibrated on held-out data or justified by a model.

These derivations verify the interval-engine role assumed in Section 4; the selection theorem remains modular to any alternative engine that satisfies the same simultaneous-coverage contract.

I Extension to Unsupported or Newly Generated Text

This section makes the strict-support boundary operational. It describes one explicit transport assumption that can widen a supported interval to a new action; without such a credible assumption, the main paper's recommendation remains partial identification or abstention.

Suppose a candidate a is not in the supported library. Let $\rho_x(a)$ be an anchor action for which a causal interval $[L_\rho, U_\rho]$ is available. A possible extension assumes the explicit transport condition

$$|\nu(a,x) - \nu(\rho_x(a), x)| \le L_x d_\phi(a, \rho_x(a)). \tag{23}$$

Under Equation (23),

$$\nu(a,x) \in \left[L_\rho - L_x d_\phi(a,\rho_x(a)),\ U_\rho + L_x d_\phi(a,\rho_x(a))\right].$$

This interval can enter PCS directly. Its width equals the anchor width plus $2L_x d_\phi$. The transport constant L_x is a substantive assumption. It may be calibrated from randomized pairs that differ by known textual edits, upper-bounded from domain constraints, or varied in a sensitivity curve. If

none of these routes is credible, the correct output is a nonidentified or abstaining decision rather than a precise causal score. This boundary separates representation sharing from identification: an embedding can define d_ϕ, but cannot by itself justify L_x.

Accordingly, the unsupported-text extension remains a sensitivity/transport route, not part of the strict-support guarantee used for the headline theorem.

J Full Algorithmic Specification

This section expands the input–operation–output flow of Section 4 into an executable procedure, including confidence allocation, nuisance fitting, interval construction, selection, and abstention.

Step 1. Fit the proposer. On $\mathcal{D}_{\text{prop}}$, fit or configure Q_θ. At deployment context x, generate a deduplicated ordered sequence $A_1, \ldots, A_{K_{\max}}$.

Step 2. Choose proposal weights. For known horizon K, use $\pi_j^{K,p} = j^{-p}/H_{K,p}$ (the primary experiments use this finite-horizon rule). For an unbounded stream, use $\pi_j^{\infty} = 1/[j(j+1)]$ or another proposal-side summable allocation. Any data-adaptive weighting must be chosen from proposal/design information included in $\mathcal{G}$, not from the held-out interval noise.

Step 3. Fit causal nuisances. On folds disjoint from each evaluation observation, estimate $m_a(x)$ and $\mu_a(x)$. Known randomization probabilities replace $\hat{\mu}$ in experimental data.

Step 4. Estimate candidate values. Compute cross-fitted scores Equation (7), localize them by Equation (22), and estimate the bias/variance ingredients required by the selected interval engine.

Step 5. Construct simultaneous intervals. For candidate j, use error level $\delta\pi_j$. Record both endpoints, the effective sample size, and any semantic or sensitivity component of the width.

Step 6. Select. Return $A_{\hat{J}}$ with $\hat{J} \in \arg\max_j L_j$. Ties are broken by earlier proposal rank.

Step 7. Abstain when necessary. A deployment may require $L_{\hat{J}} \geq \tau$ or $W_{\hat{J}} \leq w_{\max}$. If the condition fails, request an experiment, revert to a safe baseline, or ask the proposer for a supported alternative.

The value-estimation cost is application dependent. With a shared outcome model, scoring K candidates is often $O(K)$ after one model fit. Exact per-action experiments scale with the total pilot sample. The theorem does not hide those costs; it only controls the statistical consequence of selection.

This specification verifies that every object used by the selector is available from the declared proposal/design and causal-evidence resources before deployment.

K Upworthy Data and Experimental Protocol

This section expands the Upworthy deployment regime in Section 5. It records the chronological split, proposal ranker, evidence/holdout partition, intervals, metrics, tuning, and leakage controls required to reproduce the reported comparison.

K.1 Dataset construction

The confirmatory CSV contains 105,551 arms in 22,743 tests. We parse timestamps, remove all tests created from June 25, 2013 through January 9, 2014, and retain tests with at least two arms and positive impressions. The resulting temporal split is shown in Table 6.

