The Sampling Complexity of Condorcet Winner Identification in Dueling Bandits

El Mehdi Saad*
UM6P College of Computing
Rabat, Morocco
elmehdi.saad@um6p.ma

Victor Thuot*
INRAE, MISTEA, Institut Agro, Univ. Montpellier
Montpellier, France
victor.thuot@inrae.fr

Nicolas Verzelen
INRAE, MISTEA, Institut Agro, Univ. Montpellier
Montpellier, France
nicolas.verzelen@inrae.fr

Abstract

We study best-arm identification in large-scale stochastic dueling bandits under the sole assumption that a Condorcet winner exists, i.e., an arm that wins each noisy pairwise comparison with probability at least $1/2$. We introduce a new identification procedure that exploits the full gap matrix $\Delta_{i,j} = q_{i,j} - \frac{1}{2}$ (where $q_{i,j}$ is the probability that arm i beats arm j), rather than only the gaps between the Condorcet winner and the other arms. We derive high-probability, instance-dependent sample-complexity guarantees that (up to logarithmic factors) improve the best known ones by leveraging informative comparisons beyond those involving the winner. We complement these results with matching lower bounds that establish the optimality of our procedures in all regimes. Overall, our results reveal the general form of the sampling complexity, characterized by a trade-off between the cost of locating informative entries and the verification cost required to achieve the desired confidence. In particular, this complexity drastically differs from what is suggested by pure asymptotic results or by procedures that are tailored to Strongly Stochastic Transitive models.

1 Motivation and High-Level Overview

In many modern machine learning applications, obtaining trustworthy absolute feedback can be difficult, expensive, or systematically biased. By contrast, relative judgments are often easier to elicit and can be highly informative. This is especially apparent in information retrieval and recommendation systems, where users naturally compare two alternatives such as rankings or models rather than providing calibrated relevance scores [Joachims et al., 2007, Hofmann et al., 2016].

The dueling bandits framework formalizes this paradigm by allowing a learner to adaptively query pairs of arms and observe only a noisy binary outcome indicating which arm is preferred. At each round, the learner selects a pair of two arms $(i, j) \in [K] \times [K]$ and observes the outcome of their duel: the feedback is 1 if arm i is preferred to arm j, and 0 otherwise. This observation is modeled as a Bernoulli random variable with unknown parameter $q_{i,j} \in [0, 1]$. The collection of pairwise preference probabilities is represented by the matrix $\boldsymbol{Q} = (q_{i,j})_{i,j\in[K]}$. Since self-comparisons are uninformative and preferences are anti-symmetric, namely $q_{i,j} = 1 - q_{j,i}$ for all $i, j \in [K]$, the unknown matrix $\boldsymbol{Q}$ satisfies a skew-symmetry condition. Equivalently, we define the *gap matrix* $\boldsymbol{\Delta} = (\Delta_{i,j})_{i,j\in[K]}$ by $\Delta_{i,j} \coloneqq q_{i,j} - 1/2$, which satisfies $\Delta_{i,j} = -\Delta_{j,i}$ and $\Delta_{i,i} = 0$.

*Equal contribution

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

Stochastic dueling bandits have been studied extensively under a variety of structural assumptions on $\boldsymbol{Q}$ and with multiple notions of optimality [Bengs et al., 2021, Komiyama et al., 2015, Falahatgar et al., 2017, 2018, Ren et al., 2020, Jamieson et al., 2015, Zoghi et al., 2015a, Haddenhorst et al., 2021]. Unlike in the classical multi-armed bandit setting, defining an "optimal arm" is not immediate, which has led to several competing winner definitions; see the survey of Bengs et al. [2021]. In this work, we focus on instances where a distinguished arm $i^* \in [K]$ defeats, in expectation, every other arm, i.e., $q_{i^*,j} > 1/2$ for all $j \in [K] \setminus \{i^*\}$. Such an arm is called a *Condorcet winner* (CW) and is unique; we also refer to it as the optimal arm. Most existing work on dueling bandits assumes the existence of a CW [Zoghi et al., 2014, 2015b, Li et al., 2020, Komiyama et al., 2015, Chen and Frazier, 2017, Saha and Gaillard, 2022, Saha and Gupta, 2022], or even imposes the stronger requirement that the arms admit a total order [Yue et al., 2012, Yue and Joachims, 2009, Chen and Frazier, 2017]. Alternative notions of optimality are discussed in Section A.

Objective: Condorcet winner identification. Given $\delta \in (0, 1)$, the learner must output the CW i^* with probability at least $1 - \delta$ by adaptively and sequentially choosing pairs (i, j) to be compared and choosing a stopping time. We evaluate an algorithm by its (random) number of duels N_δ, called the *budget*, and we seek instance-dependent guarantees in terms of the centered gaps $\Delta_{i,j} = q_{i,j} - \frac{1}{2}$, which encode both preference direction (e.g., $\Delta_{i,j} > 0$ means i beats j) and statistical difficulty.

State-of-the-art. Although CW identification has attracted quite a lot of attention [Komiyama et al., 2015, Ailon et al., 2014, Chen and Frazier, 2017, Saha and Gaillard, 2022, Peköz et al., 2022], the optimal budget for this task remains poorly understood. Maiti et al. [2024] developed an algorithm that exploits that the CW row is the unique one with only positive gaps. They obtained a high-probability guarantee of the budget of the order of $H_{\mathrm{cw}}(\delta)$ where

$$H_{\mathrm{cw}}(\delta) := \log(1/\delta) \sum_{i \neq i^*} \frac{1}{\Delta^2_{i^*,i}} \ . \tag{1}$$

In the specific scenario where the *CW is the strongest opponent of every suboptimal arm*, that is,

$$i^* = \mathrm{argmin}_{j:\Delta_{i,j}<0} \Delta_{i,j} \ \ , \forall i \neq i^* \ , \tag{CW-SO}$$

the condition (1) is matched by the lower bound from Haddenhorst et al. [2021] which implies that the sample complexity $H_{\mathrm{cw}}(\delta)$ is optimal. Although (CW-SO) is satisfied for Strong Stochastic Transitive (SST) models[2], this condition is arguably quite strong and is not even guaranteed when there is a total order on the arms. In particular, the bound (1) can be overly conservative when some arm i is nearly tied with i^* as it largely ignores potentially informative comparisons among suboptimal arms.

Prior to that, Karnin [2016] introduced a verification-based approach. The expected budget of this procedure asymptotically satisfies

$$\mathbb{E}[N_\delta] \leq c \log\left(\frac{1}{\delta}\right) \sum_{i \neq i^*} \min_{j:\Delta_{i,j}<0} \frac{1}{\Delta^2_{i,j}} + \tilde{O}\left[\sum_{i \neq i^*} \frac{1}{\Delta^2_{i^*,i}} + \sum_{i \neq i^*} \sum_{j \neq i} \min\left(\frac{1}{\Delta^2_{i,j}}, \min_{j':\Delta_{i,j'}<0} \frac{1}{\Delta^2_{i,j'}}\right)\right], \tag{2}$$

where c is a positive numerical constant and $\tilde{O}$ hides possible logarithmic dependencies. The $\log(1/\delta)$ term in the bound (2) interprets as the sum, over all non-CW arms i, of $\log(1/\delta)/[\min_j \Delta_{i,j}]^2$ which is the minimal budget required to check whether the row $\Delta_{i,\cdot}$ is non-negative if an oracle provides to the learner the information on the best opponent of i. Indeed, Theorem 5.2 of Haddenhorst et al. [2021] establishes a lower bound of the order $\log(1/\delta)/[\min_j \Delta_{i,j}]^2$. However, in a large-scale problem where K is large compared to $\log(1/\delta)$, the right-hand side term in (2) is sizable as it scales at least like $K^2/(\max_{i \neq j} \Delta^2_{i,j})$ and its optimality is questionable. Altogether, apart from the small-scale asymptotic regime (K fixed, $\delta \to 0$) or under the restrictive condition (CW-SO), the sample complexity remains to be understood.

Open Problem. *What is the true sampling complexity of CW identification and how does it depend on gap matrix $\boldsymbol{\Delta}$ beyond the rows and columns of the Condorcet winner?*

[2] SST assumes a total order such that, whenever i is ranked above j and j above k, we have: $q_{i,k} \geq \max\{q_{i,j}, q_{j,k}\}$.

This question falls within *structured* pure exploration, where the feedback is noisy but constrained by an underlying latent object (here, a skew-symmetric matrix possessing a positive row), so the goal is to exploit structure rather than estimate all entries. We emphasize that such problems require ideas and techniques that are at the crossroads of active learning and high-dimensional statistics. Related challenges arise in noisy payoff matrix games Maiti [2025], e.g., in pure Nash equilibrium identification.

Contributions. Our main contributions are threefold: (i) we introduce new elimination-based algorithms for both the fixed budget and the fixed confidence settings and provide non-asymptotic guarantees, (ii) we establish matching lower bounds. (iii) Overall, this allows us to highlight the *trade-offs and the multiple strategies that underlie CW identification*. For the sake of simplicity, we focus in the main text on the δ-PAC setting; we also treat the fixed-budget framework and provide analogous guarantees in Appendices E and C. Our guarantees recover, up to logarithmic factors, the known optimal rate $H_{\mathrm{cw}}(\delta)$ in the favorable structured regimes such as (CW-SO). In contrast, when this restrictive assumption fails, our bounds can be drastically smaller than the existing $H_{\mathrm{cw}}(\delta)$-type guaranties of (1) or the bound (2) of Karnin [2016] by adapting to the full gaps matrix.

At a high level, our elimination-based procedure (FC-CWI), described in Algorithms (1) and (2),

iteratively scores the current candidates for CW using subroutines that (i) *search* in the gap matrix $\mathbf{\Delta}$ for informative comparisons and exploit its skew-symmetric structure, and (ii) *estimate* the signs of the discovered entries with sufficient accuracy. Candidates are then ranked by these scores and a constant fraction of arms is eliminated at each round. Our analysis reveals a delicate dependence on the full gap matrix $\mathbf{\Delta}$. Indeed, providing evidence that i^* is the CW either amounts to showing that all the CW gaps $\{\Delta_{i^*,i}\}_{i\neq i^*}$ are positive or amounts to showing that all arms $i \neq i^*$ are not CW. The evidence of sub-optimality for a given arm $i \neq i^*$ is governed both by the number of negative entries in its row, $K_{i;<0} := |\{j : \Delta_{i,j} < 0\}|$, and by the magnitudes of these gaps, denoted by the ordered values $\Delta_{i,(1)} \leq \cdots \leq \Delta_{i,(K_{i;<0})} < 0$. For each $i \neq i^*$, fix an integer $s_i \leq K_{i;<0}$ and write $\boldsymbol{s} = (s_1, \ldots, s_K)$. The following results will involve a trade-off in $\boldsymbol{s}$. Our analysis decomposes the complexity into $H_{\mathrm{cw}}(\delta)$ –see (1)–, which corresponds to the cost of separating i^* from every competitor only relying on duels with i^*, as well as two new components:

- *Exploration/Selection cost.* This term quantifies the effort required to select a negative entry whose absolute value is at least $|\Delta_{i,(s_i)}|$ in each suboptimal row.

$$H_{\mathrm{explore}}(\boldsymbol{s}, \delta) := \max_{i\neq i^*} \frac{K \log(1/\delta)}{s_i \Delta^2_{i,(s_i)}} + \sum_{i\neq i^*} \frac{K}{s_i \Delta^2_{i,(s_i)}} \ , \tag{3}$$

  Note that the second term of the right-hand-side expression is independent of δ and accounts for the fact that looking for an entry at least $|\Delta_{i,(s_i)}|$ out of K depends on both the number s_i of such entries and the magnitude $|\Delta_{i,(s_i)}|$. The $\log(1/\delta)$-dependency only arises for a single arm $i \neq i^*$.

