

Ranking-and-Selection with Multiple Correct Answers and Non-Answerable Estimates

Qiaoqiao Wang¹ and Wei You¹

¹Dept. of Industrial Engineering and Decision Analytics, The Hong Kong University of Science and Technology, Hong Kong, China

June 20, 2026

Abstract

We study fixed-precision ranking-and-selection in structured settings where the answer may be non-unique and where noisy estimates may temporarily admit no valid answer at all. This phenomenon arises naturally in problems such as multi-fidelity ranking-and-selection and identifying a Condorcet winner from pairwise comparisons. To address this, we propose a unified framework based on answer-wise acceptance sets, restricted generalized likelihood ratio stopping, and an answer-pitfall decomposition that yields a max-max-min characteristic value and a common sampling principle. We introduce ENDS, a general procedure that combines estimation, nomination, pitfall detection, and cost-aware information-directed selection. We instantiate ENDS for various problems by deriving explicit formulas. Extensive numerical experiments show that this unified recipe performs well across a broad range of pure-exploration problems and offers a practical framework and proof-of-concept algorithmic recipe.

1 Introduction

Many ranking-and-selection (R&S) problems arise in settings where information is noisy, structured, and expensive. In multi-fidelity experimentation, one can query cheap but biased proxies or expensive high-fidelity measurements; in dueling bandits, feedback arrives only through pairwise comparisons rather than direct rewards. These models are increasingly natural in engineering design, simulation optimization, preference learning for LLMs, and human-in-the-loop evaluation, where absolute scores are often unavailable or prohibitively costly and decisions must be made with a prescribed level of confidence.

What makes these settings especially challenging is that the usual single-winner template is no longer sufficient. First, the answer map may be set-valued: in good-alternative or subset-selection problems, several answers can be simultaneously correct. Second, even when the true instance is answerable, a noisy estimate may temporarily fall outside the answerable set. In multi-fidelity ranking-and-selection, empirical means may violate the bias constraints; in dueling bandits, estimated

preferences may induce ties or cycles so that no Condorcet winner exists. In such situations, the empirical-best answer can be undefined exactly when the algorithm still needs a principled way to sample, stop, and decide.

The fixed-confidence pure-exploration literature has developed sharp lower bounds and asymptotically efficient procedures for best-arm identification, and has gradually expanded to structured settings such as multiple correct answers, multi-fidelity observations, and pairwise preference feedback (Garivier and Kaufmann 2016, Degenne and Koolen 2019, Poiani et al. 2024, Chen et al. 2020). In parallel, the ranking-and-selection literature has long emphasized the close connection between fixed-precision procedures and sequential hypothesis testing, and has developed important variants around good selection, subset selection, multi-fidelity simulation, and pairwise-comparison settings (Wang et al. 2024, Hong et al. 2021, Eckman and Henderson 2018, Xu et al. 2016, Ren et al. 2020). Related questions also arise in economics (Ramdas et al. 2020, Kasy and Sautmann 2021), where sequential experimentation, information acquisition, and stopping are studied under objectives that differ from bandit sample complexity but face many of the same conceptual tensions between exploration, confidence, and irreversible choice.

Despite this progress, the existing treatments remain largely problem-specific. Standard formulations typically assume either a unique correct answer or an always-answerable estimate, and therefore do not directly address the joint presence of multiple correct answers and non-answerable estimates. This gap matters algorithmically: it affects how one defines the stopping statistic, how one nominates a provisional answer, and how one allocates sampling effort when no empirical winner is currently well defined.

This paper takes a step toward that unification. Building on the pitfall-adapted nomination perspective of Qin and You (2026), we lift the framework from unique-answer settings to answer-wise acceptance sets and answer-wise pitfalls, so that the learner reasons over all candidate answers rather than only around a single true answer. We view the paper as a proof of concept that a single structural recipe can transfer across broad models and problems. The main contributions and organization of the paper are as follows.

- In Section 2, we propose a unified framework for *cost-aware* fixed-precision R&S with a *set-valued answer map* and an *answerable region*.
- In Section 3, we extend Qin and You (2026) with an answer-wise layer: rejection regions are decomposed by candidate answers and answer-wise pitfalls, yielding a restricted-GLR stopping rule and a max-max-min characteristic value that motivates the performance limit and the corresponding sampling principle. The proposed ENDS procedure combines estimation, answer-wise nomination, pitfall detection, and cost-aware information-directed selection in a portable way across problems.
- In Section 3, we derive explicit instantiations for representative problems, including good-alternative selection, multi-fidelity ranking-and-selection, and Condorcet-winner dueling bandits.

- Through broad numerical experiments in Section 4, we show that the unified recipe is practically effective across qualitatively different exploration problems, supporting the case for a common algorithmic language beyond any one model class.

2 Problem formulation

Let Θ be the ambient space of valid parameters. We allow Θ to be structured; see Example 3. Let \mathcal{U} be a finite set of measurement types. At each round n , the learner chooses $U_n \in \mathcal{U}$ and observes $Y_n \sim P_{\theta, U_n}$. Each measurement type $u \in \mathcal{U}$ incurs measurement cost $C_u(\theta) > 0$. For every $u \in \mathcal{U}$, let $N_{n,u} \triangleq \sum_{\ell=1}^n \mathbb{1}\{U_\ell = u\}$ and $p_{n,u} \triangleq N_{n,u}/n$. We write $\mathbf{p}_n = (p_{n,u})_{u \in \mathcal{U}}$ for the measurement allocation, and

$$\bar{C}_\theta(\mathbf{p}) \triangleq \sum_{u \in \mathcal{U}} p_u C_u(\theta)$$

for the average cost per sample under allocation \mathbf{p} . The cost spent after n rounds is

$$B_n = \sum_{u \in \mathcal{U}} N_{n,u} C_u(\theta) = n \bar{C}_\theta(\mathbf{p}_n).$$

The goal is to identify a correct answer from the set of possible answers \mathcal{A} , by allocating samples to the set of measurement types \mathcal{U} and observing noisy measurements. We allow the answer map to be set-valued. For every parameter $\theta \in \Theta$, we let $\mathcal{A}^*(\theta) \subseteq \mathcal{A}$ denote the set of correct answers at θ . In multiple-correct-answer problems, one can have $|\mathcal{A}^*(\theta)| > 1$.

We say that an instance θ is *answerable* if $\mathcal{A}^*(\theta) \neq \emptyset$. Define the *answerable set* as

$$\Theta^* \triangleq \{\theta \in \Theta : \mathcal{A}^*(\theta) \neq \emptyset\}.$$

We require that the ground truth $\theta \in \text{interior}(\Theta^*)$, which excludes boundary instances at which arbitrarily small perturbations may change the set of correct answers, making exact identification statistically ill-posed. To accommodate genuinely non-answerable instances, we may augment the answer set to include a null answer $a = \emptyset$, representing the conclusion that no substantive answer exists. During learning the empirical estimate may leave the answerable set Θ^* even if the true parameter does not, in which case an empirical answer need not exist. We refer to such estimates as a *non-answerable estimate*. For each answer $a \in \mathcal{A}$, define its acceptance set by

$$\Theta_a^* \triangleq \{\vartheta \in \Theta^* : a \in \mathcal{A}^*(\vartheta)\}.$$

Let $\text{Alt}_a \triangleq \Theta^* \setminus \Theta_a^*$ denote all answerable instances for which answer a is not correct.

In this paper, we focus on the fixed-precision setting. The goal is to design a stopping rule τ_δ and a decision rule \hat{a}_{τ_δ} such that

$$\mathbb{P}_\theta(\hat{a}_{\tau_\delta} \in \mathcal{A}^*(\theta)) \geq 1 - \delta \quad \text{for every } \theta \in \Theta^*,$$

while designing a sampling rule that minimizes the expected total cost $\mathbb{E}_\theta[B_{\tau_\delta}]$.