The archive reports aggregate counts, not individual records. For an arm with N impressions and C clicks, the replay samples an early count

$$C^{(e)} \sim \text{Hypergeometric}(N, C, n_e), \qquad n_e = \max\{2, \lfloor fN \rfloor\}, \tag{24}$$

then sets $C^{(h)} = C - C^{(e)}$ and $N^{(h)} = N - n_e$. This exactly partitions the observed binary outcomes without replacement. The holdout rate $C^{(h)}/N^{(h)}$ is never used for selection.

Table 6: Chronological data split after the randomization-risk exclusion.

Split	Weeks	Tests	Arms	Impressions	Clicks
Proposal train	201303–201439	10,763	50,443	189,612,968	2,653,176
Validation	201440–201500	4,040	20,334	71,791,688	897,570
Test	201501–201517	3,065	15,051	37,316,263	409,105

K.2 Historical proposal ranker

For arm i in test t, define the Jeffreys-smoothed arm CTR and pooled test CTR

$$p_{it} = \frac{C_{it} + 1/2}{N_{it} + 1}, \qquad \bar{p}_t = \frac{\sum_{r \in t} C_{rt} + 1/2}{\sum_{r \in t} N_{rt} + 1},$$

and the clipped target

$$z_{it} = \mathrm{clip}\{\mathrm{logit}(p_{it}) - \mathrm{logit}(\bar{p}_t), -3, 3\}\,.$$

The proposal ranker predicts z_{it} from the headline only. Features are word uni- and bigram TF–IDF with 30,000 maximum features, minimum document frequency 4, sublinear term frequency, and Unicode accent stripping. The regressor is averaged SGD with Huber loss (threshold 1.35), ℓ_2 penalty 3×10^{-5}, tolerance 10^{-4}, at most 30 passes, and sample weight $\sqrt{N_{it}}$. The model is fitted only on the proposal-train period. Its prediction q orders arms within every validation and test experiment. Neither validation nor test clicks enter the ordering.

This inexpensive, reproducible ranker isolates the proposal–selection interface in the primary real-data evaluation. The frozen GPT-4-derived ordering in Appendix N.4 provides a complementary LLM-based proposal signal over the same supported candidate library.

K.3 Selectors and intervals

For fixed K, the plug-in selector is

$$\hat{J}_K = \arg\max_{j \le K} \frac{C_j^{(e)}}{n_{e,j}}.$$

For PCS, the Upworthy replay fixes the evaluator horizon at $K_{\max} = 6$ and uses the same finite-horizon allocation as the code, $\pi_j = j^{-p}/H_{6,p}$ with $(\delta, p) = (0.1, 2)$ in the primary run. An experiment with fewer than six realized arms simply leaves some preallocated ranks unused. If c clicks are observed in n early impressions, candidate j receives the two-sided Clopper–Pearson interval at level $\alpha_j = \delta\pi_j$,

$$L(c, n; \alpha) = \begin{cases} 0, & c = 0, \\ B^{-1}(\alpha/2; c, n - c + 1), & c > 0, \end{cases}$$

$$U(c, n; \alpha) = \begin{cases} 1, & c = n, \\ B^{-1}(1 - \alpha/2; c + 1, n - c), & c < n. \end{cases}$$

PCS selects the largest lower endpoint. The interval targets a stable latent arm CTR under an exchangeable Bernoulli model. The held-out empirical rate is instead a disjoint finite-population evaluation quantity. Thus the replay tests decision performance without outcome reuse; it does not assert exact Clopper–Pearson coverage for the particular realized holdout partition.

K.4 Metrics and uncertainty

For a test t and replay seed s, let v_{tjs} be the holdout CTR of arm j and $j^\star_{ts}$ its best arm. Decision regret is

$$r_{ts} = v_{tj^\star_{ts}s} - v_{t\hat{J}_{ts}s}.$$

We use 20 replay seeds on the chronological validation period to select each baseline's global tuning parameters and 100 final replay seeds on the untouched test period. Reported paired 95% intervals resample test experiments 4,000 times, preserving within-test dependence among arms and replay seeds.