- *Certification cost.* This term corresponds to the number of samples required to estimate the signs of the selected gaps (at the exploration step) at confidence level $1 - \delta$

$$H_{\mathrm{certify}}(\boldsymbol{s}, \delta) := \sum_{i\neq i^*} \frac{\log(1/\delta)}{\Delta^2_{i,(s_i)}} \ .$$

Our main upper bound shows that, with probability at least $1 - \delta$, the budget N_δ of FC-CWI satisfies

$$N_\delta \lesssim H_{\mathrm{cw}}(\delta) \wedge \min_{\substack{(s_i)_{i\neq i^*} \\ \forall i,\ s_i \leq K_{i;<0} \wedge K/8}} \{H_{\mathrm{certify}}(\boldsymbol{s}, \delta) + H_{\mathrm{explore}}(\boldsymbol{s}, \delta)\} \ , \tag{4}$$

where the notation $\lesssim$ hides logarithmic factors in K, $(\Delta_{i,(1)})_{i\neq i^*}$ and a $\log\log(1/\delta)$ factor. Under the scenario (CW-SO), our procedure still achieves budget smaller than $H_{\mathrm{cw}}(\delta)$ as in Maiti et al. [2024] but also achieves better guarantees for other gap matrices $\mathbf{\Delta}$, where the budget is driven by the right-hand side in (4). In the above infimum in (4), the smaller the s_i's are, the smaller $H_{\mathrm{certify}}(\boldsymbol{s}, \delta)$ is, but the exploration cost $H_{\mathrm{explore}}(\boldsymbol{s}, \delta)$ for localizing a good candidate can increase for small s_i's. In the following, we denote $\boldsymbol{s}^*_{\mathbf{\Delta}}$ as the vector $(s^*_i)_{i\neq i^*}$ achieving the best trade-off in Equation (4). We interpret $s^*_{\mathbf{\Delta}}$ as an effective sparsity of $\mathbf{\Delta}$, although it also depends on δ. Importantly, our algorithm does not take $\boldsymbol{s}^*_{\mathbf{\Delta}}$ as input and therefore *automatically* achieves the best balance captured by (4).

In Section 3, we provide matching lower bounds supporting the optimality of our budget in all regimes with respect to δ and K, as well as for general forms of the matrix $\mathbf{\Delta}$. Importantly, this showcases that the exploration/certification trade-off unveiled in (4) is unavoidable and intrinsic to the sample complexity of CW-identification. The benefit of our procedure is also illustrated on numerical experiments.

Emblematic regimes. As the sample complexity is quite intricate in the general case, we discuss some specific regimes to emphasize key phenomena.

- *CW has small gaps*. If some entries of the i^*-th row of Δ are small, then $H_{\text{cw}}(\delta)$ is not the minimum in (4) and our sample complexity is of the order of $H_{\text{certify}}(\boldsymbol{s}^*_{\mathbf{\Delta}}, \delta) + H_{\text{explore}}(\boldsymbol{s}^*_{\mathbf{\Delta}}, \delta)$ which *can be arbitrarily smaller* than the budget $H_{\text{cw}}(\delta)$ of Maiti et al. [2024].
- *Large-scale/fixed probability regime*. Let us consider δ as a constant and K as large. In (4), when $H_{\text{cw}}(\delta)$ is not the minimum, then the sample complexity is of the order[3] of

$$\min\left(\sum_{i\neq i^*}\frac{1}{\Delta^2_{i^*,i}}, \sum_{i\neq i^*}\frac{K}{\|\Delta_i^-\|_2^2}\right),$$

  where $\Delta^-_{i,j} := \min(\Delta_{i,j}, 0)$. In comparison to the complexity of Karnin [2016] in (2), our sample complexity is always smaller and can be smaller by a factor as larger as K.
- *Small probability regime*. This last observation is more subtle. Similarly to Karnin [2016], consider the asymptotic regime where $\log(1/\delta)$ goes to infinity, while K and $\mathbf{\Delta}$ are fixed. In this asymptotic, our lower and upper bounds on the $(1-\delta)$-quantile of the budget are of the form

$$\log\left(\frac{1}{\delta}\right)\inf_{\boldsymbol{s}}\left[\sum_{i\neq i^*}\frac{1}{\Delta^2_{i,(s_i)}} + \max_{i\neq i^*}\frac{K}{s_i\Delta^2_{i,(s_i)}}\right],$$

  whereas the bound in Karnin [2016] on the expected budget only involves the smaller quantity $\sum_{i\neq i^*}\frac{\log(1/\delta)}{\Delta^2_{i,(1)}}$. Since our bound is optimal in quantile, this establishes that there is a significant, yet intrinsic, gap between guarantees in expectation and in quantile of the budget. Especially, when there is heterogeneity in the $\Delta_{i,(s_1)}$, this gap can be as large as a factor K. This phenomenon is central for the analysis of fixed-budget algorithms - see Section C.

All the way through these two extreme regimes (δ fixed, $\delta \to 0$), both (4) and the matching lower bound illustrate a trade-off between exploration and certification. Intuitively, when $\log(1/\delta)$ increases, the effective sparsity $s^*_{\mathbf{\Delta}}$ tends to decrease so the algorithm explores other arms more thoroughly to identify stronger opponents.

Technical Innovations. Our Algorithms 1 and 2 are based on a new iterative scoring strategy that builds on the selection, for each 'active' arm i, of a strong opponent as well as the estimation of some quantile of the estimation $\Delta_{i,.}$. For that purpose, we need to introduce a new active quantile estimation algorithm achieving optimal ϵ-error simultaneously for all ϵ –see Appendix B. Our main lower bounds use novel approaches and techniques as our aim is to lower bound the $(1-\delta)$ quantile of the budget. For that purpose, we reduce the problem to an active multiple testing problem of the existence of negative entries within a vector of size $K-1$, our lower bound introduce techniques that go beyond classical MAB lower bounds are at the crossroad of high-dimensional statistics and bandit theory.

Organization. Section 2 presents our algorithms and fixed-confidence upper bounds; the fixed-budget guarantees are deferred to the appendix. Section 3 gives lower bounds for the fixed-confidence setting. Section 4 illustrates our results on numerical experiments. Section 5 concludes with implications and future directions, while related work is discussed in Section A. All proofs and fixed-budget results are provided in the appendix.

[3]To show this, we observe that $\max_{s_i=1,\dots,K_{i;<0}} s_i\Delta^2_{i,(s_i)} \asymp \|\Delta_i^-\|_2^2$, see Lemma H.1.

2 Upper Bounds: Algorithm and Guarantees

This section presents our fixed-confidence identification procedure and the upper bound announced in Section 1. The algorithm is built from a budgeted elimination-and-certification subroutine, denoted FB-CWI (Algorithm 1). Given a tentative budget scale T, a confidence level δ, this subroutine performs one elimination run using $\mathcal{O}(T)$ duels and returns a candidate arm together with a Boolean certificate indicating whether the run can be safely stopped. The budget T should be viewed only as a guessed complexity scale: if it is too small for the instance, the subroutine is allowed to fail to certify rather than output a final answer.

The fixed-confidence algorithm FC-CWI (Algorithm 2) removes the need to know this scale in advance by calling FB-CWI with geometrically increasing values of T and stopping at the first successful certificate. Thus, the main algorithm is a genuine δ-PAC procedure: the stopping time is determined by the certification step, while the budgeted subroutine only provides the candidate-generation and elimination mechanism. We describe this subroutine first, and then state the high-probability sample-complexity guarantee for the resulting doubled procedure. The corresponding fixed-budget framework and guarantees are deferred to the appendix.

FB-CWI is an elimination procedure initialized with $A_1 = [K]$: at each round k, it assigns a score $S_k(\alpha)$ to every active arm $\alpha \in A_k$, ranks the arms accordingly, and discards the bottom $1/8$ fraction (so $|A_{k+1}| = \lfloor 7|A_k|/8 \rfloor$). Therefore, the number of rounds is $O(\log K)$.

The core of FB-CWI is the score computation, whose purpose is to keep the CW ranked above the elimination threshold. At round k, we split a budget of order $T/\log(K)$ across the active set A_k. For each $\alpha \in A_k$, we devote one quarter of its share to search for a strong opponent by running Bracketed Sequential Halving (BSH), a variant of SH [Karnin et al., 2013], with the minimization convention of Section B.3, on the instance of the duels $\{(\alpha, \beta) : \beta \in [K] \setminus \{\alpha\}\}$, yielding an opponent $\alpha^{(s)}$ with a small gap $\Delta_{\alpha,\alpha^{(s)}}$, and another quarter to estimate the gap $\Delta_{\alpha,\alpha^{(s)}}$ via an empirical mean; this estimate defines the strong-opponent component of the score. We call $\alpha^{(s)}$ 'strong' because it is selected from all K arms (not only from A_k): this tends to penalize sub-optimal arms more sharply, at the price of higher uncertainty due to the larger search space. Relying only on the strong-opponent term can be brittle: if the CW is nearly tied with some arm, the selected opponent may yield a gap estimate close to zero and provide little separation with the elimination threshold. We therefore add a 'weak-opponent' term that yields adaptivity to larger gaps with the CW. More specifically, writing $\mathbf{\Delta}^{(k)} := \mathbf{\Delta}_{A_k \times A_k}$, skew-symmetry implies that at least half of the entries of $\mathbf{\Delta}^{(k)}$ are non-positive, and a simple pigeon-hole type argument implies that at least $|A_k|/4$ rows contain at least $|A_k|/4$ off-diagonal non-positive entries (Lemma H.6 in the appendix). Accordingly, for each $\alpha \in A_k$ we estimate a point whose value lies between the $1/8$- and $1/4$-quantiles of the row $(\Delta_{\alpha,\beta})_{\beta \in A_k \setminus \{\alpha\}}$ (via RANGE-QUANTILE). This lower-tail statistic is typically negative for many sub-optimal arms pushing them into the bottom-$1/8$ region, while remaining positive for the CW and leveraging the fact that most of its gaps can still be large.

Subroutines: BRACKETED SEQUENTIAL HALVING and RANGE-QUANTILE. Our score construction relies on two subroutines. For the strong-opponent search we use Bracketed Sequential Halving (BSH) [Zhao et al., 2023], an anytime and parameter-free variant of SH, [Karnin et al., 2013], designed for the data-poor regime, chosen for its adaptive guarantees on simple regret, which translate in our context into gap-dependent guarantees –see Zhao et al. [2023] and Section B.3 of the appendix. For the weak-opponent choice, we introduce Range-Quantile (Algorithm 3), a general fixed-budget procedure revealing a point in a prescribed quantile range: given N arms with means $(\mu_i)_{i \in [N]}$ ordered as $\mu_{(1)} \leq \cdots \leq \mu_{(N)}$ and indices $d < u$, it returns an estimate $\hat{t}$ that falls between the d-th and u-th means (up to an additive error ε) with error probability decaying as $\exp(-\widetilde{\Theta}(\min\{\frac{d}{N}, 1 - \frac{u}{N}\}\frac{(u-d)^2}{N^2}T\varepsilon^2))$ –see Theorem B.1). Importantly, RANGE-QUANTILE does not require ε as input, and the same output satisfies a deviation bound for every ϵ; in FB-CWI we instantiate it with $N = |A_k| - 1$, $d = \lceil (N+1)/8 \rceil$ and $u = \lceil (N+1)/4 \rceil$ and analyze it at accuracy $\varepsilon = \frac{1}{2}\Delta_{i^*,(\lceil |A_k|/8 \rceil)}$ to obtain a value between the $1/8$- and $1/4$-quantiles of $(\Delta_{\alpha,\beta})_{\beta \in A_k \setminus \{\alpha\}}$ up to this additive error. When $N \leq 3$, these ranks satisfy $d = u = 1$, and we use the fallback specified in Algorithm 3. Note that Maiti et al. [2024] gives a fixed-confidence routine that, given (δ, ε), outputs, in their decreasing-order convention, a value in $[\mu_{(N/2)} - \varepsilon,\ \mu_{(N/4+1)} + \varepsilon]$ with probability at least $1 - \delta$. Here, ε is instance-dependent and unknown, which motivates our adaptive RANGE-QUANTILE subroutine that does not take ε as input.