2.1 Examples

Example 1 (Selection of a good alternative). *Fix a tolerance $\varepsilon > 0$. We consider selection of a good alternative from a set of K alternatives (Eckman and Henderson 2018). This is also referred to as ε -best-arm identification (Jourdan et al. 2023) or (ε, δ) -PAC learning. Each measurement type is an alternative, and $\mathcal{U} = [K]$. Sampling alternative i produces a noisy observation with mean θ_i , and incurs cost $C_i(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_i)_{i \in [K]}$. The ambient parameter is Θ , the set of all possible mean vectors. For example, if the measurements are Gaussian, then the ambient parameter space is $\Theta = \mathbb{R}^K$. The set of possible answers is $\mathcal{A} = [K]$, and the set of correct answers is*

$$\mathcal{A}^*(\boldsymbol{\theta}) \triangleq \{i \in [K] : \theta_i \geq \max_{j \in [K]} \theta_j - \varepsilon\}.$$

Thus every instance is answerable, so $\Theta^ = \Theta$. This problem therefore exhibits multiple correct answers but no non-answerable estimates. The acceptance set for answer i is*

$$\Theta_i^* \triangleq \{\boldsymbol{\vartheta} \in \Theta : \vartheta_i \geq \vartheta_j - \varepsilon, \forall j \neq i\}.$$

Example 2 (Multi-fidelity ranking-and-selection). *In multi-fidelity ranking-and-selection/BAI, each alternative $i \in [K]$ can be sampled at several fidelity levels $m \in [M]$ with fidelity-dependent costs C_m , with higher fidelities being more expensive and less biased (Poiani et al. 2024). The measurement set is then $\mathcal{U} = [K] \times [M]$, with measurement $u = (a, m)$ representing that the learner samples alternative a at fidelity level m . The ambient parameter is $\boldsymbol{\theta} = (\theta_{a,m})_{a \in [K], m \in [M]} \in \Theta$, where Θ denotes the set of all possible mean tables. The multi-fidelity model class Θ_{MF} consists of all mean tables that satisfy the multi-fidelity constraints $|\theta_{i,m} - \theta_{i,M}| \leq \xi_m$, i.e.,*

$$\Theta_{\text{MF}} \triangleq \{\boldsymbol{\theta} \in \Theta : |\theta_{a,m} - \theta_{a,M}| \leq \xi_m, \forall a \in [K], m \in [M]\},$$

where ξ_m is the known bias level for fidelity m . Assume throughout that $\xi_M = 0$. The set of possible answers is $\mathcal{A} = [K]$. The goal is to identify the correct answer defined as

$$\mathcal{A}^*(\boldsymbol{\theta}) = \begin{cases} \operatorname{argmax}_{a \in [K]} \theta_{a,M} & \text{if } \boldsymbol{\theta} \in \Theta_{\text{MF}}, \\ \emptyset & \text{if } \boldsymbol{\theta} \notin \Theta_{\text{MF}}. \end{cases}$$

We assume that there is no tie at the highest fidelity.¹ The answerable set is

$$\Theta_{\text{MF}}^* \triangleq \{\boldsymbol{\theta} \in \Theta : \mathcal{A}^*(\boldsymbol{\theta}) \neq \emptyset\} = \Theta_{\text{MF}} \subsetneq \Theta,$$

i.e., the set of mean tables that satisfy the multi-fidelity constraints. The acceptance set for answer i is $\Theta_i^ \triangleq \{\boldsymbol{\vartheta} \in \Theta_{\text{MF}}^* : \vartheta_{i,M} > \vartheta_{j,M}, \forall j \neq i\}$.*

¹We exclude ties at the highest fidelity because such instances lie on the boundary of identifiability. If two alternatives are tied at fidelity M , then an arbitrarily small perturbation can make either one uniquely optimal. In these boundary cases, the GLR is zero, so exact identification is statistically impossible. A standard remedy is to relax the objective and seek an ε -good alternative instead; see Example 1.

Example 3 (Dueling bandits with a Condorcet winner). *In dueling bandits, the observations are not direct measurements of individual alternatives, but rather binary preferences between pairs of alternatives. The measurement set is the set of unordered pairs $\mathcal{U} = \{\{i, j\} : 1 \leq i < j \leq K\}$, where measurement $u = \{i, j\}$ means that the learner requests a duel between alternatives i and j . With the convention $i < j$, the observation is Bernoulli with mean $\theta_{i,j}$. The ambient parameter space is*

$$\Theta_{\text{DB}} \triangleq \{\boldsymbol{\theta} = (\theta_{i,j})_{1 \leq i, j \leq K} \in [0, 1]^{K \times K} : \theta_{i,j} = 1 - \theta_{j,i}, \forall i, j\}.$$

The set of possible answers is $\mathcal{A} = [K]$. The goal is to identify the Condorcet winner

$$\mathcal{A}^*(\boldsymbol{\theta}) \triangleq \{i \in [K] : \theta_{i,j} > 1/2, \forall j \neq i\}.$$

A Condorcet winner need not exist if the preference matrix contains cycles, but it is unique if it exists.² The answerable set is

$$\Theta_{\text{CW}}^* \triangleq \{\boldsymbol{\theta} \in \Theta_{\text{DB}} : |\mathcal{A}^*(\boldsymbol{\theta})| = 1\} \subsetneq \Theta_{\text{DB}}.$$

The acceptance set for answer i is

$$\Theta_i^* \triangleq \{\boldsymbol{\vartheta} \in \Theta_{\text{CW}}^* : \vartheta_{i,j} > 1/2, \forall j \neq i\}.$$

See also Chen et al. (2020), Haddenhorst et al. (2021).

In Examples 2 and 3, the answerable set $\Theta^* = \bigcup_{a \in \mathcal{A}} \Theta_a^*$ is a proper subset of the ambient model Θ . Consequently, during learning the empirical or posterior estimate may leave Θ^* , in which case an empirical answer need not exist. We refer to such estimates as *non-answerable estimates*.

2.2 Stopping rule via GLR statistics

We follow the sequential hypothesis-testing perspective in Hong et al. (2021), Wang et al. (2024). In the presence of non-answerable estimates, the natural testing statistic is the restricted generalized likelihood ratio (GLR) that compares the best fit inside an acceptance set with the best *answerable parameter* outside.

Let $\widehat{\boldsymbol{\theta}}_n$ denote an unconstrained maximum likelihood estimator (MLE) in the ambient parameter space Θ . For the one-parameter exponential-family models, the negative log-likelihood ratio relative to $\widehat{\boldsymbol{\theta}}_n$ can be written as a sum of coordinatewise Kullback–Leibler (KL) divergences, i.e., $\Gamma(\mathbf{N}_n; \widehat{\boldsymbol{\theta}}_n, \boldsymbol{\vartheta})$, where

$$\Gamma(\mathbf{w}; \boldsymbol{\theta}, \boldsymbol{\vartheta}) \triangleq \sum_{u \in \mathcal{U}} w_u \text{KL}(P_{\boldsymbol{\theta}, u} \| P_{\boldsymbol{\vartheta}, u}), \quad \text{for } \mathbf{w} \in \mathbb{R}_+^{\mathcal{U}},$$

where $\text{KL}(\cdot \| \cdot)$ is the KL divergence between probability distributions. The GLR statistic for answer a is

$$Z_n(a) \triangleq Z(a; \mathbf{N}_n, \widehat{\boldsymbol{\theta}}_n), \quad \text{where } Z(a; \mathbf{p}, \boldsymbol{\theta}) \triangleq \inf_{\boldsymbol{\vartheta} \in \text{Alt}_a} \Gamma(\mathbf{p}; \boldsymbol{\theta}, \boldsymbol{\vartheta}) - \inf_{\boldsymbol{\vartheta} \in \Theta_a^*} \Gamma(\mathbf{p}; \boldsymbol{\theta}, \boldsymbol{\vartheta}). \quad (1)$$

²We can accommodate genuinely non-answerable instances by augmenting the answer set with a null answer $a = \emptyset$.

The statistic $Z_n(a)$ is large when the data are much better explained by Θ_a^* than by any alternative answer. Let $L_n(\boldsymbol{\vartheta}) \triangleq \Gamma(\mathbf{N}_n; \widehat{\boldsymbol{\theta}}_n, \boldsymbol{\vartheta})$, then $Z_n(a) = \inf_{\boldsymbol{\vartheta} \in \text{Alt}_a} L_n(\boldsymbol{\vartheta}) - \inf_{\boldsymbol{\vartheta} \in \Theta_a^*} L_n(\boldsymbol{\vartheta})$.

Whenever $\widehat{\boldsymbol{\theta}}_n \in \Theta_a^*$, we have $\inf_{\boldsymbol{\vartheta} \in \Theta_a^*} L_n(\boldsymbol{\vartheta}) = 0$ and hence $Z_n(a) = \inf_{\boldsymbol{\vartheta} \in \text{Alt}_a} L_n(\boldsymbol{\vartheta})$. Thus, for a correct answer a and a consistent estimator that eventually enters Θ_a^* , the GLR reduces to the usual Chernoff-type quantity used in fixed-confidence pure exploration (Garivier and Kaufmann 2016). In this sense, the GLR is a robustified version of the usual Chernoff information to handle non-answerable estimates.