K.5 Reproducibility controls

All random-number generators are explicitly seeded. The proposal model is trained once per temporal split. Fixed-K baselines and PCS share the same proposal order, evidence partition, holdout partition, and test contexts. The validation period is used only to choose the global fixed K; the final test period is untouched until all method choices are fixed. The archive's noncausal metadata fields "winner," "first place," and "significance" are not used.

L Complete Baseline Audit and Experimental Results

This section verifies the baseline-fairness and statistical claims summarized in Section 5. All selectors share the proposal order and evidence/holdout inputs within a replay; tuning uses only the designated validation information.

L.1 Persuasion replay construction

The source study contains 25,982 participants overall. For the causal-selection replay we retain only AI-treatment observations with supported generated messages, giving 19,529 observations from 17,713 participants, 720 messages, 24 language models, and ten issues. For held-out issue h, each model receives a historical proposal score equal to its average issue-centered randomized outcome on the other nine issues. Messages in h inherit their model's score; released task-completion, on-topic, legibility, and valence annotations break within-model ties. No participant outcome from h enters its proposal order.

For every message and replay seed, half of participants are reserved as a fixed evaluation set. The other half is the evidence pool, of which 25%, 50%, or 100% is exposed to the selector; budgets are nested within a seed. All 72 supported messages per issue remain eligible. The primary replay uses 200 seeds, $\delta = 0.2$, and $p = 2$.

Persuasion interval engine. For candidate j with evidence mean $\bar{Y}_j$, standard error $s_j/\sqrt{n_j}$, and confidence weight π_j, the lower endpoint is the Studentized bound

$$L_j = \bar{Y}_j - t_{1-\delta\pi_j/2,n_j-1}\frac{s_j}{\sqrt{n_j}}.$$

Pointwise LCB sets $\pi_j = 1$, uniform LCB uses $\pi_j = 1/K$, and PCS uses the rank-power allocation. The held-out half is never used to construct these scores.

L.2 Baseline definitions and validation budgets

All baselines receive exactly the same order, evidence partition, and holdout. Tuned fixed depth selects the validation-optimal K from a prespecified grid. Variance penalty selects $\arg\max_j\{\hat{\nu}_j - \lambda\widehat{\mathrm{se}}_j\}$, and rank+variance additionally subtracts $\gamma \log j$. Upworthy tunes on its chronological validation period; persuasion tunes separately for every held-out issue using only the other nine issues (leave-one-issue-out). Thus no target-issue holdout is used to tune its comparator.

L.3 Full Upworthy strong-baseline table

The paired PCS improvement over tuned fixed depth is 0.042 bp (95% bootstrap interval $[-0.050, 0.137]$), 0.283 $[0.127, 0.442]$, 0.651 $[0.424, 0.889]$, and 0.949 bp $[0.671, 1.230]$ as evidence increases. The validation-selected rank+variance control is the strongest point estimate at 0.5% and 2%, preventing a claim of uniform empirical dominance.

L.4 Full persuasion strong-baseline table

Relative to leave-one-issue-out tuned fixed depth, PCS changes held-out value by $+0.43$, $+1.00$, and $+1.48$ points. The ten-issue bootstrap intervals are $[-1.67, 2.44]$, $[-1.29, 3.33]$, and $[-0.56, 3.68]$; the corresponding paired-t intervals are $[-2.07, 2.93]$, $[-1.76, 3.77]$, and $[-1.09, 4.05]$. Six of ten issue-level gains are positive at every budget, and exact sign-flip p-values are 0.760, 0.424, and 0.230.

Table 7: Upworthy mean decision regret in basis points over 100 final replay seeds. Lower is better. Bold marks the best point estimate in each evidence regime.