Given a prescribed rank interval and a fixed sampling budget, quantile estimation must balance inspecting enough arms to locate the interval against measuring their means accurately. RANGE-QUANTILE adapts to this trade-off without a prescribed accuracy parameter: a Lepski-type comparison across sampling scales produces one estimate whose deviation probability is controlled for every $\varepsilon \in (0,1)$. The construction and proof are given in Appendix B.

Algorithm 1 FB-CWI + Certification

Input: Fixed budget (T), Certification(δ, T, c).
If $T < 8K\log_{8/7}(K)$, return `False` and an arbitrary arm $I \in [K]$.
$k \leftarrow 1$, $A_1 \leftarrow [K]$, $n \leftarrow 1 + \left\lceil \log_2\left(\frac{T}{8K\log(K)}\right)\right\rceil$.
$\phi_1, \phi_2 \leftarrow$ `True`
while $|A_k| > 1$ **do**
  Let $B_k \leftarrow \left\lfloor \frac{T}{|A_k|\log_{8/7}(K)} \right\rfloor$.
  for $\alpha \in A_k$ **do**
    `/* Finding a strong opponent */`
    • Run Bracketed SH as specified in Section B.3 with a budget $\lceil B_k/4 \rceil$ and where the candidate arms are $\{(\alpha, \beta)$ for $\beta \in [K] \setminus \{\alpha\}\}$. Let $(\alpha, \alpha^{(s)})$ denote the output.
    • Query $\lceil B_k/4 \rceil$ samples of $(\alpha; \alpha^{(s)})$ and compute the quantity $Z_k^{(s)}(\alpha)$ (corresponding to the empirical mean of the gaps).
    `/* Computing a weak score */`
    • Run Range-Quantile on duels between α and arms in $A_k \setminus \{\alpha\}$, with a budget $\lceil B_k/2 \rceil$ and quantiles $(\lceil |A_k|/8 \rceil, \lceil |A_k|/4 \rceil)$ let $Z_k^{(w)}(\alpha)$ denote the output.
  end for
  Compute the scores $S_k(\alpha) = \min\{Z_k^{(s)}(\alpha), 0\} + Z_k^{(w)}(\alpha)$ for each $\alpha \in A_k$.
  Rank the arms in A_k following the scores $S_k(\cdot)$ and put in A_{k+1} the top $|A_k| - \lceil |A_k|/8 \rceil$ arms.
  `/* Check fixed confidence */`
  Let $\bar{\alpha}$ denote the arm ranked $|A_k| - \lceil |A_k|/8 \rceil + 1$ according to the scores $S_k(\cdot)$.
  $\phi_1 \leftarrow \phi_1 \cdot \mathbb{1}\left(S_k(\bar{\alpha}) < -\sqrt{\frac{2c\log(T)}{\lceil B_k/4 \rceil}\log\left(8K^2\log_{8/7}(K)\log(T)\cdot\frac{n(n+1)}{\delta}\right)}\right)$
  $k \leftarrow k+1$.
end while
Let I denote the unique arm in A_k.
`/* Use` T `queries to test the sign of gaps of` I `(unique arm in` A_k`) at confidence` δ `*/`
Run Test-CW with inputs (I, δ, T) to check the sign of gaps of arm I, let ϕ_2 denote its output.
Return $\phi_1 \vee \phi_2$ and I.

Algorithm 2 FC-CWI

Input: c, confidence input δ.
$\varphi \leftarrow$ `False`, $T \leftarrow 8K\log(K)$.
while $\neg\varphi$ **do**
  Run FB-CWI procedure (Algorithm 1) with inputs δ, T, c. Let φ, I be its output.
  $T \leftarrow 2T$.
end while
Return I.

From a budgeted subroutine to fixed confidence. Algorithm 1 is used as a certificate-driven subroutine: for a tentative budget scale T and confidence level δ, it spends $\mathcal{O}(T)$ duels and returns a candidate I together with a success flag $\Psi = \phi_1 \vee \phi_2$. The direct certificate ϕ_2 verifies that I is the CW, leading to the classical $H_{\mathrm{cw}}(\delta)$ regime. The elimination certificate ϕ_1 instead checks that the elimination frontier remains safely negative, thereby exploiting informative comparisons among suboptimal arms. The analysis shows that, up to logarithmic factors, certification succeeds once

$$T \gtrsim H_{\mathrm{cw}}(\delta) \wedge \min_{\substack{(s_i)_{i\neq i^*} \\ s_i \in \{1,\dots,K_{i;<0}\}}} \left\{H_{\text{certify}}(\boldsymbol{s},\delta) + H_{\text{explore}}(\boldsymbol{s},\delta)\right\} . \tag{5}$$

Since these quantities are unknown, we run FB-CWI under a doubling schedule on T and stop as soon as a budgeted verification succeeds. It uses at most $2T$ queries per stage (in Algorithm 1, ϕ_1 is computed from the same samples as the scores, and ϕ_2 runs TEST-CW with at most T extra queries). A direct verification is to certify that the returned arm I is Condorcet by testing $\Delta_{I,j} > 0$ for all $j \neq I$ at level $1-\delta$, which, when $I = i^*$, costs $\sum_{j \neq I} \log(1/\delta)/\Delta_{I,j}^2$ up to logarithmic factors and matches the $H_{\mathrm{cw}}(\delta)$ regime. When the second term in (5) is dominant, our identification relies instead on certifying sub-optimality: many sub-optimal arms have scores $S_k(\cdot)$ that are typically negative while the CW remains positive. This motivates the second verification ϕ_1, which checks that the elimination frontier is on the negative side. We stop at the first success of ϕ_1 or ϕ_2, which guarantees δ-correctness. The complete proof of Theorem 2.1 is presented in Section D of the appendix.

Theorem 2.1. *There exists a constant c_0 such that the following holds for any $\delta \in (0, 1/6)$. Under the assumption of the existence of a CW, the output of Algorithm 2, denoted ψ, with input (δ, c) where $c \geq c_0$ satisfies: $\mathbb{P}(\psi_\delta \neq i^*) \leq \delta$.*

Moreover, the total number of queries N_δ satisfies, with probability at least $1-\delta$,

$$N_\delta \leq \tilde{c} \cdot \Big(H_{cw}(\delta) \wedge \min_{\substack{(s_i)_{i \neq i^*} \\ \forall i \neq i^*,\ s_i \in \{1,\dots,K_{i,<0}\}}} \big\{ H_{certify}(\boldsymbol{s}, \delta) + H_{explore}(\boldsymbol{s}, \delta) \big\} \Big) ,$$

where $\tilde{c}$ is proportional to c and hides logarithmic factors in K and the instance-dependent complexity terms, including the $\log^6$ exploration warm-up factor, and a $\log\log(1/\delta)$ factor. The quantities $H_{cw}, H_{certify}, H_{explore}$ are defined in Section 1 .

Remarks (On the constant c in the stopping rule)**.** *The stopping condition in Algorithm 1 involves a numerical constant c_0 inherited from the high-probability analysis of Corollary B.2. This* absolute *value of c_0 can be made explicit by tracking constants in the proof. Since we did not optimize numerical factors, we keep c_0 symbolic for readability.*

Guarantees: intuition. The proof of Theorem 2.1 analyzes one call of the budgeted subroutine FB-CWI at a tentative scale T. In the outer fixed-confidence algorithm, a call that does not certify is harmless: FC-CWI simply doubles T and tries again. Thus, the key point is to show that once T exceeds the relevant instance-dependent scale, the CW is not eliminated and one of the certificates succeeds with high probability. Conditionally on $i^* \in A_k$, the only way the elimination step can be harmful is if the CW score falls below the bottom-$1/8$ cutoff in some round. The analysis therefore controls the separation between the CW score and the elimination frontier across the $O(\log K)$ rounds.

The weak-opponent term gives a baseline separation. At round k, the CW benefits from a positive lower-tail gap of order $\Delta_{i^*,(\lceil |A_k|/8 \rceil)}$, estimated with $B_k = \Theta(T/(|A_k| \log K))$ samples. Concentration bounds, together with the RANGE-QUANTILE guarantee, yield an error of order $\exp(-\widetilde{\Theta}(B_k \Delta^2_{i^*,(\lceil |A_k|/8 \rceil)}))$. The worst round is controlled by

$$\max_k \frac{|A_k|/8}{\Delta^2_{i^*,(\lceil |A_k|/8 \rceil)}} \leq \max_{i \in [K-1]} \frac{i}{\Delta^2_{i^*,(i)}} \leq \sum_{i \neq i^*} \frac{1}{\Delta^2_{i,i^*}} ,$$

which leads, after the final CW verification, to the classical $H_{\mathrm{cw}}(\delta)$ regime.

The strong-opponent term yields the matrix-adaptive regime by actively finding negative witnesses for suboptimal arms. Fix integers $1 \leq s_i \leq K_{i,<0}$. For each $i \neq i^*$, Bracketed Sequential Halving needs budget of order $K/(s_i \Delta^2_{i,(s_i)})$ to locate an opponent whose gap is at most $\frac{7}{8}\Delta_{i,(s_i)}$, while certifying the sign of this gap costs $1/\Delta^2_{i,(s_i)}$, up to logarithmic factors. The aggregate cost of locating such witnesses gives to $\sum_{i \neq i^*} \frac{K}{s_i \Delta^2_{i,(s_i)}}$ whereas the high-confidence part is governed by $\sum_{i \neq i^*} \frac{1}{\Delta^2_{i,(s_i)}} + \max_{i \neq i^*} \frac{K}{s_i \Delta^2_{i,(s_i)}}$. Thus, certification succeeds once T is larger than the bound stated in Theorem 2.1.

3 Instance Dependent Lower Bounds

Beyond the instance-dependent lower bound (2) from Haddenhorst et al. [2021] that we complement with a corresponding quantile-bound and restate as Proposition 3.1, we establish in Theorem 3.2

the first lower bound that highlight the intrinsic cost of detecting informative entries for CWI. In Appendix E, we also derive fixed-budget minimax lower bounds. Denote by $\mathbb{D}_{\mathrm{cw}}$ the class of dueling bandit environments that admits a CW[4] :

$$\mathbb{D}_{\mathrm{cw}} := \left\{ \boldsymbol{\Delta} \in [-\tfrac{1}{4}, \tfrac{1}{4}]^{K\times K} : \boldsymbol{\Delta} = -\boldsymbol{\Delta}^T \text{ and } \exists\, i^* \in [K] \text{ such that } \forall j \neq i^*,\ \Delta_{i^*,j} > 0 \right\} . \quad (6)$$

We say that an algorithm A is δ-correct for CW identification if, for any $\boldsymbol{\Delta} \in \mathbb{D}_{\mathrm{cw}}$, it identifies i^* with error probability at most δ, that is, $\mathbb{P}_{\boldsymbol{\Delta},A}(\hat{i} \neq i^*(\boldsymbol{\Delta})) \leqslant \delta$, where $\mathbb{P}_{\boldsymbol{\Delta},A}$ denotes the probability[5] induced by the interaction between A and the environment with gap matrix $\boldsymbol{\Delta}$.