Recall that $Z_n(a)$ quantifies the evidence for a to be correct. The overall evidence of the data for *some* answer to be correct is then $\max_{a \in \mathcal{A}} Z_n(a)$. Accordingly, a GLR stopping rule is

$$\tau_\delta \triangleq \inf\{n \geq 1 : \max_{a \in \mathcal{A}} Z_n(a) \geq \beta_{n,\delta}\}, \quad (2)$$

where $\beta_{n,\delta}$ is a threshold typically in the order of $\mathcal{O}(\log(1/\delta))$; see Ding et al. (2025) for discussions. The corresponding decision rule is to nominate the answer best-supported by the data:

$$\hat{a}_{\tau_\delta} \in \operatorname{argmax}_{a \in \mathcal{A}} Z_{\tau_\delta}(a). \quad (3)$$

This decision rule differs from the usual empirical-best-answer rule. We choose it for essentially free finite-sample robustness, since the noisy estimates may be non-answerable and the empirical-best answer may then be undefined. Such transient non-answerability vanishes asymptotically under accurate estimation.

Remark 1 (On the threshold $\beta_{n,\delta}$ and δ -correctness). *One needs a threshold sequence $\beta_{n,\delta}$ such that $\mathbb{P}_\theta(\exists a \notin \mathcal{A}^*(\theta), \exists n \geq 1 : Z_n(a) \geq \beta_{n,\delta}) \leq \delta$ for every $\theta \in \Theta^*$. This is the usual generalized-likelihood-ratio logic behind fixed-confidence pure exploration. The exact threshold calibration is generally open and problem-dependent; see Garivier and Kaufmann (2016), Ramdas et al. (2020), and Ding et al. (2025) for related discussions from the machine learning, economics and simulation communities. For the purposes of this paper, we keep the exact threshold abstract and focus instead on efficient sampling rules. In the numerical experiments, we use the heuristic threshold $\beta_{n,\delta} = \log((\log n + 1)/\delta)$.*

2.3 Answer-wise pitfalls and the max-max-min characteristic value

To state the lower bound in a way that covers both non-answerable estimates and multiple-correct-answer problems, we propose an answer-wise generalization of pitfalls in Qin and You (2026). For each potential answer $a \in \mathcal{A}$, we assume that the rejection region Alt_a can be decomposed into finitely many simpler alternative sets. More precisely, for every candidate answer $a \in \mathcal{A}$, let

$$\text{Alt}_a = \bigcup_{x \in \mathcal{X}_a} \text{Alt}_{a,x},$$

where \mathcal{X}_a is a finite pitfall index set for answer a and each $\text{Alt}_{a,x} \subseteq \text{Alt}_a$ is a convex set of (answerable) alternative instances associated with pitfall x for answer a ; see Sections 3.1–3.3) for examples. A pitfall $x \in \mathcal{X}_a$ describes one concrete way in which answer a can fail. In ranking-and-selection, a

pitfall is a challenger alternative that beats the current nominee. In dueling bandits, a pitfall is an opponent that prevents the nominee from being a Condorcet winner.

In Qin and You (2026), pitfalls are indexed through the unique correct answer at the true instance. Here we instead index pitfalls by each candidate answer a . This answer-wise indexing is what lets the framework handle both multiple correct answers and temporary non-answerable estimates.

For lower-bound statements, we consider only answerable instances $\theta \in \Theta^*$. For any allocation $\mathbf{p} \in \Delta_{\mathcal{U}}$, true instance $\theta \in \Theta^*$, and alternative instance $\vartheta \in \Theta$, define the cost-normalized discrimination rate

$$\Gamma^c(\mathbf{p}; \theta, \vartheta) \triangleq \frac{\sum_{u \in \mathcal{U}} p_u \text{KL}(P_{\theta,u} \| P_{\vartheta,u})}{\bar{C}_{\theta}(\mathbf{p})}.$$

This is the statistical distance between θ and ϑ under allocation \mathbf{p} , normalized by the average cost per sample. For each answer-pitfall pair (a, x) , define the cost-normalized information rate for ruling out pitfall x when certifying answer a as

$$D_{a,x}(\mathbf{p}; \theta) \triangleq \inf_{\vartheta \in \text{Alt}_{a,x}} \Gamma^c(\mathbf{p}; \theta, \vartheta).$$

Let $\vartheta^{a,x}(\mathbf{p}, \theta) \in \text{argmin}_{\vartheta \in \overline{\text{Alt}}_{a,x}} \Gamma^c(\mathbf{p}; \theta, \vartheta)$ denote the most-confusing alternative instance.

The learner must rule out *all* pitfalls, so the bottleneck information rate is $\min_{x \in \mathcal{X}_a} D_{a,x}(\mathbf{p}; \theta)$. Among all potential answers in \mathcal{A} , the learner may nominate the one that is easiest to certify, namely the one with the largest bottleneck information rate. The learner then chooses the allocation \mathbf{p} to maximize that rate. This leads to the characteristic value

$$\Gamma^*(\theta) \triangleq \max_{\mathbf{p} \in \Delta_{\mathcal{U}}} \max_{a \in \mathcal{A}} \min_{x \in \mathcal{X}_a} D_{a,x}(\mathbf{p}; \theta).$$

For problems with a unique answer, the second maximization is vacuous, as it is always attained by that answer. Accordingly, the characteristic value reduces to the one studied in Qin and You (2026). Nevertheless, the second maximization remains important when temporary estimates fall outside the answerable set. In multiple-correct-answer problems (Degenne and Koolen 2019), it is essential even for answerable instances, because the answer set is set-valued.

Under any allocation rule that guarantees consistent estimation of the parameters, one expects $Z_n(a)$ to grow approximately at rate $B_n \cdot \min_{x \in \mathcal{X}_a} D_{a,x}(\mathbf{p}; \theta)$. Optimizing over \mathbf{p} and a suggests the heuristic scaling $\max_{a \in \mathcal{A}} Z_n(a) \approx B_n \cdot \Gamma^*(\theta)$. In view of the stopping rule (2), this leads to the heuristic lower bound:

$$\mathbb{E}_{\theta}[B_{\tau_{\delta}}] \gtrsim \frac{\log(1/\delta)}{\Gamma^*(\theta)}.$$

This is the direct analogue of the classical characteristic time in fixed-confidence pure exploration (Garivier and Kaufmann 2016), except that the normalization is by spent budget rather than by the number of samples. We refer to Qin and Russo (2024) for a rigorous discussion.

3 The ENDS algorithm

We now generalize Qin and You (2026) to accommodate multiple correct answers and non-answerable estimates. At round n , let $\theta_n \in \Theta$ be an estimator of θ . This estimator may be the sample mean, a posterior draw, or a projection onto the answerable set. During sampling, we nominate

$$a_n \in \arg \max_{a \in \mathcal{A}} Z(a; \mathbf{p}_n, \theta_n)$$

as the most plausible answer. (Here, we use a variant of GLR, namely $Z(a; \mathbf{p}_n, \theta_n)$, which uses the user-specified estimator θ_n instead of the ambient MLE $\hat{\theta}_n$. In our numerical experiments, we use posterior sampling as the default estimation routine for the sampling rule. Drawing θ_n from the current posterior naturally encourages exploration and mitigates the influence of rare, high-leverage early observations on nomination and pitfall detection.) Once the nominee a_n has been selected, the pitfall detector chooses

$$x_n \in \operatorname{argmin}_{x \in \mathcal{X}_{a_n}} D_{a_n, x}(\mathbf{p}_n; \theta_n).$$

The principal pitfall x_n is the most likely way for answer a_n to fail under the current estimate and allocation. The cost-aware information-directed selection (IDS) distribution is then

$$H_u^{a_n, x_n}(\mathbf{p}_n; \theta_n) \triangleq \frac{p_{n, u} \operatorname{KL}(P_{\theta_n, u} \| P_{\vartheta^{a_n, x_n, u}}) / C_u(\theta_n)}{\sum_{v \in \mathcal{U}} p_{n, v} \operatorname{KL}(P_{\theta_n, v} \| P_{\vartheta^{a_n, x_n, v}}) / C_v(\theta_n)}, \quad (4)$$

where $\vartheta^{a_n, x_n} = \vartheta^{a_n, x_n}(\mathbf{p}_n, \theta_n)$ is the most confusing alternative instance for that pitfall, namely

$$\vartheta^{a_n, x_n}(\mathbf{p}_n, \theta_n) \in \operatorname{argmin}_{\vartheta \in \overline{\operatorname{Alt}}_{a_n, x_n}} \Gamma^c(\mathbf{p}_n; \theta_n, \vartheta).$$

The IDS (4) defines a valid distribution over measurement types; see Qin and You (2026) for the case with unit-cost. The next measurement is drawn from this distribution. The stopping rule is given by (2) and final decision is the nominee in (3); both based on the ambient-MLE-based GLR (1). This yields a conceptually simple four-step scheme: Estimate–Nominate–Detect–Select. We refer to this algorithm as ENDS and summarize its fixed-precision variant in Algorithm 1. The sampling rule is anytime, so the algorithm can be turned into a fixed-budget algorithm by replacing the stopping rule with a fixed budget.