Method	0.5%	1%	2%	5%
Tuned fixed-K	32.98	32.58	32.07	30.12
Plug-in full	33.06	32.65	32.16	30.12
Pointwise LCB	33.06	32.65	32.16	30.12
Uniform LCB	33.06	32.65	32.16	30.12
CV variance penalty	33.06	32.65	32.16	30.12
CV rank+variance	**32.78**	32.37	**31.32**	29.39
PCS	32.93	**32.30**	31.42	**29.17**

Table 8: Held-out attitude value on the randomized LLM-persuasion replay over 200 seeds. Higher is better.

Method	25%	50%	100%
Tuned fixed-K	57.33	57.08	57.38
Plug-in full	56.64	55.47	54.89
Pointwise LCB	56.67	56.11	56.18
Uniform LCB	57.50	57.78	58.46
CV variance penalty	56.85	56.03	56.57
CV rank+variance	**58.09**	57.30	56.10
PCS	57.75	**58.08**	**58.86**

The point estimates support the proposed reliability pattern, but the small number of issue-level contexts does not resolve a real-data superiority claim.

M Full Sensitivity, Robustness, Efficiency, and Error Analyses

This section expands the stress axes reported in Secs. 5.3–5.4. The purpose is to quantify where the mechanism persists or fails, not to upgrade the bounded main-text claims into universal robustness or generalization statements.

M.1 Controlled logging-support stress on an RCT-derived shell

We impose outcome-independent rank-based thinning on the persuasion evidence pool while leaving the randomized holdout untouched. At full evidence the resulting held-out values are: The simultaneous selectors remain much less exposed to the optimistic plug-in errors created by sparse late-rank evidence. The experiment is a controlled support intervention, not a claim about naturally occurring confounding.

Early:late support skew	Tuned fixed-K	Plug-in	Uniform LCB	PCS
1:1	57.43	55.29	58.75	**59.07**
2:1	56.93	53.14	57.73	**58.09**
4:1	57.09	54.32	57.81	**57.98**
8:1	57.69	54.18	57.59	**57.83**

Table 9: RCT-derived support imbalance. Selection evidence is thinned without using outcomes; the audit half is unchanged.

M.2 Small-cluster inference audit for persuasion

The persuasion study provides ten issue-level deployment contexts. For PCS versus tuned fixed depth, the issue-bootstrap, paired-t, leave-one-issue-out, and exact sign-flip diagnostics are all generated from the same ten paired issue effects. At full evidence the mean gain is $+1.48$ points,

but the bootstrap interval $[-0.56, 3.68]$ and sign-flip $p = 0.230$ remain unresolved. The resulting evidence is therefore reported as mechanism-consistent validation rather than as a statistically resolved performance gain.

M.3 Why several Upworthy uncertainty baselines coincide

Within many Upworthy tests the archived arm allocations are close to balanced, and the validation-selected variance-only penalty is zero in all four primary regimes. Consequently, plug-in, pointwise LCB, uniform LCB, and the variance-only control can choose the same arm often enough to have nearly identical aggregate regret. The separately tuned rank+variance control remains distinct and is a stronger comparator in the 0.5% and 2% regimes.

M.4 Hyperparameter sensitivity

On persuasion we evaluate every combination $\delta \in \{0.05, 0.1, 0.2, 0.4\}$ and $p \in \{1, 1.5, 2, 2.5, 3\}$. The point-estimate PCS gain over the leave-one-issue-out tuned fixed-depth selector ranges from -0.28 to $+0.54$ at 25% evidence, $+0.89$ to $+1.26$ at 50%, and $+1.22$ to $+1.57$ at full evidence. We do not attach twenty post-hoc significance claims to this grid; its purpose is sensitivity of the qualitative regime, not multiplicity-free hypothesis testing.

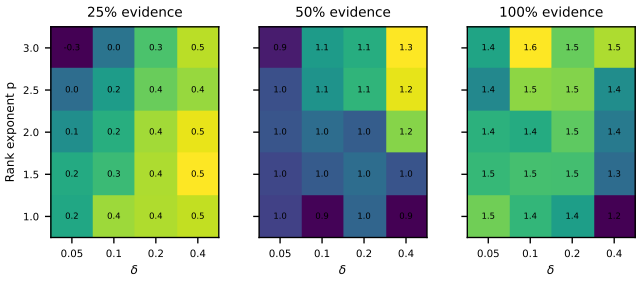

Figure 6: Full confidence-spending sensitivity grid for persuasion. Values are point-estimate gains over the leave-one-issue-out tuned fixed-depth selector across family error δ and rank exponent p.