Proposition 3.1. *Let $K \geqslant 2$ and $\delta \in (0, 1/6)$ and consider any gap matrix $\boldsymbol{\Delta} \in \mathbb{D}_{\mathrm{cw}}$. For any algorithm A that is δ-correct on $\mathbb{D}_{\mathrm{cw}}$, the budget N_δ satisfies*

$$\mathbb{E}_{\boldsymbol{\Delta},A}[N_\delta] \ \geqslant\ \frac{1}{4}\sum_{i\neq i^*} \frac{\log(1/(4\delta))}{\Delta^2_{i,(1)}} \text{ , and } \quad \mathbb{P}_{\boldsymbol{\Delta},A}\left(N_\delta \geqslant \frac{1}{3}\sum_{i\neq i^*} \frac{\log\left(1/(6\delta)\right)}{\Delta^2_{i,(1)}} \right) \geqslant \delta. . \quad (7)$$

The lower Bound of the expected budget is a particular case of Theorem 5.2 in Haddenhorst et al. [2021]. We complement it here with a quantile-bound. This proposition certifies optimality of the complexity $H_{\mathrm{cw}}(\delta)$ in the CW–SO regime. Indeed, under (CW-SO), the strongest negative entry of every suboptimal row is the comparison with the Condorcet winner, so $\Delta_{i,(1)} = \Delta_{i,i^\star} = -\Delta_{i^\star,i}$. Hence the lower bound in Proposition 3.1 is, up to constants, $\sum_{i\neq i^\star} \log(1/\delta)/\Delta^2_{i^\star,i} = H_{\mathrm{cw}}(\delta)$.

To characterize the optimality of (4), we establish a lower bound on the δ-quantile of any algorithm under a local class of instances. The budget condition (4) only depends on the gap matrix $\boldsymbol{\Delta}$ through three vectors: (i) the row i^* of the CW $\Delta_{i^*,.}$, (ii) the effective sparsity $\boldsymbol{s}^*_{\boldsymbol{\Delta}}$, and (iii) the gaps at the sparsity level $\boldsymbol{s}^*$: $(\Delta_{i,(s^*_i)})_{i\neq i^*}$. Given any gap matrix $\boldsymbol{\Delta}$, we define the collection $\mathbb{D}(\boldsymbol{\Delta})$ of gap matrices $\tilde{\boldsymbol{\Delta}}$ that leave the Condorcet winner $i^*_{\boldsymbol{\Delta}}$, the effective sparsity $\boldsymbol{s}^*_{\boldsymbol{\Delta}}$, and the gaps $(\Delta_{i,(s^*_i)})_{i\neq i^*}$ unchanged

$$\mathbb{D}(\boldsymbol{\Delta}) := \{\tilde{\boldsymbol{\Delta}} \text{ s.t. } \ i^*_{\tilde{\boldsymbol{\Delta}}} = i^*_{\boldsymbol{\Delta}},\ \boldsymbol{s}^*_{\tilde{\boldsymbol{\Delta}}} = \boldsymbol{s}^*_{\boldsymbol{\Delta}},\ (\tilde{\Delta}_{i,(s^*_i)})_{i\neq i^*} = (\Delta_{i,(s^*_i)})_{i\neq i^*}\} \ . \quad (8)$$

Theorem 3.2. *Let A be a δ-correct algorithm for CW identification, with $\delta \leqslant 1/12$, and let $\boldsymbol{\Delta} \in \mathbb{D}_{\mathrm{cw}}$. Assume that $\boldsymbol{\Delta}$ has no ties, that is, $\forall i \neq j$, $\Delta_{i,j} \neq 0$. For this matrix $\boldsymbol{\Delta}$, one can construct a matrix $\tilde{\boldsymbol{\Delta}}$ by permuting the entries of $\boldsymbol{\Delta}$ in such a way that $\tilde{\boldsymbol{\Delta}} \in \mathbb{D}(\boldsymbol{\Delta})$, and such that*

$$\mathbb{P}_{\tilde{\boldsymbol{\Delta}},A}\left(N_\delta \ \geqslant\ \frac{1}{3}\max_{i\neq i^*} \frac{K_{i;<0}}{\|\Delta_i^-\|^2}\log\left(\frac{1}{6\delta}\right) \ \vee\ \frac{1}{37\log(2K)}\sum_{i\neq i^*}\frac{K_{i;<0}}{\|\Delta_i^-\|^2} \right) \ \geqslant\ \delta \ . \quad (9)$$

Moreover, for all $i \neq i^*$, the rows satisfy $(\tilde{\Delta}_{i,(j)})_{j\leq K_{i;<0}} = (\Delta_{i,(j)})_{j\leq K_{i;<0}}$, i.e., $\tilde{\Delta}_{i,.}$ and $\Delta_{i,.}$ share the same $K_{i;<0}$ negative entries, up to permutation.

Remarks. *While the bound* (7) *from Proposition 3.1 captures the cost of certification, the lower bound* (9) *captures the intrinsic need of exploration. Indeed, from the classical bound of Lemma H.1, we have $\min_{\boldsymbol{s}} H_{explore}(\boldsymbol{s},\delta) = \max_{i\neq i^*} K\log(1/\delta)/\|\Delta_i^-\|^2 + \sum_{i\neq i^*} K/\|\Delta_i^-\|^2$, up to a factor $\log(2K)$. Importantly, the two lower bounds in Proposition 3.1 and Theorem 3.2 imply the following.*

Corollary 3.3. *Let A be a δ-correct algorithm for CW identification, with $\delta \leqslant 1/12$, and let $\boldsymbol{\Delta} \in \mathbb{D}_{\mathrm{cw}}$. Then, one can construct $\tilde{\boldsymbol{\Delta}} \in \mathbb{D}(\boldsymbol{\Delta})$, such that*

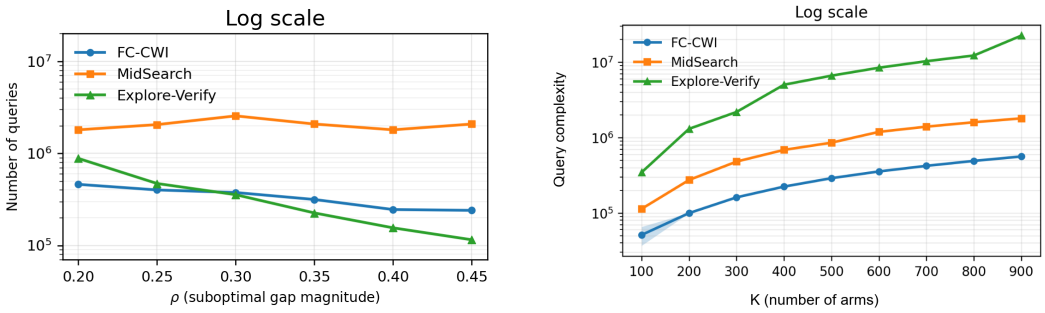
$$\mathbb{P}_{\tilde{\boldsymbol{\Delta}},A}\left(N_\delta \gtrsim H_{certify}(\boldsymbol{s}^*,\delta) + H_{explore}(\boldsymbol{s}^*,\delta)\right) \geqslant \delta \ , \quad (10)$$

where $\gtrsim$ hides a log term in K and a numerical constant.

This corollary establishes, up to logarithmic factors, the local minimax optimality of $H_{\mathrm{certify}}(\boldsymbol{s}^*,\delta) + H_{\mathrm{explore}}(\boldsymbol{s}^*,\delta)$ over $\mathbb{D}(\boldsymbol{\Delta})$. By definition, the hard instance $\tilde{\boldsymbol{\Delta}} \in \mathbb{D}(\boldsymbol{\Delta})$ has the same Condorcet winner i^*, effective sparsities $\boldsymbol{s}^*$, and selected gaps $(\Delta_{i,(s^*_i)})_{i\neq i^*}$ as $\boldsymbol{\Delta}$. Its construction first builds an auxiliary matrix with these parameters and then applies the permutation argument of Theorem 3.2.

[4]The restriction to gaps in away from $-1/2$ and $1/2$ is standard in lower bounds with Bernoulli rewards.

[5]denote $\mathbb{E}_{\boldsymbol{\Delta},A}$ for the corresponding expectation



(a) Informativeness sweep.

(b) Large-scale behavior.

Figure 1: Left: query complexity in the sparse gaps matrix model with $K = 64$, $\alpha = 0.3$, and $\delta = 0.1$, as the suboptimal gap magnitude ρ varies. Right: query complexity as a function of K in the sparse model with $\gamma = 0.3$, $\rho = 0.4$, and $\alpha = 0.3$, shown on a logarithmic scale.

Proof Sketch Theorem 3.2. The proof reduces CW identification to a collection of active signal-detection tasks: for each non-winner row, the learner must certify the existence of at least one negative entry under a global error constraint. The key technical novelty is an adversarial permutation construction that hides the locations of negative gaps and reduces the problem to multiple hypothesis testing; this yields the complexity term $\frac{K_{i;<0}}{\|\Delta_i^-\|^2}\log(1/\delta)$ via active signal-detection lower-bound arguments. Finally, the two terms in (9) are obtained through new multiple-hypothesis techniques.

4 Numerical simulations

We compare FC-CWI with two standard baselines: MidSearch [Maiti et al., 2024] and Explore–Verify [Karnin, 2016]. Our experiments target two predictions of the theory. First, we vary the strength ρ of informative suboptimal-suboptimal comparisons, while keeping the direct CW gaps fixed, to isolate the benefit of exploiting the full gap matrix. Second, we study the scaling with the number of arms K in the same sparse informative regime. Each method is run with a fixed numerical prefactor, chosen once to compensate for conservative theoretical constants and then kept unchanged across all main-text experiments; details and robustness checks are given in Appendix I.

All instances are sparse skew-symmetric gap matrices with arm 0 as the Condorcet winner. The CW beats every other arm with margin at least γ, while the suboptimal block contains an α fraction of nonzero entries per row, with gaps equal to $\pm\rho$ and all remaining entries set to zero. Thus, γ controls the difficulty of direct CW verification, whereas ρ controls the usefulness of suboptimal-suboptimal comparisons. The precise construction is deferred to Appendix I.

We fix $K = 64$, $\alpha = 0.3$, and $\delta = 0.1$, and vary ρ in the range $[0.2, 0.45]$ with step 0.05. For the large-scale diagnostic, we fix $\gamma = 0.3$, $\rho = 0.4$, and $\alpha = 0.3$, and vary $K \in \{100, 200, \dots, 900\}$. Figure 1 summarizes the two diagnostics. In the left panel, FC-CWI remains more sample-efficient than MidSearch for different ρ values. Explore–Verify also improves as ρ increases and can become competitive in this small-scale setting, where finding and verifying strong opponents is cheap. This behavior is consistent with the theory: the cost of Explore–Verify can be favorable at small K, but its sample complexity carries a stronger dependence on K. This is visible in the right panel, where, as K grows, Explore–Verify scales much faster than FC-CWI, while FC-CWI keeps the most favorable query growth on the logarithmic scale.

5 Discussion

In this manuscript, we consider δ-PAC and fixed-budget Condorcet-winner identification in stochastic dueling bandits under the sole assumption that a CW exists. We derive instance-dependent, high-probability sample-complexity guarantees that exploit the full gap matrix, and complement them with new lower bounds covering both large-scale regimes, where K is large, and small probability regimes, where δ is small. We further improve over the state-of-the-art under additional structural

assumptions such as weak stochastic transitivity (WST), both on the upper- and lower-bound sides; see Section A.

More broadly, CWI is a structured pure-exploration problem over a latent matrix, with close connections to noisy payoff matrix games and Nash equilibrium identification [Zhao et al., 2023, Maiti et al., 2024, 2025, Ito et al., 2025]. We believe that the exploration-certification decomposition and the tail-oriented lower-bound viewpoint may help clarify analogous expectation-versus-high-probability separations in equilibrium learning, whose optimal instance-dependent query complexity remains largely open; see Maiti [2025].

Acknowledgments and Disclosure of Funding

This work has partially been supported by ANR-21-CE23-0035 (ASCAI, ANR).

References

Nir Ailon, Zohar Karnin, and Thorsten Joachims. Reducing dueling bandits to cardinal bandits. In *International Conference on Machine Learning*, pages 856–864. PMLR, 2014.

Jean-Yves Audibert, Sébastien Bubeck, and Rémi Munos. Best arm identification in multi-armed bandits. In *COLT*, pages 41–53, 2010.