To instantiate this algorithm for a specific problem, one needs only to specify the explicit form of the alternative sets $\operatorname{Alt}_{a, x}$, the cost-normalized information rates $D_{a, x}(\mathbf{p}; \theta)$, and the IDS distribution $H^{a, x}(\mathbf{p}; \theta)$. We now provide these explicit forms for Examples 1–3.

3.1 Example: Selection of a good alternative

Consider the selection of a good alternative in Example 1. A natural pitfall for answer a is a challenger alternative $x \neq a$ whose mean exceeds that of a by more than ε . Accordingly, for $x \in [K] \setminus \{a\}$ define $\operatorname{Alt}_{a, x}^\varepsilon \triangleq \{\vartheta \in \Theta : \vartheta_x > \vartheta_a + \varepsilon\}$. Then $\operatorname{Alt}_a = \bigcup_{x \neq a} \operatorname{Alt}_{a, x}^\varepsilon$ and $\mathcal{X}_a = [K] \setminus \{a\}$. The characteristic value is

$$\Gamma_\varepsilon^*(\theta) = \max_{\mathbf{p} \in \Delta_K} \max_{a \in [K]} \min_{x \neq a} D_{a, x}^\varepsilon(\mathbf{p}; \theta), \quad \text{where} \quad D_{a, x}^\varepsilon(\mathbf{p}; \theta) \triangleq \inf_{\vartheta \in \operatorname{Alt}_{a, x}^\varepsilon} \Gamma^c(\mathbf{p}; \theta, \vartheta).$$

Algorithm 1 ENDS

Require: Confidence level δ , threshold sequence $\{\beta_{n,\delta}\}_{n \geq 1}$, estimation routine **est** (default: posterior sampling)

- 1: **Initialize:** sample each measurement type $u \in \mathcal{U}$ once
 - 2: **while** $\max_{a \in \mathcal{A}} Z_n(a) < \beta_{n,\delta}$ **do**
 - 3: $\boldsymbol{\theta}_n \leftarrow \mathbf{est}(\mathcal{H}_n)$ ▷ Estimate
 - 4: $a_n \in \arg \max_{a \in \mathcal{A}} Z(a; \mathbf{p}_n, \boldsymbol{\theta}_n)$ ▷ Nominate
 - 5: $x_n \in \arg \min_{x \in \mathcal{X}_{a_n}} D_{a_n,x}(\mathbf{p}_n; \boldsymbol{\theta}_n)$ ▷ Detect
 - 6: draw $U_{n+1} \sim \mathbf{H}^{a_n, x_n}(\mathbf{p}_n; \boldsymbol{\theta}_n)$, where $\mathbf{H}^{a_n, x_n} = (H_u^{a_n, x_n})_{u \in \mathcal{U}}$ is defined in (4) ▷ Select
 - 7: measure U_{n+1} , observe Y_{n+1} and update \mathcal{H}_{n+1}
 - 8: **end while**
 - 9: **return** $\hat{a}_n \in \arg \max_{a \in \mathcal{A}} Z_n(a)$
-

Assume that all alternatives belong to the same one-parameter exponential family with mean domain \mathbf{M} and KL divergence d . For $\mu \in \mathbf{M}$, define the upper-feasible interval $J_\varepsilon(\mu) \triangleq (-\infty, \mu + \varepsilon] \cap \mathbf{M}$. For $z \in \mathbf{M}$, let $\Pi_{J_\varepsilon(\mu)}(z)$ denote the projection of z onto $J_\varepsilon(\mu)$. For any weight vector $\mathbf{w} = (w_i)_{i \in [K]} \in \mathbb{R}_+^K$ and any mean vector $\mathbf{z} = (z_i)_{i \in [K]} \in \mathbf{M}^K$, define

$$G_{a,x}^\varepsilon(\mathbf{w}, \mathbf{z}) \triangleq \inf_{\substack{\mu, \nu \in \mathbf{M} \\ \nu \geq \mu + \varepsilon}} \left\{ w_a d(z_a, \mu) + w_x d(z_x, \nu) \right\}$$

and

$$A_a^\varepsilon(\mathbf{w}, \mathbf{z}) \triangleq \inf_{\mu \in \mathbf{M}} \left[w_a d(z_a, \mu) + \sum_{j \neq a} w_j d(z_j, \Pi_{J_\varepsilon(\mu)}(z_j)) \right].$$

The quantity $G_{a,x}^\varepsilon(\mathbf{w}, \mathbf{z})$ is the weighted KL distance from \mathbf{z} to the pitfall set $\text{Alt}_{a,x}^\varepsilon$, while $A_a^\varepsilon(\mathbf{w}, \mathbf{z})$ is the weighted KL distance from \mathbf{z} to the acceptance set Θ_a^* . Then, $D_{a,x}^\varepsilon(\mathbf{p}; \boldsymbol{\theta}) = G_{a,x}^\varepsilon(\mathbf{p}, \boldsymbol{\theta}) / \bar{C}_\theta(\mathbf{p})$. Let $(\mu_{a,x}^*, \nu_{a,x}^*) \in \arg \min_{\mu, \nu \in \mathbf{M}: \nu \geq \mu + \varepsilon} \{p_a d(\theta_a, \mu) + p_x d(\theta_x, \nu)\}$, and define $v_i^{a,x}(\mathbf{p}, \boldsymbol{\theta}) = \mu_{a,x}^*$ if $i = a$, $\nu_{a,x}^*$ if $i = x$, and θ_i otherwise.

Proposition 1. *The vector $\boldsymbol{\vartheta}^{a,x}(\mathbf{p}, \boldsymbol{\theta})$ attains the value $G_{a,x}^\varepsilon(\mathbf{p}, \boldsymbol{\theta})$. Plugging $\boldsymbol{\vartheta}^{a,x}$ into (4), we obtain the explicit IDS distribution $\mathbf{H}^{a,x}(\mathbf{p}; \boldsymbol{\theta})$. Specifically, $H_i^{a,x}(\mathbf{p}; \boldsymbol{\theta}) = 0$ for every $i \notin \{a, x\}$. Let $\mathbf{N}_n = (N_{n,i})_{i \in [K]}$ and let $\hat{\boldsymbol{\theta}}_n$ be the ambient MLE. The explicit form of $Z_n(a)$ is obtained by*

$$\inf_{\boldsymbol{\vartheta} \in \Theta_a^*} L_n(\boldsymbol{\vartheta}) = A_a^\varepsilon(\mathbf{N}_n, \hat{\boldsymbol{\theta}}_n), \quad \inf_{\boldsymbol{\vartheta} \in \text{Alt}_a} L_n(\boldsymbol{\vartheta}) = \min_{x \neq a} G_{a,x}^\varepsilon(\mathbf{N}_n, \hat{\boldsymbol{\theta}}_n).$$

Remark 2. *In the Gaussian unit-variance, unit-cost case, $d(u, v) = (u - v)^2/2$ and $\mathbf{M} = \mathbb{R}$. If $\theta_x < \theta_a + \varepsilon$, then $G_{a,x}^\varepsilon(\mathbf{p}, \boldsymbol{\theta}) = \frac{(\theta_a - \theta_x + \varepsilon)^2}{2(1/p_a + 1/p_x)}$. The optimizer is $\mu_{a,x}^* = \frac{p_a \theta_a + p_x (\theta_x - \varepsilon)}{p_a + p_x}$ and $\nu_{a,x}^* = \mu_{a,x}^* + \varepsilon$, and the IDS rule reduces to $H_a^{a,x}(\mathbf{p}; \boldsymbol{\theta}) = \frac{p_x}{p_a + p_x}$ and $H_x^{a,x}(\mathbf{p}; \boldsymbol{\theta}) = \frac{p_a}{p_a + p_x}$. The outer maximization is attained by $a = I^*$, and*

$$\Gamma_\varepsilon^*(\boldsymbol{\theta}) = \max_{\mathbf{p} \in \Delta_K} \min_{x \neq I^*} \frac{(\theta_{I^*} - \theta_x + \varepsilon)^2}{2(1/p_{I^*} + 1/p_x)}.$$

3.2 Example: Multi-fidelity ranking-and-selection

Consider the multi-fidelity ranking-and-selection problem described in Example 2. A natural pitfall for answer a is a challenger alternative $x \neq a$ whose top-fidelity mean strictly exceeds that of alternative a . Accordingly, for $x \in [K] \setminus \{a\}$, define $\text{Alt}_{a,x}^{\text{MF}} \triangleq \{\boldsymbol{\vartheta} \in \Theta_{\text{MF}} : \vartheta_{x,M} > \vartheta_{a,M}\}$. The characteristic value is

$$\Gamma_{\text{MF}}^*(\boldsymbol{\theta}) = \max_{\mathbf{p} \in \Delta_{\mathcal{U}}} \max_{a \in \mathcal{A}} \min_{x \neq a} D_{a,x}^{\text{MF}}(\mathbf{p}; \boldsymbol{\theta}), \quad \text{where} \quad D_{a,x}^{\text{MF}}(\mathbf{p}; \boldsymbol{\theta}) \triangleq \inf_{\boldsymbol{\vartheta} \in \text{Alt}_{a,x}^{\text{MF}}} \Gamma^c(\mathbf{p}; \boldsymbol{\theta}, \boldsymbol{\vartheta}). \quad (5)$$

At answerable instances with a unique best alternative, the maximum over a is vacuous because it is attained by that alternative. For non-answerable instances, the maximization is useful for nomination.