M.5 Proposal-order ablation

Replacing only the learned/historical proposal order by an independent random order raises Upworthy PCS regret by 2.95, 2.25, 1.50, and 1.21 bp across 0.5%, 1%, 2%, and 5% evidence. On persuasion, the learned-order advantage is 1.04, 1.13, and 1.40 held-out points across 25%, 50%, and 100% evidence. This isolates the value of proposal information from the value of simultaneous certification.

M.6 Action-source-family robustness

Table 10: PCS gain over a separately leave-one-issue-out tuned fixed-depth selector after restricting the persuasion action library to one model family. Mixed signs are a scope boundary.

Family	25%	50%	100%
Pythia	+0.20	+2.00	+3.32
Qwen1.5	+0.91	+2.52	+0.42
Llama	-0.68	+0.45	-0.12
Yi	-0.10	-0.57	-0.38
Falcon	+3.03	+2.58	+2.42

The mixed signs prevent a generic cross-family improvement claim. They are consistent with the theory's dependence on proposal quality and support rather than on a specific language-model family.

M.7 CONTROLLED EVIDENCE-NOISE STRESS

At full persuasion evidence, the historical rank-1 action is already strong. Adding mean-zero Gaussian noise to evidence scores while leaving the disjoint holdout unchanged gives PCS minus rank-1 values of -0.28, -0.56, -0.84, and -1.04 points at 0, 0.25, 0.5, and 1.0 issue-SD corruption. This deliberately exposes a failure regime: when the proposer has already concentrated value near the first rank and the evidence engine is noisy, opening further comparisons can hurt.

M.8 SELECTOR-ONLY EFFICIENCY

The main text reports the claim-relevant selector latency and memory headline. The full scaling curve is shown here; candidate generation, causal estimation, and data loading are excluded, and the adaptive allocation solver is accounted for separately.

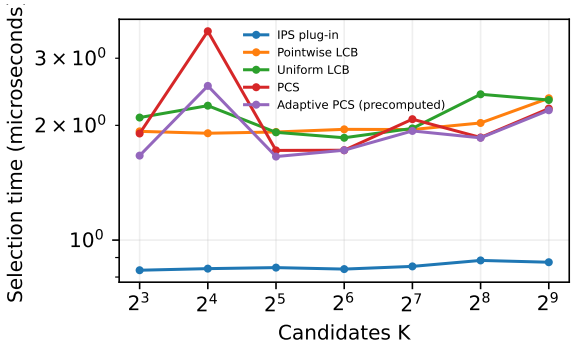

Figure 7: Selector-only CPU scaling. Adaptive allocation is precomputed for the selection-time curve; its separate solver cost is reported in the accompanying text/table.

Across $K = 32$–512, canonical PCS takes approximately 1.7–2.2 microseconds per selection in the reported microbenchmark and stays comparable to uniform LCB. At $K = 512$, traced peak working memory is about 8.2 KiB. The width-optimal adaptive allocation is computed separately from held-out outcomes; the reference nested-bisection solver takes roughly 47–107 ms over $K = 32$–512 and is not included in the vanilla PCS timing. All methods have zero trainable selector parameters. The benchmark excludes candidate generation, data collection, proposer training, and evaluator fitting.

M.9 QUALITATIVE CASES AND FAILURE TAXONOMY

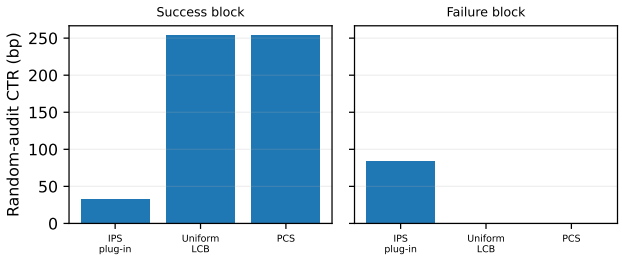

Figure 8: One OBD block where pessimistic selection avoids an optimistic production estimate and one block where conservatism misses the better randomized-audit action.