Viktor Bengs, Róbert Busa-Fekete, Adil El Mesaoudi-Paul, and Eyke Hüllermeier. Preference-based online learning with dueling bandits: A survey. *The Journal of Machine Learning Research*, 22(1): 278–385, 2021.

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

V Buldygin and K Moskvichova. The sub-gaussian norm of a binary random variable. *Theory of probability and mathematical statistics*, 86:33–49, 2013.

RM Castro. Adaptive sensing performance lower bounds for sparse signal detection and support estimation. *Bernoulli*, 20(4):2217–2246, 2014.

Bangrui Chen and Peter I Frazier. Dueling bandits with weak regret. In *International Conference on Machine Learning*, pages 731–739. PMLR, 2017.

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.

Moein Falahatgar, Alon Orlitsky, Venkatadheeraj Pichapati, and Ananda Theertha Suresh. Maximum selection and ranking under noisy comparisons. In *International Conference on Machine Learning*, pages 1088–1096. PMLR, 2017.

Moein Falahatgar, Ayush Jain, Alon Orlitsky, Venkatadheeraj Pichapati, and Vaishakh Ravindrakumar. The limits of maxing, ranking, and preference learning. In *International conference on machine learning*, pages 1427–1436. PMLR, 2018.

Sébastien Gerchinovitz, Pierre Ménard, and Gilles Stoltz. Fano's inequality for random variables. *Statistical science*, 35(2):178–201, 2020.

Maximilian Graf, Victor Thuot, and Nicolas Verzelen. Clustering items through bandit feedback: Finding the right feature out of many. In *Forty-second International Conference on Machine Learning, ICML 2025, Vancouver, BC, Canada, July 13-19, 2025*, PMLR, 2025.

Björn Haddenhorst, 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.

Katja Hofmann, Lihong Li, and Filip Radlinski. Online evaluation for information retrieval. *Found. Trends Inf. Retr.*, 10:1–117, 2016. URL `https://api.semanticscholar.org/CorpusID:34529647`.

Shinji Ito, Haipeng Luo, Taira Tsuchiya, and Yue Wu. Instance-dependent regret bounds for learning two-player zero-sum games with bandit feedback. *arXiv preprint arXiv:2502.17625*, 2025.

Kevin Jamieson, Sumeet Katariya, Atul Deshpande, and Robert Nowak. Sparse dueling bandits. In *Artificial Intelligence and Statistics*, pages 416–424. PMLR, 2015.

Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, Filip Radlinski, and Geri Gay. Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search. *ACM Transactions on Information Systems (TOIS)*, 25(2):7–es, 2007.

Zohar Karnin, Tomer Koren, and Oren Somekh. Almost optimal exploration in multi-armed bandits. In *International conference on machine learning*, pages 1238–1246. PMLR, 2013.

Zohar S Karnin. Verification based solution for structured mab problems. *Advances in Neural Information Processing Systems*, 29, 2016.

Junpei Komiyama, Junya Honda, Hisashi Kashima, and Hiroshi Nakagawa. Regret lower bound and optimal algorithm in dueling bandit problem. In *Conference on learning theory*, pages 1141–1154. PMLR, 2015.

Junpei Komiyama, Junya Honda, and Hiroshi Nakagawa. Copeland dueling bandit problem: Regret lower bound, optimal algorithm, and computationally efficient algorithm. In *International Conference on Machine Learning*, pages 1235–1244. PMLR, 2016.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Chang Li, Ilya Markov, Maarten De Rijke, and Masrour Zoghi. Mergedts: A method for effective large-scale online ranker evaluation. *ACM Transactions on Information Systems (TOIS)*, 38(4): 1–28, 2020.

R Duncan Luce et al. *Individual choice behavior*, volume 4. Wiley New York, 1959.

Arnab Maiti. Open problem: Optimal instance-dependent sample complexity for finding nash equilibrium in two player zero-sum matrix games. In *The Thirty Eighth Annual Conference on Learning Theory*, pages 6230–6234. PMLR, 2025.

Arnab Maiti, Ross Boczar, Kevin Jamieson, and Lillian Ratliff. Near-optimal pure exploration in matrix games: A generalization of stochastic bandits & dueling bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 2602–2610. PMLR, 2024.

Arnab Maiti, Ross Boczar, Kevin Jamieson, and Lillian Ratliff. Query-efficient algorithm to find all nash equilibria in a two-player zero-sum matrix game. *ACM Transactions on Economics and Computation*, 13(3):1–18, 2025.

Erol Peköz, Sheldon M Ross, and Zhengyu Zhang. Dueling bandit problems. *Probability in the Engineering and Informational Sciences*, 36(2):264–275, 2022.

Robin L Plackett. The analysis of permutations. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 24(2):193–202, 1975.

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.

El Mehdi Saad, Nicolas Verzelen, and Alexandra Carpentier. Active ranking of experts based on their performances in many tasks. In *International Conference on Machine Learning*, pages 29490–29513. PMLR, 2023.

Aadirupa Saha and Pierre Gaillard. Versatile dueling bandits: Best-of-both world analyses for learning from relative preferences. In *International Conference on Machine Learning*, pages 19011–19026. PMLR, 2022.

Aadirupa Saha and Shubham Gupta. Optimal and efficient dynamic regret algorithms for non-stationary dueling bandits. In *International Conference on Machine Learning*, pages 19027–19049. PMLR, 2022.

Max Simchowitz, Kevin Jamieson, and Benjamin Recht. The simulator: Understanding adaptive sampling in the moderate-confidence regime. In *Conference on Learning Theory*, pages 1794–1834. PMLR, 2017.

Louis L Thurstone. A law of comparative judgment. In *Scaling*, pages 81–92. Routledge, 2017.

Yisong Yue and Thorsten Joachims. Interactively optimizing information retrieval systems as a dueling bandits problem. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 1201–1208, 2009.

Yisong Yue, Josef Broder, Robert Kleinberg, and Thorsten Joachims. The k-armed dueling bandits problem. *Journal of Computer and System Sciences*, 78(5):1538–1556, 2012.

Yao Zhao, Connor Stephens, Csaba Szepesvári, and Kwang-Sung Jun. Revisiting simple regret: Fast rates for returning a good arm. In *International Conference on Machine Learning*, pages 42110–42158. PMLR, 2023.

Masrour Zoghi, Shimon A Whiteson, Maarten De Rijke, and Remi Munos. Relative confidence sampling for efficient on-line ranker evaluation. In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 73–82, 2014.

Masrour Zoghi, Zohar S Karnin, Shimon Whiteson, and Maarten De Rijke. Copeland dueling bandits. *Advances in neural information processing systems*, 28, 2015a.

Masrour Zoghi, Shimon Whiteson, and Maarten de Rijke. Mergerucb: A method for large-scale online ranker evaluation. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, pages 17–26, 2015b.

A Comparison to related framework and hypotheses in dueling bandits

Total order implies a Condorcet winner. Weak stochastic transitivity (WST) is a standard sufficient condition for a latent total order: for any $i, j, k \in [K]$, $\Delta_{i,j} \geq 0$ and $\Delta_{j,k} \geq 0$ imply $\Delta_{i,k} \geq 0$. Thus (up to tie-breaking) preferences are transitive and admit a total order, whose maximal element is a Condorcet winner (CW). CW/optimal-arm identification in such total-order regimes is studied in, e.g., Falahatgar et al. [2017, 2018], which provide worst-case guarantees under WST (notably an $\mathcal{O}(K\varepsilon^{-2}\log(K/\delta))$ bound for (ε, δ)-maxing, i.e., returning an arm I such that $\mathbb{P}(\Delta_{I,i^*} \leq -\varepsilon) \leq \delta)$). Since WST is strictly stronger than merely assuming the existence of a CW, our CW-based bounds directly apply under WST. In particular, our bound for FC-CW1 in Theorem 2.1 improves over the state-of-the art under WST without relying on that assumption. Furthermore, the proofs of Proposition 3.1 and Theorem 3.2 also yield lower bounds for algorithms that are δ-correct over the class of WST instances without ties. Indeed, the specific permutation construction used in Theorem 3.2 preserves every pairwise sign and hence WST. A stronger assumption is strong stochastic transitivity (SST), which adds magnitude constraints (for ordered $i \succ j \succ k$, $\Delta_{i,k} \geq \max\{\Delta_{i,j}, \Delta_{j,k}\}$); in particular, SST implies a CW and ensures that each suboptimal arm attains its largest loss against the CW. Instance-dependent sample complexity under SST (often with mild regularity such as STI) is characterized in Falahatgar et al. [2017], Ren et al. [2020], and our upper bounds recover these guarantees (of the order $H_{\text{cw}}(\delta)$) up to logarithmic factors without even relying on the SST assumption. Finally, parametric random-utility models such as Bradley–Terry–Luce, Plackett–Luce, and Thurstone [Bradley and Terry, 1952, Luce et al., 1959, Plackett, 1975, Thurstone, 2017] are more restrictive than SST and therefore fall within our scope; in particular, our bounds recover existing guarantees for BTL (see, e.g., Ren et al. [2020]). See the survey of Bengs et al. [2021] for a broader overview.

Works on other types of winners. Beyond the CW objective [Komiyama et al., 2015, Ailon et al., 2014, Chen and Frazier, 2017, Saha and Gaillard, 2022, Peköz et al., 2022], alternative notions include the *Borda winner* [Chen et al., 2020, Jamieson et al., 2015] (maximizing average pairwise advantage) and the *Copeland winner* [Zoghi et al., 2015a, Komiyama et al., 2016] (maximizing the number of beaten opponents). Since the Borda and Condorcet winners can differ, Borda-specific guarantees do not transfer to our setting. Copeland winners always exist (possibly non-unique) and generalize the CW; fixed-confidence identification is studied in Zoghi et al. [2015a], but when specialized to CW instances the resulting complexity is $\mathcal{O}\left(\max_{i\in[K]} \sum_{j\neq i} \log(1/\delta)/\Delta_{i,j}^2\right)$, which is looser than our bounds.

B Intermediate results: Adaptive quantile estimation

In this section, we present an algorithm for quantile estimation with a fixed budget of samples, and a high-probability guarantee on the estimation error. This result is of independent interest and will be used as a key building block in the proofs of the main results.

B.1 Quantile Bracketing

Consider a classical K-armed bandit setting, where we are given K arms with means $(\mu_i)_{i\in[K]}$. We assume that the samples from the arms are range-bounded[6] by 1, and without loss of generality that the means are distinct. Denote by $\mu_{(1)} < \mu_{(2)} < \ldots < \mu_{(K)}$ the ordered means of the K arms. For two integers $d \leq u$ in $[K]$, our objective is to find a point in the interval $[\mu_{(d)}, \mu_{(u)}]$.

For this task, a learner is given a fixed budget of T queries, after which the learner outputs a quantity q_T. In this framework, we are targeting an 'adaptive' guarantee in the following sense. Given the budget T as input, we want the output to satisfy

$$\forall \epsilon > 0, \quad \mathbb{P}\left(q_T \notin \left[\mu_{(d)} - \epsilon, \mu_{(u)} + \epsilon\right]\right) \leq \exp\left(-c \cdot r\epsilon^2 \frac{T}{\log(T)}\right) ,$$

where c is a positive universal constant, and r is a nonnegative quantity depending only on d, u and K.

[6]The result can be extended to sub-Gaussian variables

The allocation strategy we develop requires a sufficiently large budget. Specifically, we assume that

$$T \geq \frac{128K}{u-d} \log_2 \left(\frac{128K}{u-d} \right) .$$

When T falls below this threshold, the resulting guarantee, including its logarithmic prefactor, becomes vacuous (the stated upper bound on the error probability exceeds 1). In this regime, we therefore resort to an arbitrary heuristic. Algorithm 3 implements this by explicitly branching between the small-budget case $T < \frac{128K}{u-d} \log_2\left(\frac{128K}{u-d}\right)$ and the regime where the budget is large enough for the analysis to be meaningful.