Assume that each fidelity level belongs to a one-parameter exponential family with mean domain \mathbf{M} and KL divergence d . For $\mu \in \mathbf{M}$, define the valid interval at fidelity m by $I_m(\mu) \triangleq [\mu - \xi_m, \mu + \xi_m] \cap \mathbf{M}$ for all $m \in [M]$. For $z \in \mathbf{M}$, let $\Pi_{I_m(\mu)}(z)$ denote the projection of z onto $I_m(\mu)$. For any weight $\mathbf{w} = (w_{i,m})_{i \in [K], m \in [M]}$ and any mean table $\mathbf{z} = (z_{i,m})_{i \in [K], m \in [M]} \in \mathbf{M}^{K \times M}$, define $\Psi_i(\mu; \mathbf{w}, \mathbf{z}) \triangleq \sum_{m=1}^M w_{i,m} d(z_{i,m}, \Pi_{I_m(\mu)}(z_{i,m}))$ for $i \in [K]$. This is the weighted KL distance from \mathbf{z} to the set of mean tables that satisfy the multi-fidelity constraints for alternative i with top-fidelity mean μ .

For $a \in [K]$ and $x \neq a$, define the weighted KL distance from \mathbf{z} to the pitfall set $\text{Alt}_{a,x}^{\text{MF}}$

$$\begin{aligned} G_{a,x}(\mathbf{w}, \mathbf{z}) &\triangleq \inf_{\boldsymbol{\vartheta} \in \text{Alt}_{a,x}^{\text{MF}}} \sum_{\substack{i \in [K] \\ m \in [M]}} w_{i,m} d(z_{i,m}, \vartheta_{i,m}) \\ &= \sum_{i \notin \{a,x\}} \inf_{\mu \in \mathbf{M}} \Psi_i(\mu; \mathbf{w}, \mathbf{z}) + \inf_{\substack{\mu_a, \mu_x \in \mathbf{M} \\ \mu_x \geq \mu_a}} \left\{ \Psi_a(\mu_a; \mathbf{w}, \mathbf{z}) + \Psi_x(\mu_x; \mathbf{w}, \mathbf{z}) \right\}. \end{aligned}$$

For each $j \neq a$, fix any minimizer $\ell_j \in \arg \min_{\nu \in \mathbf{M}} \Psi_j(\nu; \mathbf{w}, \mathbf{z})$, and define

$$A_a(\mathbf{w}, \mathbf{z}) \triangleq \inf_{\mu \in \mathbf{M}} \left[\Psi_a(\mu; \mathbf{w}, \mathbf{z}) + \sum_{j \neq a} \Psi_j(\mu \wedge \ell_j; \mathbf{w}, \mathbf{z}) \right].$$

Since each $\Psi_j(\cdot; \mathbf{w}, \mathbf{z})$ is convex, $\inf_{\nu \leq \mu} \Psi_j(\nu; \mathbf{w}, \mathbf{z}) = \Psi_j(\mu \wedge \ell_j; \mathbf{w}, \mathbf{z})$ for all $j \neq a$, so $A_a(\mathbf{w}, \mathbf{z})$ is exactly the weighted KL distance from \mathbf{z} to the acceptance set Θ_a^* , written as a one-dimensional convex optimization. Finally, $D_{a,x}^{\text{MF}}(\mathbf{p}; \boldsymbol{\theta}) = G_{a,x}(\mathbf{p}, \boldsymbol{\theta}) / \bar{C}(\mathbf{p})$ with $\bar{C}(\mathbf{p}) \triangleq \sum_{i=1}^K \sum_{m=1}^M p_{i,m} C_m$.

Let $\mu_i^* \in \arg \min_{\mu \in \mathbf{M}} \Psi_i(\mu; \mathbf{p}, \boldsymbol{\theta})$ for each $i \notin \{a, x\}$, let

$$(\mu_a^*, \mu_x^*) \in \arg \min_{\substack{\mu_a, \mu_x \in \mathbf{M} \\ \mu_x \geq \mu_a}} \left\{ \Psi_a(\mu_a; \mathbf{p}, \boldsymbol{\theta}) + \Psi_x(\mu_x; \mathbf{p}, \boldsymbol{\theta}) \right\},$$

fix $\ell_j^* \in \arg \min_{\nu \in \mathbf{M}} \Psi_j(\nu; \mathbf{p}, \boldsymbol{\theta})$ for each $j \neq a$, and let

$$\lambda_a^* \in \arg \min_{\mu \in \mathbf{M}} \left[\Psi_a(\mu; \mathbf{p}, \boldsymbol{\theta}) + \sum_{j \neq a} \Psi_j(\mu \wedge \ell_j^*; \mathbf{p}, \boldsymbol{\theta}) \right].$$

Proposition 2. The table $\vartheta^{a,x}(\mathbf{p}, \boldsymbol{\theta})$ defined by $\vartheta_{i,m}^{a,x}(\mathbf{p}, \boldsymbol{\theta}) \triangleq \Pi_{I_m(\mu_i^*)}(\theta_{i,m})$ for $i \in [K]$, $m \in [M]$ attains the value $G_{a,x}(\mathbf{p}, \boldsymbol{\theta})$. Plugging $\vartheta^{a,x}$ into (4), we obtain the explicit IDS distribution $\mathbf{H}^{a,x}(\mathbf{p}; \boldsymbol{\theta})$. The table $\vartheta^{a,*}(\mathbf{p}, \boldsymbol{\theta})$ defined below belongs to Θ_a^* and attains the value $A_a(\mathbf{p}, \boldsymbol{\theta})$:

$$\vartheta_{i,m}^{a,*}(\mathbf{p}, \boldsymbol{\theta}) \triangleq \begin{cases} \Pi_{I_m(\lambda_a^*)}(\theta_{a,m}), & i = a, \\ \Pi_{I_m(\lambda_a^* \wedge \ell_i^*)}(\theta_{i,m}), & i \neq a, \end{cases} \quad i \in [K], m \in [M].$$

Let $\mathbf{N}_n = (N_{n,i,m})_{i,m}$ and let $\hat{\boldsymbol{\theta}}_n$ be the ambient MLE. The explicit form of $Z_n(a)$ is obtained by

$$\inf_{\boldsymbol{\vartheta} \in \Theta_a^*} L_n(\boldsymbol{\vartheta}) = A_a(\mathbf{N}_n, \hat{\boldsymbol{\theta}}_n), \quad \inf_{\boldsymbol{\vartheta} \in \text{Alt}_a} L_n(\boldsymbol{\vartheta}) = \min_{x \neq a} G_{a,x}(\mathbf{N}_n, \hat{\boldsymbol{\theta}}_n).$$

Remark 3 (Gaussian special case). In the Gaussian unit-variance case, $\mathbf{M} = \mathbb{R}$ and $d(u, v) = (u - v)^2/2$, so

$$\Psi_i(\mu; \mathbf{w}, \mathbf{z}) = \frac{1}{2} \sum_{m=1}^M w_{i,m} (|z_{i,m} - \mu| - \xi_m)_+^2, \quad i \in [K].$$

Therefore, for any choice of minimizers $\ell_j \in \arg \min_{\nu \in \mathbb{R}} \Psi_j(\nu; \mathbf{w}, \mathbf{z})$,

$$A_a(\mathbf{w}, \mathbf{z}) = \inf_{\mu \in \mathbb{R}} \left[\frac{1}{2} \sum_{m=1}^M w_{a,m} (|z_{a,m} - \mu| - \xi_m)_+^2 + \sum_{j \neq a} \frac{1}{2} \sum_{m=1}^M w_{j,m} (|z_{j,m} - (\mu \wedge \ell_j)| - \xi_m)_+^2 \right].$$

Remark 4. If $\hat{\boldsymbol{\theta}}_n \in \Theta_a^*$, then $A_a(\mathbf{N}_n, \hat{\boldsymbol{\theta}}_n) = 0$ and, for every $x \neq a$,

$$G_{a,x}(\mathbf{N}_n, \hat{\boldsymbol{\theta}}_n) = \inf_{\mu \in \mathbf{M}} \{ \Psi_a(\mu; \mathbf{N}_n, \hat{\boldsymbol{\theta}}_n) + \Psi_x(\mu; \mathbf{N}_n, \hat{\boldsymbol{\theta}}_n) \}.$$

Hence $Z_n(a) = \min_{x \neq a} \inf_{\mu \in \mathbf{M}} \{ \Psi_a(\mu; \mathbf{N}_n, \hat{\boldsymbol{\theta}}_n) + \Psi_x(\mu; \mathbf{N}_n, \hat{\boldsymbol{\theta}}_n) \}$.