In the success block (men, position 2, 2019-11-30), plug-in selects item 8 with production estimate 0.0249 and SE 0.0128; the randomized audit CTR is 0.00325. PCS and uniform LCB select item 13 with estimate 0.00838 and SE 0.000385; its randomized audit CTR is 0.0254. In the failure block (men, position 1, 2019-11-26), PCS stays at rank 1 (item 31) and audits at 0, whereas plug-in selects rank 30 (item 29) and audits at 0.00836. The error taxonomy is therefore structural: proposal miss, noisy optimistic winner, or conservative rejection of a genuinely valuable uncertain action.

N THEORY-FACING AND EXTERNAL-EVIDENCE AUDITS

This section supplies detailed numbers and auxiliary checks for the theory-facing curves and the external randomized audit summarized in the main text. It verifies observable consequences and calibration sensitivity without changing the theorem assumptions.

N.1 MULTIPLICITY-SCALING DIAGNOSTIC

Figure 4 promotes the two central Gaussian diagnostics to the main paper so the observable consequences of Theorem **??** are visible without opening the appendix. With $\Delta = \sigma = 1$, the exact plug-in construction increases expected regret from 0.159 at $K = 2$ to essentially 1 by $K = 64$. In a separate Monte Carlo best-arm identification experiment, the empirical gap required for 90% recovery increases from 0.26 at $K = 2$ to about 0.63 at $K = 512$, consistent with the multiplicity scaling in Theorem 4.

N.2 PROPOSAL-PRIOR/SUPPORT DIFFERENTIAL STRESS

The simulation varies retained proposal signal in $\{0, 0.25, 0.5, 0.75, 1\}$ and early:late support skew in $\{1, 2, 4, 8\}$. Uniform LCB is the zero reference. Adaptive proposal/support allocation is negative when proposal information is absent, near zero around intermediate signal, and positive when the proposal order is highly informative. At signal 1 its value gain over uniform is about 0.012–0.017 across the four support skews. This is an empirical reminder that Theorem 7 optimizes its stated width criterion conditional on q_j, a_j; misspecified proposal priors can still worsen downstream selection.

N.3 MEASURED-CONFOUNDING REWEIGHTING AUDIT

We use the randomized persuasion data as a controlled shell and impose deterministic selection tilts based on observed participant ideology and action identity. This creates a known measured-confounding mechanism without changing the untouched randomized target. Mean absolute distortion of naive action means rises from 0 at $\kappa = 0$ to 5.64, 10.32, 15.80, and 18.30 outcome points at $\kappa = 0.5, 1, 2, 3$. Applying the known inverse tilt returns the randomized action means up to numerical precision. The experiment verifies the role of measured-confounder adjustment; it is not evidence about naturally hidden confounding.

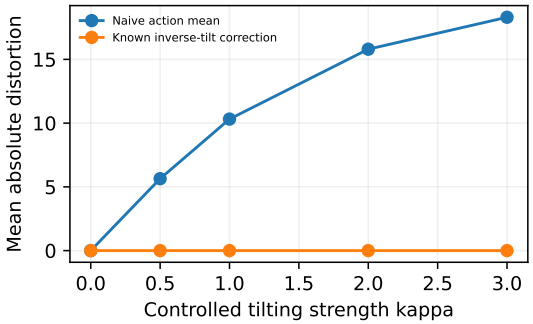

Figure 9: Controlled measured-confounding diagnostic derived from the randomized persuasion sample.

N.4 RELEASED GPT-LABEL PROPOSER

The persuasion release contains outcome-free GPT-based quality annotations for each supported message. We use task completion, topicality, legibility, and stance/valence correctness lexicographically to define a frozen alternative proposal order; no new model call and no held-out participant outcome enters this order. Averaged over issues and replays, prefix-1 oracle value is 59.22 for the GPT-label order, 59.14 for the historical order, and 58.37 for a fixed random order. The ordering is not uniformly better: historical/random prefixes overtake it at several larger K, and all orders converge at $K = 72$.