Solving this problem requires balancing the tasks of locating the arms whose means fall within the desired rank range, and estimating these means accurately. The algorithm runs a multi-step scheme indexed by ℓ. At each level ℓ, it draws a random multi-set $\mathcal{A}_\ell$ of arms large enough to contain, with good probability, representatives of the $[d, u]$ quantile range. Then, it allocates $\Theta(\epsilon_\ell^{-2})$ samples per selected arm so that empirical means are accurate up to ϵ_ℓ.

Then we form three empirical quantiles $\hat{t}_\ell^{(1)}, \hat{t}_\ell^{(2)}, \hat{t}_\ell^{(3)}$ corresponding to ranks slightly below, near the middle of, and slightly above $[d, u]$, yielding a noisy bracket around the target interval. Finally, a Lepski-type stability rule selects the earliest level $\bar{\ell}$ whose middle estimate $\hat{t}_\ell^{(2)}$ remains consistent with the brackets produced at all finer levels (up to the tolerance $2\epsilon_{\ell'}$), and returns $\hat{t}_{\bar{\ell}}^{(2)}$.

Theorem B.1 states that with budget T, the output lies in $[\mu_{(d)} - \epsilon, \mu_{(u)} + \epsilon]$ with high probability for every $\epsilon \in (0, 1)$, and the failure probability decays essentially as $\exp(-\Theta(\epsilon^2 T / \log T))$ (up to the multiplicative factor $40 \log_2(T)$ and the extra $\log(\frac{16K}{u-d})$ term). The quantity

$$r = \min\left\{ \frac{d}{K}, 1 - \frac{u}{K} \right\} \cdot \left(\frac{u-d}{K} \right)^2 ,$$

captures the difficulty of the target rank range as it decreases when the interval is narrower ($u - d$ small) or when it is close to the extremes (small d or large u). When d and u are constant fractions of K, we have $r = \Theta(1)$ and the bound simplifies to $\log(T) \exp(-c\epsilon^2 T / \log T)$.

In the early-return case, we also write $\hat{t}_{\bar{\ell}}^{(2)}$ for the returned value.

Theorem B.1. *Consider Algorithm 3 with inputs (K, d, u, T), such that $u > d$ and $T \geq 2$. Then, the output satisfies for any $\epsilon \in (0, 1)$:*

$$\mathbb{P}\left(\hat{t}_{\bar{\ell}}^{(2)} \notin [\mu_{(d)} - \epsilon, \mu_{(u)} + \epsilon] \right) \leq 40 \log_2(T) \exp\left(-c \cdot r \cdot \frac{\epsilon^2 T}{\log\left(\frac{16K}{u-d}\right) \log_2(T)} \right) , \tag{12}$$

where $r = \min\left\{ \frac{d}{K}, 1 - \frac{u}{K} \right\} \left(\frac{u-d}{K} \right)^2$ is a nonnegative quantity depending only on d, u and K, and c is an absolute numerical constant. When $u = K$, we have $r = 0$, and the bound is vacuous.

Here, we did not try to optimize the constants. Next, we state a corollary that will be used in the proofs of the main theorems.

Corollary B.2. *Consider Algorithm 3 with inputs $(K, \lceil K/8 \rceil, \lceil K/4 \rceil, T)$, where $T \geq 4$. Then, the output satisfies for any $\epsilon \in (0, 1)$:*

$$\mathbb{P}\left(\hat{t}_{\bar{\ell}}^{(2)} \notin [\mu_{(\lceil K/8 \rceil)} - \epsilon, \mu_{(\lceil K/4 \rceil)} + \epsilon] \right) \leq \log(T) \exp\left(-c \cdot \frac{\epsilon^2 T}{\log(T)} \right) ,$$

where c is an absolute numerical constant smaller than 1. The same bound holds for Algorithm 3 with inputs $(N, \lceil (N+1)/8 \rceil, \lceil (N+1)/4 \rceil, t)$, for every $N \geq 1$ and $t \geq 4$: writing $d = \lceil (N+1)/8 \rceil$ and $u = \lceil (N+1)/4 \rceil$, for any $\epsilon \in (0, 1)$,

$$\mathbb{P}\left(\hat{t}_{\bar{\ell}}^{(2)} \notin [\mu_{(d)} - \epsilon, \mu_{(u)} + \epsilon] \right) \leq \log(t) \exp\left(-c_1 \frac{\epsilon^2 t}{\log(t)} \right) ,$$

where $c_1 > 0$ is an absolute numerical constant.

Algorithm 3 Range-Quantile(K, d, u, T)

Input: K number of arms and a budget T, integers $d \leq u$ in $[K]$.
`/* Consider separately the case where we have a small budget */`
if $u = d$ **then**
  Set $T_0 = +\infty$
else
  Set $T_0 = \frac{128K}{u-d} \log_2\left(\frac{128K}{u-d}\right)$
end if
if $T \leq T_0$ **then**
  If $T < K$, return 0. Otherwise, distribute the budget as evenly as possible among the K arms and return the average of the empirical means with ranks $d, \dots, u$
end if
$L = \lfloor \log_2(T/\log_2(T)) \rfloor$ and $\ell_{\min} = \left\lceil \log_2\left(\frac{16K}{u-d}\right) \right\rceil$
`/* Otherwise if we have large budget: */`
for $\ell = \ell_{\min}, \dots, L-1$ **do**
  Let $\epsilon_\ell = 2 \cdot 2^{-(L-\ell)/2}$.
  Sample a set $\mathcal{A}_\ell$ of $\left\lfloor \frac{\epsilon_\ell^2 T}{\log\left(\frac{16K}{u-d}\right)\log_2(T)} \right\rfloor$ arms from $[K]$ with replacement (duplicates are treated as different arms).
  Allocate $\left\lceil \frac{\log\left(\frac{16K}{u-d}\right)}{2\epsilon_\ell^2} \right\rceil$ samples to each arm $a \in \mathcal{A}_\ell$, and compute its empirical mean $\hat{\mu}_a$.
  Rank the empirical means in $\mathcal{A}_\ell$ in increasing order: $\hat{\mu}_{(1)} \leq \cdots \leq \hat{\mu}_{(|\mathcal{A}_\ell|)}$.
  Let:

$$\hat{t}_\ell^{(1)} = \hat{\mu}_{(\lceil \frac{3d+u}{4K}|\mathcal{A}_\ell| \rceil)},\ \hat{t}_\ell^{(2)} = \hat{\mu}_{(\lceil \frac{d+u}{2K}|\mathcal{A}_\ell| \rceil)},\ \hat{t}_\ell^{(3)} = \hat{\mu}_{(\lceil \frac{d+3u}{4K}|\mathcal{A}_\ell| \rceil)}. \tag{11}$$

end for
Let $\bar{\ell} = \min_{\ell \in [\![\ell_{\min}, L-1]\!]} \left\{ \forall \ell' \in \{\ell, \dots, L-1\} : \hat{t}_\ell^{(2)} \in \left[\hat{t}_{\ell'}^{(1)} - 2\epsilon_{\ell'},\ \hat{t}_{\ell'}^{(3)} + 2\epsilon_{\ell'} \right] \right\}$,
Return $\hat{t}_{\bar{\ell}}^{(2)}$.

of Corollary B.2. Suppose $K \geq 5$, then $\lceil K/4 \rceil > \lceil K/8 \rceil$. Let $\epsilon > 0$ and $I = [\mu_{(\lceil K/8 \rceil)} - \epsilon, \mu_{(\lceil K/4 \rceil)} + \epsilon]$. We have

$$\min\left\{ \frac{\lceil K/8 \rceil}{K}, 1 - \frac{\lceil K/4 \rceil}{K} \right\} \geq \min\left\{ \frac{1}{8}, 1 - \frac{K/4+1}{K} \right\} \geq \frac{1}{8}\,. \tag{13}$$

Moreover, using $K \geq 5$ and $K = 8q + r$ where $q \in \mathbb{N}$ and $r \in \{0, \dots, 7\}$, we show that

$$\left(\frac{\lceil K/4 \rceil - \lceil K/8 \rceil}{K} \right)^2 \geq \frac{1}{144}\,. \tag{14}$$

Thus $\log\left(\frac{16K}{u-d}\right) \leq \log(192) < 6$ for $d = \lceil K/8 \rceil$ and $u = \lceil K/4 \rceil$. Applying Theorem B.1 with $u = \lceil K/4 \rceil$, $d = \lceil K/8 \rceil$ and using the bounds (13) and (14) we obtain

$$\begin{aligned}
\mathbb{P}\left(\hat{t}_{\bar{\ell}}^{(2)} \notin I\right) &\leq \min\left\{1,\ 40\log_2(T)\exp\left(-c\,\frac{1}{8}\cdot\frac{1}{144}\cdot\frac{\epsilon^2 T}{6\log_2(T)}\right)\right\} \\
&\leq \log(T)\exp\left(-c'\frac{\epsilon^2 T}{\log(T)}\right)\,,
\end{aligned}$$

where c' is a numerical constant. The last line follows by absorbing all numerical constants into $c' > 0$, using $T \geq 4$.

Suppose now that $K \in \{2, 3, 4\}$, then $\lceil K/4 \rceil = \lceil K/8 \rceil = 1$. In this case, Algorithm 3 allocates at least $\lfloor T/K \rfloor \geq T/(2K) \geq T/8$ samples to each arm and outputs the minimal empirical mean. Let $a \in [K]$ denote the index corresponding to the arm with the smallest true mean, we therefore have

(let $\hat{t}$ denote the output)

$$\begin{aligned}\mathbb{P}\big(\hat{t} \notin [\mu_{(\lceil K/8 \rceil)} - \epsilon, \mu_{(\lceil K/4 \rceil)} + \epsilon]\big) &= \mathbb{P}\big(\hat{t} < \mu_{(1)} - \epsilon\big) + \mathbb{P}\big(\hat{t} > \mu_{(1)} + \epsilon\big) \\ &\leq \mathbb{P}(\hat{\mu}_a > \mu_a + \epsilon) + \sum_{i=1}^{K} \mathbb{P}(\hat{\mu}_i < \mu_i - \epsilon) \ ,\end{aligned}$$

where $\hat{\mu}_i$ denotes the empirical mean of arm i. We conclude using Hoeffding's inequality, with the fact that each arm receives at least $\lfloor T/K \rfloor \geq T/(2K) \geq T/8$ samples that

$$\mathbb{P}\big(\hat{t} \notin [\mu_{(\lceil K/8 \rceil)} - \epsilon, \mu_{(\lceil K/4 \rceil)} + \epsilon]\big) \leq 5 \exp\left(-\epsilon^2 \frac{T}{4}\right) \ ,$$

which corresponds to the result.

For the shifted ranks, if $N \geq 4$, then $\min\{d/N, 1 - u/N\} \geq 1/8$ and $(u-d)/N \geq 1/12$, the latter by (14) with $K = N + 1$. Thus $\log(16N/(u-d)) \leq \log(192) < 6$, and Theorem B.1 gives the same bound after decreasing the numerical constant. If $N \in \{1, 2, 3\}$, then $d = u = 1$; the fallback samples each arm at least $\lfloor t/N \rfloor \geq t/(2N) \geq t/6$ times and returns the smallest empirical mean. The preceding Hoeffding argument gives

$$\mathbb{P}\left(\hat{t}_{\bar{\ell}}^{(2)} \notin [\mu_{(1)} - \epsilon, \mu_{(1)} + \epsilon]\right) \leq 4 \exp(-\epsilon^2 t/3) \ ,$$

which has the claimed form by the same constant-absorption argument. □

B.2 Proof of Theorem B.1

Since the right-hand side of (12) does not depend on the values of $\mu_{(1)}, \ldots, \mu_{(K)}$, it suffices to treat the strictly ordered case where $\mu_{(1)} < \mu_{(2)} < \cdots < \mu_{(K)}$; the case where some values are perhaps identical follows by a continuity argument.