3.3 Example: Dueling bandits with a Condorcet winner

Consider the dueling bandit problem described in Example 3. A natural pitfall for answer a is a challenger alternative $x \neq a$ that itself becomes a Condorcet winner. Accordingly, for $x \in [K] \setminus \{a\}$, define $\text{Alt}_{a,x}^{\text{CW}} \triangleq \{ \boldsymbol{\vartheta} \in \Theta_{\text{DB}} : \vartheta_{x,j} > 1/2, \forall j \neq x \}$. Since x beating every opponent already implies uniqueness of the Condorcet winner, we have $\text{Alt}_{a,x}^{\text{CW}} = \Theta_x^*$. Hence $\text{Alt}_a = \bigcup_{x \neq a} \text{Alt}_{a,x}^{\text{CW}}$ with $\mathcal{X}_a = [K] \setminus \{a\}$. The characteristic value is

$$\Gamma_{\text{CW}}^*(\boldsymbol{\theta}) = \max_{\mathbf{p} \in \Delta_{\mathcal{U}}} \max_{a \in [K]} \min_{x \neq a} D_{a,x}^{\text{CW}}(\mathbf{p}; \boldsymbol{\theta}), \quad \text{where} \quad D_{a,x}^{\text{CW}}(\mathbf{p}; \boldsymbol{\theta}) \triangleq \inf_{\boldsymbol{\vartheta} \in \text{Alt}_{a,x}^{\text{CW}}} \Gamma^c(\mathbf{p}; \boldsymbol{\theta}, \boldsymbol{\vartheta}).$$

Let d denote the Bernoulli KL divergence. For any weight vector $\mathbf{w} = (w_{\{i,j\}})_{\{i,j\} \in \mathcal{U}} \in \mathbb{R}_+^{\mathcal{U}}$, extend it symmetrically by setting $w_{i,j} \triangleq w_{\{i,j\}} = w_{j,i}$ for $i \neq j$. Define $I_{\text{CW}} \triangleq [1/2, 1]$ and projection $\Pi_{\text{CW}}(q) \triangleq \Pi_{I_{\text{CW}}}(q) = q \vee 1/2$ for $q \in [0, 1]$. For any preference matrix $\mathbf{z} \in \Theta_{\text{DB}}$ and each alternative $i \in [K]$, define

$$\Psi_i(\mathbf{w}, \mathbf{z}) \triangleq \sum_{j \neq i} w_{i,j} d(z_{i,j}, \Pi_{\text{CW}}(z_{i,j})) = \sum_{j: z_{i,j} \leq 1/2} w_{i,j} d(z_{i,j}, 1/2).$$

Consequently, $G_{a,x}^{\text{CW}}(\mathbf{w}, \mathbf{z}) = \Psi_x(\mathbf{w}, \mathbf{z})$, $A_a^{\text{CW}}(\mathbf{w}, \mathbf{z}) = \Psi_a(\mathbf{w}, \mathbf{z})$, and

$$D_{a,x}^{\text{CW}}(\mathbf{p}; \boldsymbol{\theta}) = \frac{\Psi_x(\mathbf{p}, \boldsymbol{\theta})}{\bar{C}_{\boldsymbol{\theta}}(\mathbf{p})} = \frac{\sum_{j \neq x} p_{\{x,j\}} d(\theta_{x,j}, \theta_{x,j} \vee 1/2)}{\bar{C}_{\boldsymbol{\theta}}(\mathbf{p})} = \frac{\sum_{j: \theta_{x,j} \leq 1/2} p_{\{x,j\}} d(\theta_{x,j}, 1/2)}{\bar{C}_{\boldsymbol{\theta}}(\mathbf{p})}.$$

The most confusing alternative for pitfall x is therefore obtained by leaving every duel not involving x unchanged and pushing each losing duel of x exactly to the boundary $1/2$. For $x \neq a$, define the boundary representative $\boldsymbol{\vartheta}^{a,x}(\mathbf{p}, \boldsymbol{\theta}) \in \Theta_{\text{DB}}$ by $\vartheta_{x,j}^{a,x}(\mathbf{p}, \boldsymbol{\theta}) \triangleq \theta_{x,j} \vee \frac{1}{2}$, $\vartheta_{j,x}^{a,x}(\mathbf{p}, \boldsymbol{\theta}) \triangleq 1 - \vartheta_{x,j}^{a,x}(\mathbf{p}, \boldsymbol{\theta})$, $j \neq x$, and $\vartheta_{i,j}^{a,x}(\mathbf{p}, \boldsymbol{\theta}) \triangleq \theta_{i,j}$ for all $i, j \neq x$. Similarly, define $\boldsymbol{\vartheta}^{a,*}(\mathbf{p}, \boldsymbol{\theta})$ by replacing x with a in the display above.

Proposition 3. *The matrix $\boldsymbol{\vartheta}^{a,x}(\mathbf{p}, \boldsymbol{\theta})$ attains the same objective value as the infimum defining $G_{a,x}^{\text{CW}}(\mathbf{p}, \boldsymbol{\theta})$, and $\boldsymbol{\vartheta}^{a,*}(\mathbf{p}, \boldsymbol{\theta})$ attains the same objective value as the infimum defining $A_a^{\text{CW}}(\mathbf{p}, \boldsymbol{\theta})$. Plugging $\boldsymbol{\vartheta}^{a,x}$ into (4), we obtain the explicit IDS distribution $\mathbf{H}^{a,x}(\mathbf{p}; \boldsymbol{\theta})$. Let $\mathbf{N}_n = (N_{n,\{i,j\}})_{\{i,j\} \in \mathcal{U}}$ and let $\hat{\boldsymbol{\theta}}_n$ be the ambient MLE. The explicit form of $Z_n(a)$ is obtained by*

$$\inf_{\boldsymbol{\vartheta} \in \Theta_a^*} L_n(\boldsymbol{\vartheta}) = A_a^{\text{CW}}(\mathbf{N}_n, \hat{\boldsymbol{\theta}}_n) = \Psi_a(\mathbf{N}_n, \hat{\boldsymbol{\theta}}_n)$$

and

$$\inf_{\boldsymbol{\vartheta} \in \text{Alt}_a} L_n(\boldsymbol{\vartheta}) = \min_{x \neq a} G_{a,x}^{\text{CW}}(\mathbf{N}_n, \hat{\boldsymbol{\theta}}_n) = \min_{x \neq a} \Psi_x(\mathbf{N}_n, \hat{\boldsymbol{\theta}}_n).$$

Remark 5. *If $\hat{\boldsymbol{\theta}}_n \in \Theta_a^*$, then $\Psi_a(\mathbf{N}_n, \hat{\boldsymbol{\theta}}_n) = 0$ and*

$$Z_n(a) = \min_{x \neq a} \sum_{j: \hat{\theta}_{n,x,j} \leq 1/2} N_{n,\{x,j\}} d(\hat{\theta}_{n,x,j}, 1/2).$$

If $\boldsymbol{\theta} \in \Theta_{\text{CW}}^$ has Condorcet winner a^* , then the outer maximization in $\Gamma_{\text{CW}}^*(\boldsymbol{\theta})$ is attained by a^* . In the unit-cost setting,*

$$\Gamma_{\text{CW}}^*(\boldsymbol{\theta}) = \max_{\mathbf{p} \in \Delta_{\mathcal{U}}} \min_{x \neq a^*} \sum_{j: \theta_{x,j} \leq 1/2} p_{\{x,j\}} d(\theta_{x,j}, 1/2).$$

4 Numerical experiments

In this section, we evaluate the proposed algorithm ENDS against state-of-the-art methods for selection of a good alternative, multi-fidelity R&S, and dueling bandits. Our numerical results show that ENDS (with posterior sampling as the default estimation routine), as a unified algorithm, is highly robust and competitive. We refer readers to Qin and You (2026) for more experiments on conventional R&S, whose algorithm is a special case of ENDS with unique answers.