This experiment tests the proposal–selection interface with a released model-derived signal while preserving randomized support.

N.5 Natural production logging with randomized audit

Outcome separation and estimand. OBD contains 24,200,424 BTS rows and 2,691,861 Uniform Random rows, totaling 26,892,285 impressions. For each campaign, position, day, and item, current-day BTS observations form the IPS estimate

$$\hat{V}_j = \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{1}\{A_i = j\}Y_i}{\mu_i(j \mid X_i)}$$

and an empirical standard error over the same n logged rows, including the implicit zeros for rows assigned to other items. Historical BTS days provide only the proposal order. Rank-power PCS uses $\pi_j \propto j^{-2}$. Adaptive PCS uses the theorem's proposal-prior/support solver with a decreasing rank prior and the production-side variance scale. Uniform LCB sets $\pi_j = 1/K$; pointwise LCB uses a marginal critical value. These empirical IPS intervals are selection scores, not a finite-sample proof of Equation (9). The time-matched Uniform Random outcome table is accessed only after the item is selected.

Rolling and static audits. The primary rolling audit uses each day after the first as a target day, producing $3 \times 3 \times 6 = 54$ campaign-position-day blocks. For inference, positions are averaged within each campaign-day, leaving 18 clusters. A secondary static audit uses the first three BTS days for proposal ranking and the final four days for selection/audit, producing nine campaign-position settings.

Table 11: Full OBD rolling audit. Random-audit CTR is computed only after selection.

Selector	Blocks	Random-audit CTR	Audit regret	Mean rank
Rank-1	54	0.4258%	137.98 bp	1.00
IPS plug-in	54	0.3939%	141.17 bp	26.70
Pointwise LCB	54	0.5488%	125.67 bp	19.17
Uniform LCB	54	0.6780%	112.76 bp	10.30
PCS	54	0.6609%	114.47 bp	7.67
Adaptive PCS	54	**0.6781%**	**112.75 bp**	9.11

Relative to IPS plug-in, PCS gains 26.70 bp with a 95% 18-cluster interval $[10.99, 42.42]$; adaptive PCS gains 28.42 bp $[12.52, 44.32]$; uniform LCB gains 28.41 bp $[13.19, 43.64]$. The adaptive-minus-uniform point difference is only about 0.008 bp, so OBD supports simultaneous uncertainty control but not nonuniform-allocation superiority. With only six calendar-day clusters, the PCS interval widens to $[1.78, 51.63]$ while preserving the direction.

Interval-scale sensitivity. Multiplying every production standard error by $c \in \{0.5, 0.75, 1, 1.25, 1.5, 2\}$ preserves a positive PCS gain over plug-in. The 18-cluster intervals range from $+21.11$ bp $[9.13, 33.09]$ at $c = 0.5$ to $+29.59$ bp $[11.57, 47.61]$ at $c = 2$. This is a calibration stress, not a replacement for formal interval validity.

O Full Experimental Reproducibility Map

This section documents the intended end-to-end experimental pipeline and the artifact boundary relevant to reproduction. The components below specify the programs required to regenerate the reported analyses from raw data; panel-level reproduction and reference-code checks use a smaller artifact set:

- `code/run_upworthy.py`: archive filtering, chronological split, headline-only proposal ranker, validation tuning, hypergeometric evidence/holdout replay, strong baselines, paired bootstrap, and order ablation.

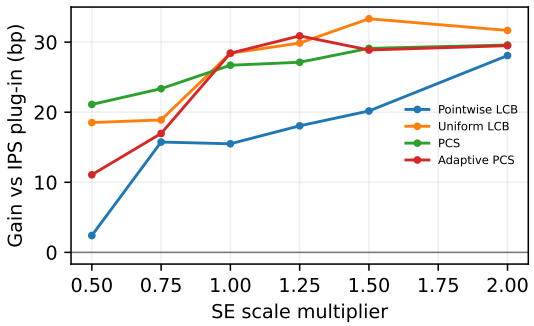

Figure 10: OBD randomized-audit gain versus IPS plug-in as the production standard-error scale is perturbed.