Proof. If $T \leq \frac{128K}{u-d} \log_2\left(\frac{128K}{u-d}\right)$ the bound is vacuous, including when the fallback returns 0: in this range $rT/\log_2(T) \leq 128$, so the right-hand side of (12) exceeds 1 for sufficiently small c. Assume that $T > \frac{128K}{u-d} \log_2\left(\frac{128K}{u-d}\right)$ so that $\log_2(\log_2(T)) \geq 1$ and $\ell_{\min} \leq L - 1$. We introduce the following additional notation, for each $\ell \in \{\ell_{\min}, \ldots, L-1\}$ where $\ell_{\min} = \left\lceil \log_2\left(\frac{16K}{u-d}\right) \right\rceil$, let

$$N_\ell := |\mathcal{A}_\ell| = \left\lfloor \frac{\epsilon_\ell^2 T}{\log\left(\frac{16K}{u-d}\right) \log_2(T)} \right\rfloor \qquad \text{and} \qquad T_\ell := \left\lceil \frac{\log\left(\frac{16K}{u-d}\right)}{2\epsilon_\ell^2} \right\rceil \ , \tag{15}$$

and let

$$r_0 := \frac{d}{K} \quad , \quad r_1 := \frac{3d+u}{4K} \quad , \quad r_2 := \frac{d+u}{2K} \quad , \quad r_3 := \frac{d+3u}{4K} \quad \text{and} \quad r_4 := \frac{u}{K} \ .$$

The proof follows the steps below

- We start by a sanity check, verifying that the total number of queries made by Algorithm 3 is at most T, and that $\bar{\ell}$ exists.
- Next, we show an intermediary result about the quantities $\hat{t}_\ell^{(i)}$ for $i \in \{1, 2, 3\}$ in the form of an upper-bound on

$$\mathbb{P}\left(\hat{t}_\ell^{(i)} \notin \left[\mu_{\left(\left\lceil \frac{r_{i-1}+r_i}{2} \cdot K \right\rceil\right)} - \epsilon_\ell, \ \mu_{\left(\left\lceil \frac{r_i+r_{i+1}}{2} \cdot K \right\rceil\right)} + \epsilon_\ell\right]\right) \ .$$

- Finally, we build on the obtained intermediary result to prove that the way $\bar{\ell}$ is defined allows to have the stated guarantees.

Sanity checks: Recall the expressions $L = \lfloor \log_2(T/\log_2(T)) \rfloor$, $\epsilon_\ell = 2.\ 2^{-(L-\ell)/2}$ and $|\mathcal{A}_\ell| = \left\lfloor \frac{\epsilon_\ell^2 T}{\log\left(\frac{16K}{u-d}\right)\log_2(T)} \right\rfloor$. Algorithm 3 comprises $L - \ell_{\min}$ iterations, for each iteration $\ell \in \{\ell_{\min}, \dots, L-1\}$ it makes $|\mathcal{A}_\ell| T_\ell$ queries. Thus, the total number of queries is

$$\begin{aligned}
\sum_{\ell=\ell_{\min}}^{L-1} |\mathcal{A}_\ell| \cdot \left\lceil \frac{\log\left(\frac{16K}{u-d}\right)}{2\epsilon_\ell^2} \right\rceil &= \sum_{\ell=\ell_{\min}}^{L-1} \left\lfloor \frac{\epsilon_\ell^2 T}{\log\left(\frac{16K}{u-d}\right)\log_2(T)} \right\rfloor \cdot \left\lceil \frac{\log\left(\frac{16K}{u-d}\right)}{2\epsilon_\ell^2} \right\rceil \\
&\le \sum_{\ell=\ell_{\min}}^{L-1} \frac{\epsilon_\ell^2 T}{\log\left(\frac{16K}{u-d}\right)\log_2(T)} \cdot \left(\frac{\log\left(\frac{16K}{u-d}\right)}{2\epsilon_\ell^2} + 1 \right) \\
&= \sum_{\ell=\ell_{\min}}^{L-1} \frac{T}{2\log_2(T)} + \frac{\epsilon_\ell^2 T}{\log\left(\frac{16K}{u-d}\right)\log_2(T)} \\
&< \frac{T}{2} + \frac{T}{\log\left(\frac{16K}{u-d}\right)\log_2(T)} \sum_{\ell=\ell_{\min}}^{L-1} \epsilon_\ell^2 \\
&\le \frac{T}{2} + \frac{4T}{\log\left(\frac{16K}{u-d}\right)\log_2(T)} \le T\,,
\end{aligned}$$

where we used in the last line the threshold condition on T, giving

$$\log(16K/(u-d))\log_2(T) \ge 8\,.$$

For the definition of the quantity $\bar{\ell}$, note that the set over which the minimum is taken is not empty, since it always contains $L-1$.

Step 2: Per-level quantile control.

In this step we will prove that for every level $\ell \in \{\ell_{\min}, \dots, L-1\}$ and every $i \in \{1,2,3\}$

$$\mathbb{P}\Big(\hat{t}_\ell^{(i)} \notin C_{\ell,i}\Big) \le p_\ell\,, \tag{16}$$

where we define (for each ℓ and $i \in \{1,2,3\}$)

$$\begin{aligned}
C_{\ell,i} &\coloneqq \left[\mu_{\left(\left\lceil \frac{r_{i-1}+r_i}{2} K \right\rceil\right)} - \epsilon_\ell,\ \mu_{\left(\left\lceil \frac{r_i+r_{i+1}}{2} K \right\rceil\right)} + \epsilon_\ell \right], \\
\kappa_{d,u} &\coloneqq \min\left\{ \frac{d}{K}, 1 - \frac{u}{K} \right\} \left(\frac{u-d}{60K} \right)^2, \qquad p_\ell \coloneqq 4\exp(-\kappa_{d,u} N_\ell)\,.
\end{aligned}$$

Throughout this step, fix $\ell \in \{\ell_{\min}, \dots, L-1\}$ and $i \in \{1,2,3\}$. Define the two (random-sample) ranks

$$r_- \coloneqq \left\lceil \frac{r_{i-1} + 2r_i}{3} N_\ell \right\rceil, \qquad r_+ \coloneqq \left\lceil \frac{2r_i + r_{i+1}}{3} N_\ell \right\rceil,$$

and define the two (population) bracket points

$$m_- \coloneqq \mu_{\left(\left\lceil \frac{r_{i-1}+r_i}{2} K \right\rceil\right)}, \qquad m_+ \coloneqq \mu_{\left(\left\lceil \frac{r_i+r_{i+1}}{2} K \right\rceil\right)}\,.$$

Let $\gamma_1, \dots, \gamma_{N_\ell}$ be the (random) true means of the sampled multiset $\mathcal{A}_\ell$, and let $\gamma_{(1)} \le \dots \le \gamma_{(N_\ell)}$ be their order statistics (ties broken arbitrarily).

Next, introduce the event $E^{(i)}$ given by

$$E^{(i)} \coloneqq \left\{ \gamma_{(r_-)} < m_- \right\} \cup \left\{ \gamma_{(r_+)} > m_+ \right\}.$$

Then, by a union bound,

$$\mathbb{P}\Big(\hat{t}_\ell^{(i)} \notin C_{\ell,i}\Big) \le \underbrace{\mathbb{P}\Big(E^{(i)}\Big)}_{\text{Term 1}} + \underbrace{\mathbb{P}\Big(\hat{t}_\ell^{(i)} \notin C_{\ell,i} \text{ and } \neg E^{(i)}\Big)}_{\text{Term 2}}. \tag{17}$$

Next we will control the probability of $E^{(i)}$ (Term 1). Define the counts

$$M_1 \coloneqq |\{j \in [N_\ell] : \gamma_j < m_-\}|, \qquad M_2 \coloneqq |\{j \in [N_\ell] : \gamma_j > m_+\}| \,.$$

Since $\mathcal{A}_\ell$ is obtained by sampling arms i.i.d. with replacement from $[K]$, these are binomials:

$$M_1 \sim \mathrm{Bin}\left(N_\ell,\ \frac{\left\lceil \frac{r_{i-1}+r_i}{2} K \right\rceil - 1}{K}\right), \qquad M_2 \sim \mathrm{Bin}\left(N_\ell,\ 1 - \frac{\left\lceil \frac{r_i+r_{i+1}}{2} K \right\rceil}{K}\right).$$

Moreover, by definition of order statistics we have

$$\{\gamma_{(r_-)} < m_-\} \subseteq \{M_1 \geq r_-\}, \qquad \{\gamma_{(r_+)} > m_+\} \subseteq \{M_2 \geq N_\ell - r_+ + 1\} \,.$$

Therefore,

$$\mathbb{P}\Big(E^{(i)}\Big) \leq \mathbb{P}(M_1 \geq r_-) \ + \ \mathbb{P}(M_2 \geq N_\ell - r_+ + 1) \,.$$

Let us bound the two terms in the upper bound above using binomial tail bounds. Using $\lceil x \rceil - 1 \leq x$ and $\lceil x \rceil \geq x$, we have the parameter bounds

$$\frac{\left\lceil \frac{r_{i-1}+r_i}{2} K \right\rceil - 1}{K} \ \leq \ \frac{r_{i-1}+r_i}{2}, \qquad 1 - \frac{\left\lceil \frac{r_i+r_{i+1}}{2} K \right\rceil}{K} \ \leq \ 1 - \frac{r_i + r_{i+1}}{2} \,.$$

Hence, M_1 and M_2 are stochastically dominated by $\mathrm{Bin}(N_\ell, \frac{r_{i-1}+r_i}{2})$ and $\mathrm{Bin}(N_\ell, 1 - \frac{r_i+r_{i+1}}{2})$, respectively. Also, by construction we have

$$\frac{r_{i-1}+2r_i}{3} - \frac{r_{i-1}+r_i}{2} = \frac{r_i - r_{i-1}}{6}, \qquad \Big(1 - \frac{2r_i + r_{i+1}}{3}\Big) - \Big(1 - \frac{r_i + r_{i+1}}{2}\Big) = \frac{r_{i+1} - r_i}{6} \,.$$

Applying Hoeffding's inequality to these dominating binomials yields

$$\mathbb{P}(M_1 \geq r_-) \leq \exp\left(-2N_\ell \Big(\frac{r_i - r_{i-1}}{6}\Big)^2\right),$$
$$\mathbb{P}(M_2 \geq N_\ell - r_+ + 1) \leq \exp\left(-2N_\ell \Big(\frac{r_{i+1} - r_i}{6}\Big)^2\right).$$

Since for each $j \in \{1,2,3,4\}$ one has $r_j - r_{j-1} \geq \frac{u-d}{4K}$, we obtain

$$\text{Term 1} = \mathbb{P}\Big(E^{(i)}\Big) \leq 2\exp\left(-\frac{N_\ell (u-d)^2}{288K^2}\right). \tag{18}$$

Next, let us upper bound Term 2 in (17). On $\neg E^{(i)}$ we have $\gamma_{(r_-)} \geq m_-$ and $\gamma_{(r_+)} \leq m_+$, hence

$$[m_- - \epsilon_\ell,\ m_+ + \epsilon_\ell] \supseteq [\gamma_{(r_-)} - \epsilon_\ell,\ \gamma_{(r_+)} + \epsilon_\ell].$$

Therefore,

$$\begin{aligned}\text{Term 2} &= \mathbb{P}\Big(\hat{t}_\ell^{(i)} \notin [m_- - \epsilon_\ell, m_+ + \epsilon_\ell] \text{ and } \neg E^{(i)}\Big) \\ &\leq \mathbb{P}\Big(\hat{t}_\ell^{(i)} \notin [\gamma_{(r_-)} - \epsilon_\ell,\ \gamma_{(r_+)} + \epsilon_\ell]\Big) \,.\end{aligned}$$

By Lemma B.4, this implies

$$\text{Term 2} \leq 2\exp(-\kappa_{d,u} N_\ell) \,. \tag{19}$$

Combining (17), (18) and (19), and using that $\kappa_{d,u} \leq (u-d)^2/(288K^2)$, we obtain

$$\mathbb{P}\Big(\hat{t}_\ell^{(i)} \notin C_{\ell,i}\Big) \leq 4\exp(-\kappa_{d,u} N_\ell) = p_\ell \,,$$

which is exactly (16).