4.1 Selection of a good alternative

For Example 1, we evaluate ENDS on 100 randomly generated Gaussian instances following Zhong and Hong (2018). In each instance, the means of the K arms are drawn i.i.d. from $\mathcal{N}(0, 16\epsilon^2)$ with $\epsilon = 0.1$, and the observation variances are all equal to 1 but treated as unknown. As baselines, we implement the unknown-variance version of Algorithm 2 of Kim and Eckman (2024) (coupled with

our GLR stopping), KN procedure (Kim and Nelson 2001), Paulson’s procedure (Paulson 1964), Procedure 2 of Zhong and Hong (2018), and uniform allocation. For ENDS, we plug the empirical sample variances into the formulas in Section 3.1. A fully rigorous treatment of unknown variance would require extending the GLR to two-parameter exponential-family models; see Section 4.1 of Qin and You (2026). ENDS and the uniform-allocation baseline use the stopping rule in (2) with the heuristic threshold $\beta_{n,\delta} = \log((\log n + 1)/\delta)$.

Table 1: Comparison of expected stopping budget under confidence level $\delta = 0.05$ and unknown observation variance $\sigma_i^2 = 1$. PCS omitted, since all algorithms achieve a perfect PCS of 1.

K	# good systems	ENDS	KE-Alg2	KN	Paulson	ZH-Proc2	Uniform
		$\mathbb{E}[B_{\tau_\delta}] \pm \text{SE}$	$\mathbb{E}[B_{\tau_\delta}] \pm \text{SE}$	$\mathbb{E}[B_{\tau_\delta}] \pm \text{SE}$	$\mathbb{E}[B_{\tau_\delta}] \pm \text{SE}$	$\mathbb{E}[B_{\tau_\delta}] \pm \text{SE}$	$\mathbb{E}[B_{\tau_\delta}] \pm \text{SE}$
10	1.43	1035 \pm 122	1723 \pm 305	2545 \pm 159	3236 \pm 199	6243 \pm 563	3737 \pm 414
100	1.86	4563 \pm 279	9148 \pm 1702	21453 \pm 482	27707 \pm 701	42028 \pm 1772	60964 \pm 4536
1000	2.28	19917 \pm 892	48412 \pm 8630	186094 \pm 1944	265476 \pm 4276	272350 \pm 4978	702645 \pm 54655

The results are reported in Table 1. All methods achieve essentially perfect PGS on this benchmark, while ENDS consistently requires the smallest stopping budget across all values of K . The advantage becomes more pronounced as the number of alternatives increases, suggesting that ENDS scales favorably in this unknown-variance setting. We note, however, that ENDS is not designed for large-scale problems, where the computational burden can become substantial.

4.2 Multi-fidelity R&S

We consider four synthetic instances. In each instance, the configuration is specified by the fidelity-wise mean vectors $\boldsymbol{\theta}_{\cdot,m}$, the bias vector $\boldsymbol{\xi}$, and the sampling cost vector \mathbf{C} . The top fidelity determines the true best arm. Instance 1 has $(K, M) = (5, 2)$, $\boldsymbol{\theta}_{\cdot,1} = (0.4, 0.4, 0.4, 0.4, 0.5)$, and $\boldsymbol{\theta}_{\cdot,2} = (0.5, 0.5, 0.5, 0.5, 0.6)$, with $\boldsymbol{\xi} = (0.1, 0)$ and $\mathbf{C} = (0.5, 5)$. In this instance, the optimal allocation balances sampling across the two fidelities. Instance 2 has $(K, M) = (4, 2)$, $\boldsymbol{\theta}_{\cdot,1} = (0.9, 0.4, 0.4, 0.3)$, and $\boldsymbol{\theta}_{\cdot,2} = (0.5, 0.6, 0.7, 0.85)$, with $\boldsymbol{\xi} = (0.55, 0)$ and $\mathbf{C} = (0.1, 5)$. Here the low fidelity acts as a misleading trap, and the optimal allocation samples exclusively from the high fidelity. Instance 3 has $(K, M) = (4, 2)$, $\boldsymbol{\theta}_{\cdot,1} = (0.15, 0.25, 0.55, 0.85)$, and $\boldsymbol{\theta}_{\cdot,2} = (0.1, 0.2, 0.6, 0.8)$, with $\boldsymbol{\xi} = (0.1, 0)$ and $\mathbf{C} = (0.5, 5)$. In this case, the optimal allocation samples exclusively from the low fidelity. We also include Instance 4, randomly generated with $(K, M) = (5, 4)$, following the setup in Table 2 of Poiani et al. (2024). For all instances, we fix target confidence level $\delta = 0.05$ and Gaussian observation with known variance $\sigma^2 = 1$.

We compare ENDS with MF-GRAD (Poiani et al. 2024), TS-KKT-IDS (Qin and You 2026) applied only at the highest fidelity, GRAD (Ménard 2019) applied only at the highest fidelity, and two uniform allocation schemes: uniform sampling and uniform cost allocation. To ensure a fair comparison, all algorithms use the same stopping rule in (2) with a heuristic stopping threshold $\beta_{n,\delta} = \log((\log n + 1)/\delta)$.

Table 2: Comparison of expected stopping cost (\pm Monte Carlo standard error) under confidence level $\delta = 0.05$ and observation variance $\sigma^2 = 1$. PCS omitted, since all algorithms achieve a PCS of > 0.998 .

Inst.	ENDS	MF-GRAD	TS-KKT-IDS	GRAD	Uniform Sample	Uniform Cost
1	32237 \pm 531	65611 \pm 1088	53841 \pm 712	77602 \pm 1086	56952 \pm 814	67366 \pm 1012
2	11504 \pm 199	38790 \pm 671	10794 \pm 203	17215 \pm 352	18883 \pm 347	38980 \pm 656
3	3550 \pm 84	4231 \pm 103	5085 \pm 101	8938 \pm 196	6844 \pm 153	4539 \pm 103
4	3613 \pm 65	3953 \pm 78	21815 \pm 330	28922 \pm 455	11765 \pm 184	3734 \pm 69

We make several observations from these experiments. First, ENDS achieves the lowest expected stopping cost in almost all instances while maintaining PCS equal to 1. The only exception is Instance 2, where it is slightly outperformed by TS-KKT-IDS. In Instance 2, the optimal allocation places all mass on the highest fidelity, so restricting attention to that fidelity is advantageous; indeed, ENDS restricted to the highest fidelity coincides exactly with TS-KKT-IDS. Second, although MF-GRAD is based on the same optimal allocation problem in (5), it treats sampling allocation as a continuous optimization problem followed by a tracking step. By contrast, ENDS acts directly as a bandit sampling rule, which appears markedly more efficient in our experiments. Third, Instance 2 shows that ENDS is robust to misleading low-fidelity information. When the low fidelity distorts the arm ranking, MF-GRAD incurs much larger cost because it converges slowly to the optimal allocation that excludes the low fidelity. ENDS instead performs nearly as well as TS-KKT-IDS, which effectively benefits from oracle knowledge that the low fidelity should be ignored. Fourth, in Instance 4, most of the discriminative information lies in the lower fidelities. This explains the poor performance of TS-KKT-IDS and GRAD, which use only the highest fidelity, and the competitive performance of uniform cost allocation, which assigns very little mass to the highest fidelity. Finally, all algorithms achieve nearly perfect PCS, suggesting that the heuristic stopping rules are conservative. Sharper threshold calibration is left for future work, see also Ding et al. (2025).

4.3 Dueling bandits

We compare ENDS with state-of-the-art dueling bandit algorithms on random instances drawn from the family considered by Haddenhorst et al. (2021). As baselines, we implement SAVAGE (Urvoy et al. 2013), DKWT (Haddenhorst et al. 2021), SEEBS (Ren et al. 2020), Explore-then-Verify (EtV, Karnin 2016), and a uniform-allocation baseline. For each algorithm, we averaged over 1000 independently sampled instances, sampled uniformly from the set of K -armed preference matrices such that: (i) there is a unique Condorcet winner; (ii) every off-diagonal entry is at least $h/2$ away from $1/2$. The parameter h controls the hardness of the instance, with smaller values corresponding to more difficult problems. ENDS and Uniform-allocation use the GLR stopping (2), while all other algorithms use their own stopping rules.