- `code/run_persuasion.py`: supported AI-treatment filtering, leave-one-issue-out proposal/tuning protocol, strong baselines, issue-level inference, sensitivity, order ablation, model-family restrictions, and released GPT-label proposer coverage.
- `code/run_persuasion_stress.py`: outcome-independent support thinning, controlled evidence-noise stress, and measured-confounding reweighting diagnostic.
- `code/aggregate_obd_fast.sh` and `code/audit_obd_from_stats.py`: streaming aggregation of the full OBD ZIP, rolling/static BTS selection with delayed Uniform-Random audit, cluster inference, and interval-scale sensitivity.
- `code/run_simulations.py`: exact winner's-curse curve, multiplicity Monte Carlo, and proposal-signal/support differential stress.
- `code/benchmark_efficiency.py`: selector-only CPU timing and traced memory; adaptive allocation cost is reported separately.
- `code/make_figures.py`: all manuscript figures from the generated CSV results.
- `code/verify_results.py` and `tests/test_core.py`: raw-count/result integrity checks and core selector/allocation/interval unit tests.

A complete rerun should set explicit random seeds, record SHA256 hashes and exact input paths for the raw-data archives in `DATA_PROVENANCE.md`, and populate `results/` with the generated aggregate CSVs. The current artifact boundary therefore distinguishes panel/formula verification from an end-to-end rerun that also requires the original experimental programs, provenance metadata, and third-party raw archives.

P Supporting Equal-Support Statements

The following two source statements support Section 4. Their mathematical content is unchanged; their full proofs remain in Appendices E and G. They compare different risk criteria and therefore support rate dependence, not constant-matched minimax optimality.

Corollary 5 (Minimax simple-regret price)**.** *For $K \geq 16$, consider the same Gaussian observation model and the K instances with one best arm of gap*

$$\Delta_K = \frac{\sigma}{2}\sqrt{\frac{\log K}{n}}.$$

Then every selector obeys

$$\inf_{\widehat{J}} \sup_{v \in [K]} \mathbb{E}_v[\mu_v - \mu_{\widehat{J}}] \geq \frac{\sigma}{4}\sqrt{\frac{\log K}{n}}. \tag{25}$$

Corollary 6 (Matching comparison rate under equal support)**.** *Fix $\delta \in (0,1)$. Suppose the best action is among a finite family of K supported candidates, each candidate has n independent Gaussian*

observations with variance σ^2, $B_j = 0$, and the family budget is uniform, $\pi_j = 1/K$. Then, with probability at least $1 - \delta$,

$$R_{\mathrm{PCS},K} \le 2\sigma \sqrt{\frac{2\log(2K/\delta)}{n}}. \tag{26}$$

*For fixed δ and $K \ge 16$, Equation (**??**) share the same $\sqrt{\log K/n}$ dependence. This compares a high-probability PCS upper bound with an expected-regret minimax lower bound, so the claim is matching K, n scaling rather than constant-matched minimax optimality under one common risk criterion.*

Q Visual Provenance and Artifact Boundaries

Figures 1–3 are explanatory schematics. The four-action mechanism uses illustrative estimates $(0.31, 0.58, 0.82, 0.50)$ and radii $(0.07, 0.10, 0.36, 0.25)$; it is not an experiment. Figure 4(a) evaluates the unchanged analytic formula. Panel (b) uses the nine reported Monte Carlo display points reconstructed from the available vector figure and calibrated against its reference curve. These coordinates are treated as displayed diagnostic values rather than raw simulation output. Figure 5 uses the reported means and does not add unreported standard errors. The original sixteen vector figures are retained in the companion archive.

Each newly drawn panel has a standalone Python entry point and PDF/SVG/PNG outputs. The artifact manifest distinguishes reported summaries, analytic evaluations, schematic values, and recovered display geometry. Independent reruns of the original simulations and raw-data replays require the missing original programs, seed lists, and input/result provenance described in Appendix O. The new reference-code tests verify only the mathematical operations they exercise; they do not validate the missing empirical pipeline.