Table 3 reports the expected stopping cost $\mathbb{E}[B_{\tau_\delta}]$ (\pm Monte Carlo standard error). Across all settings, ENDS is the most cost-efficient method. The baselines with their native stopping rules

Table 3: Comparison of the expected stopping budget $\mathbb{E}[B_{\tau_\delta}]$ (\pm Monte Carlo standard error) under different configurations of (m, h) and $\delta = 0.05$. PCS omitted, since all algorithms achieve a perfect PCS of 1.

(m, h)	ENDS	SAVAGE	Uniform	DKWT	SEEBS	EtV
(5, 0.20)	173 \pm 4	953 \pm 15	501 \pm 20	6506 \pm 103	7535 \pm 134	8751 \pm 176
(5, 0.15)	206 \pm 5	1246 \pm 24	670 \pm 30	8293 \pm 142	11779 \pm 239	11183 \pm 280
(5, 0.10)	303 \pm 10	1825 \pm 47	1058 \pm 57	14198 \pm 396	19404 \pm 472	20464 \pm 1803 ^{†2}
(5, 0.05)	479 \pm 27	3580 \pm 146	2081 \pm 158	31291 \pm 1255	37198 \pm 1133	50368 \pm 4220 ^{†14}
(10, 0.20)	295 \pm 4	2805 \pm 22	809 \pm 32	14509 \pm 155	17520 \pm 194	26281 \pm 317
(10, 0.15)	336 \pm 5	3276 \pm 33	1065 \pm 57	18475 \pm 200	27447 \pm 347	33373 \pm 730
(10, 0.10)	409 \pm 7	4321 \pm 63	1471 \pm 90	32826 \pm 596	45535 \pm 689	58731 \pm 3330 ^{†7}
(10, 0.05)	494 \pm 11	6888 \pm 177	2701 \pm 446	74341 \pm 1910	81790 \pm 1567	154396 \pm 8501 ^{†76}

^{†x} indicates that the algorithm was forcibly terminated at 10^6 steps on x out of the 1000 instances.

incur substantially larger costs. This comparison also suggests that the GLR-based stopping rule, which is decoupled from the sampling rule, can be considerably more efficient than elimination-type stopping rules designed jointly with their sampling rules. Thus, the observed advantage should not be attributed solely to the ENDS sampling rule; it is partly intertwined with the use of a stronger stopping rule. The same caveat applies to the comparisons with KN, Paulson, and ZH-Proc2 (but not KE-Alg2) in Section 4.1. In contrast, all algorithms in Section 4.2 use the same stopping rule.

5 Conclusion

We studied cost-aware fixed-precision ranking-and-selection problems in which the correct answer may be non-unique and intermediate estimates may be non-answerable. To handle this setting, we proposed a unified framework based on answer-wise acceptance sets, restricted generalized likelihood ratio stopping, and an answer-pitfall decomposition, which leads to a common max-max-min characteristic value and the ENDS procedure. We derived explicit instantiations for good-alternative selection, multi-fidelity ranking-and-selection, and Condorcet-winner dueling bandits, and numerical experiments suggest that this unified recipe is robust and competitive across these qualitatively different problems.

Several directions remain open. A natural next step is to establish stronger theoretical guarantees for ENDS, such as asymptotic, instance-dependent upper bounds that match the characteristic value. A second is to develop a sharper understanding of the stopping rule, including less conservative threshold calibration and nonasymptotic δ -correctness guarantees. It would also be interesting to extend the framework to broader structured settings with richer feedback models, such as contextual problems.

Acknowledgments

Wei You’s research is generously supported by the Hong Kong Research Grants Council [Grant GRF 16212823] and [Theme-based Research Project T32-615/24-R].

References

- Wei Chen, Yihan Du, Longbo Huang, and Haoyu Zhao. Combinatorial pure exploration for dueling bandit. In *International Conference on Machine Learning*, pages 1531–1541. PMLR, 2020.
- Rémy Degenne and Wouter M Koolen. Pure exploration with multiple correct answers. *Advances in Neural Information Processing Systems*, 32:14591–14600, 2019.
- Mingrui Ding, Siyang Gao, and Qihong Zhao. Stopping rules for sampling in precision medicine. In *2025 Winter Simulation Conference (WSC)*, pages 929–940. IEEE, 2025. doi: 10.1109/WSC68292.2025.11338934.
- David J Eckman and Shane G Henderson. Guarantees on the probability of good selection. In *2018 Winter Simulation Conference (WSC)*, pages 351–365. IEEE, 2018. doi: 10.1109/WSC.2018.8632345.
- Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 998–1027. PMLR, 2016.
- Björn Haddendorst, Viktor Bengs, and Eyke Hüllermeier. Identification of the generalized condorcet winner in multi-dueling bandits. *Advances in Neural Information Processing Systems*, 34:25904–25916, 2021.
- L. Jeff Hong, Weiwei Fan, and Jun Luo. Review on ranking and selection: A new perspective. *Frontiers of Engineering Management*, 8(3):321–343, 2021. doi: 10.1007/s42524-021-0152-6.
- Marc Jourdan, Rémy Degenne, and Emilie Kaufmann. An ε -best-arm identification algorithm for fixed-confidence and beyond. *Advances in Neural Information Processing Systems*, 36:16578–16649, 2023. doi: 10.52202/075280-0727.
- Zohar S Karnin. Verification based solution for structured MAB problems. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper_files/paper/2016/file/65b9eea6e1cc6bb9f0cd2a47751a186f-Paper.pdf.
- Maximilian Kasy and Anja Sautmann. Adaptive treatment assignment in experiments for policy choice. *Econometrica*, 89(1):113–132, 2021. doi: 10.3982/ECTA17527.
- Seong-Hee Kim and Barry L Nelson. A fully sequential procedure for indifference-zone selection in simulation. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 11(3):251–273, 2001. doi: 10.1145/502109.502111.
- Taeho Kim and David J Eckman. Rate-optimal budget allocation for the probability of good selection. In *2024 Winter Simulation Conference (WSC)*, pages 3324–3335. IEEE, 2024.
- Pierre Ménard. Gradient ascent for active exploration in bandit problems. *arXiv preprint arXiv:1905.08165*, 2019.
- Edward Paulson. A sequential procedure for selecting the population with the largest mean from k normal populations. *The Annals of Mathematical Statistics*, pages 174–180, 1964.

- Riccardo Poiani, Rémy Degenne, Emilie Kaufmann, Alberto Maria Metelli, and Marcello Restelli. Optimal multi-fidelity best-arm identification. *Advances in Neural Information Processing Systems*, 37:121882–121927, 2024. doi: 10.52202/079017-3874.
- Chao Qin and Daniel Russo. Optimizing adaptive experiments: A unified approach to regret minimization and best-arm identification. *arXiv preprint arXiv:2402.10592*, 2024.
- Chao Qin and Wei You. Dual-directed algorithm design for efficient pure exploration. *Operations Research*, 74(2):573–1152, 2026. doi: 10.1287/opre.2023.0590.
- Aaditya Ramdas, Johannes Ruf, Martin Larsson, and Wouter Koolen. Admissible anytime-valid sequential inference must rely on nonnegative martingales. *arXiv preprint arXiv:2009.03167*, 2020.
- Wenbo Ren, Jia Liu, and Ness Shroff. The sample complexity of best- k items selection from pairwise comparisons. In *International Conference on Machine Learning*, pages 8051–8072. PMLR, 2020.
- Tanguy Urvoy, Fabrice Clerot, Raphael Féraud, and Sami Naamane. Generic exploration and k-armed voting bandits. In *International conference on machine learning*, pages 91–99. PMLR, 2013. doi: 10.5555/3042817.3042904.
- Wenyu Wang, Hong Wan, and Xi Chen. Bonferroni-free and indifference-zone-flexible sequential elimination procedures for ranking and selection. *Operations Research*, 72(5):2119–2134, 2024.
- Jie Xu, Si Zhang, Edward Huang, Chun-Hung Chen, Loo Hay Lee, and Nurcin Celik. Mo²tos: Multi-fidelity optimization with ordinal transformation and optimal sampling. *Asia-Pacific Journal of Operational Research*, 33(3):1650017, 2016. doi: 10.1142/S0217595916500172.
- Ying Zhong and L Jeff Hong. Fully sequential ranking and selection procedures with pac guarantee. In *2018 Winter Simulation Conference (WSC)*, pages 1898–1908. IEEE, 2018. doi: 10.1109/WSC.2018.8632439.