Step 3: Conclusion.

If $\epsilon < 3\epsilon_{\ell_{\min}}$, then the upper bound of the theorem is greater than 1 and the bound is vacuous. Assume that $\epsilon \geq 3\epsilon_{\ell_{\min}}$. Let $\ell^\star$ be the largest level such that

$$3\epsilon_{\ell^\star} \leq \epsilon \ .$$

This implies in particular, since $\ell^\star + 1$ violates the condition above, that $\epsilon < 3\epsilon_{\ell^\star+1} = 3\sqrt{2}\epsilon_{\ell^\star}$, therefore

$$\epsilon_{\ell^\star} \geq \frac{\epsilon}{3\sqrt{2}} \ . \tag{20}$$

Next, we will prove that for any $\ell \in \{\ell_{\min}, \dots, L-1\}$, we have

$$\mathbb{P}(\hat{t}^{(2)}_{\bar{\ell}} \notin [\mu_{(d)} - 3\epsilon_\ell,\ \mu_{(u)} + 3\epsilon_\ell]) \leq 2(4L+1)p_\ell \ .$$

Let $\ell \in \{\ell_{\min}, \dots, L-1\}$, recall that by definition of $\bar{\ell}$, for $\ell' \geqslant \bar{\ell}$, one has $\hat{t}^{(2)}_{\bar{\ell}} \leqslant \hat{t}^{(3)}_{\ell'} + 2\epsilon_{\ell'}$. Then, it holds that

$$\mathbb{P}\left(\hat{t}^{(2)}_{\bar{\ell}} > \mu_{(u)} + 3\epsilon_\ell\right) \leq \mathbb{P}(\bar{\ell} > \ell) + \mathbb{P}(\hat{t}^{(3)}_{\ell} > \mu_{(u)} + \epsilon_\ell) \leq 4Lp_\ell + p_\ell \ ,$$

where we use Lemma B.3, which ensures that for every $\ell \in \{\ell_{\min}, \dots, L-1\}$ we have $\mathbb{P}(\bar{\ell} > \ell) \leq 4Lp_\ell$, and we use the Bound 16 from step 2 with $i = 3$. The second bound

$$\mathbb{P}\left(\hat{t}^{(2)}_{\bar{\ell}} < \mu_{(d)} - 3\epsilon_\ell\right) \leq (4L+1)p_\ell \ ,$$

is proven using the same arguments (in particular Bound 16 with $i = 1$).

Applying this bound to $\ell^\star$, using $3\epsilon_{\ell^\star} \leqslant \epsilon$, we have

$$\begin{aligned}\mathbb{P}\left(\hat{t}^{(2)}_{\bar{\ell}} \notin [\mu_{(d)} - \epsilon,\ \mu_{(u)} + \epsilon]\right) &\leq \mathbb{P}\left(\hat{t}^{(2)}_{\bar{\ell}} \notin [\mu_{(d)} - 3\epsilon_{\ell^\star},\ \mu_{(u)} + 3\epsilon_{\ell^\star}]\right) \\ &\leq 2(4L+1)p_{\ell^\star} \ .\end{aligned}$$

Next, in order to upper bound $p_{\ell^\star}$ we use the following lower bound on $N_{\ell^\star}$

$$\begin{aligned}N_{\ell^\star} = \left\lfloor \frac{\epsilon^2_{\ell^\star} T}{\log\left(\frac{16K}{u-d}\right)\log_2(T)} \right\rfloor &\geq \frac{1}{2}\frac{\epsilon^2_{\ell^\star} T}{\log\left(\frac{16K}{u-d}\right)\log_2(T)} \\ &\geq \frac{\epsilon^2 T}{36\log\left(\frac{16K}{u-d}\right)\log_2(T)} \ ,\end{aligned}$$

where we use the fact that from the assumption on the budget T, $N_{\ell^\star} \geqslant 2$, and $\epsilon_{\ell^\star} \geqslant \epsilon/3\sqrt{2}$ (see (20)). Therefore, using the definition of p_ℓ,

$$\begin{aligned}p_{\ell^\star} &= 4\exp(-\kappa_{d,u} \cdot N_{\ell^\star}) \\ &\leq 4\exp\left(-\min\left\{\frac{d}{K},\ 1-\frac{u}{K}\right\}\left(\frac{u-d}{60K}\right)^2 \cdot \frac{\epsilon^2 T}{36\log\left(\frac{16K}{u-d}\right)\log_2(T)}\right) \\ &= 4\exp\left(-\frac{cr\epsilon^2 T}{\log\left(\frac{16K}{u-d}\right)\log_2(T)}\right) \ ,\end{aligned}$$

where $c > 0$ is an absolute constant, and $r = \min\left\{\frac{d}{K},\ 1-\frac{u}{K}\right\}\left(\frac{u-d}{K}\right)^2$. Finally, given that $L \leq \log_2(T)$ and $2(4L+1)\cdot 4 \leq 40\log_2(T)$ for $T \geq 2$,

$$\mathbb{P}\left(\hat{t}^{(2)}_{\bar{\ell}} \notin [\mu_{(d)} - \epsilon,\ \mu_{(u)} + \epsilon]\right) \leq 40\log_2(T)\exp\left(-\frac{cr\epsilon^2 T}{\log\left(\frac{16K}{u-d}\right)\log_2(T)}\right) \ ,$$

which is the desired bound. □

Below are two technical lemmas deferred here to avoid cluttering the proof above.

Lemma B.3. *For every $\ell \in \{\ell_{\min}, \dots, L-1\}$, we have $\mathbb{P}(\bar{\ell} > \ell) \leq 4Lp_\ell$.*

Proof. Suppose that $\bar{\ell} > \ell$, then using the definition of $\bar{\ell}$ we have necessarily that there exists $\ell' \geqslant \ell$ such that $\hat{t}_\ell^{(2)} \notin I_{\ell'}$, with $I_{\ell'} = \left[\hat{t}_{\ell'}^{(1)} - 2\epsilon_{\ell'},\ \hat{t}_{\ell'}^{(3)} + 2\epsilon_{\ell'}\right]$. Therefore,

$$\mathbb{P}(\bar{\ell} > \ell) \leq \sum_{\ell' \geqslant \ell} \mathbb{P}\left(\hat{t}_\ell^{(2)} < \hat{t}_{\ell'}^{(1)} - 2\epsilon_{\ell'}\right) + \sum_{\ell' \geqslant \ell} \mathbb{P}\left(\hat{t}_\ell^{(2)} > \hat{t}_{\ell'}^{(3)} + 2\epsilon_{\ell'}\right) .$$

Let

$$m_- \coloneqq \mu_{(\lceil \frac{r_1+r_2}{2} K \rceil)}, \quad m_+ \coloneqq \mu_{(\lceil \frac{r_2+r_3}{2} K \rceil)} .$$

Let $\ell' \geqslant \ell$, the event $\hat{t}_\ell^{(2)} < \hat{t}_{\ell'}^{(1)} - 2\epsilon_{\ell'}$ implies $\hat{t}_{\ell'}^{(1)} > m_- + \epsilon_{\ell'}$ or $\hat{t}_\ell^{(2)} < m_- - \epsilon_{\ell'}$. Since we have $\epsilon_{\ell'} \geq \epsilon_\ell$ for $\ell' \geqslant \ell$, Bound 16 yields $\mathbb{P}(\hat{t}_{\ell'}^{(1)} > m_- + \epsilon_{\ell'}) \leq p_{\ell'}$ and

$$\begin{aligned}\mathbb{P}(\hat{t}_\ell^{(2)} < m_- - \epsilon_{\ell'}) &\leq \mathbb{P}(\hat{t}_\ell^{(2)} < m_- - \epsilon_\ell) \\ &\leq p_\ell .\end{aligned}$$

Therefore,

$$\mathbb{P}(\hat{t}_\ell^{(2)} < \hat{t}_{\ell'}^{(1)} - 2\epsilon_{\ell'}) \leq p_\ell + p_{\ell'} .$$

Similarly,

$$\mathbb{P}(\hat{t}_\ell^{(2)} > \hat{t}_{\ell'}^{(3)} + 2\epsilon_{\ell'}) \leq p_\ell + p_{\ell'} .$$

Therefore, for every $\ell' \geqslant \ell$,

$$\mathbb{P}(\hat{t}_\ell^{(2)} \notin I_{\ell'}) \leq 2p_\ell + 2p_{\ell'} .$$

Summing over $\ell' \geqslant \ell$,

$$\mathbb{P}(\bar{\ell} > \ell) \leq \sum_{\ell' \geqslant \ell} (2p_\ell + 2p_{\ell'}) \leq 2Lp_\ell + 2\sum_{\ell' \geqslant \ell} p_{\ell'} .$$

Since $N_{\ell'}$ increases with ℓ' (thus $p_{\ell'}$ is decreasing), we have $\sum_{\ell' \geqslant \ell} p_{\ell'} \leq Lp_\ell$, which gives $\mathbb{P}(\bar{\ell} > \ell) \leq 4Lp_\ell$. □

Lemma B.4. *Let $\ell \in \{\ell_{\min}, \dots, L-1\}$ and consider the notation introduced in the proof of Theorem B.1. For each $i \in \{1,2,3\}$, define the two indices*

$$r_- \coloneqq \left\lceil \frac{r_{i-1} + 2r_i}{3} N_\ell \right\rceil \quad \textit{and} \quad r_+ \coloneqq \left\lceil \frac{2r_i + r_{i+1}}{3} N_\ell \right\rceil .$$

Then

$$\mathbb{P}\left(\hat{t}_\ell^{(i)} \notin \left[\gamma_{(r_-)} - \epsilon_\ell, \gamma_{(r_+)} + \epsilon_\ell\right] \mid \mathcal{A}_\ell\right) \leq 2\exp\left(-\kappa_{d,u} \cdot N_\ell\right) .$$

Proof. Fix $\ell \in \{\ell_{\min}, \dots, L-1\}$ and $i \in \{1,2,3\}$. Define $q = \lceil r_i N_\ell \rceil$. Given that we have $T_\ell = \lceil \log(16K/(u-d))/(2\epsilon_\ell^2) \rceil$, let

$$\begin{aligned}\delta_\ell &\coloneqq \exp\left(-2T_\ell \epsilon_\ell^2\right) \\ &\leq \exp\left(-\log(16K/(u-d))\right) = \frac{u-d}{16K} . \end{aligned} \tag{21}$$

Moreover, for every arm j, given that samples are 1-range bounded and $\hat{\mu}_j$ is computed with T_ℓ samples, Hoeffding's inequality gives

$$\mathbb{P}(\hat{\mu}_j \leq \gamma_j - \epsilon_\ell \mid \mathcal{A}_\ell) \leq \delta_\ell, \quad \mathbb{P}(\hat{\mu}_j \geq \gamma_j + \epsilon_\ell \mid \mathcal{A}_\ell) \leq \delta_\ell . \tag{22}$$

Next, we prove the lower tail of the claimed bound. Recall that $\hat{t}_\ell^{(i)} = \hat{\mu}_{(q)}$ and $r_- \coloneqq \left\lceil \frac{r_{i-1}+2r_i}{3} N_\ell \right\rceil$ and define the event

$$\mathcal{E}_- \coloneqq \{\hat{\mu}_{(q)} < \gamma_{(r_-)} - \epsilon_\ell\} .$$

If $\mathcal{E}_-$ occurs, then at least q empirical means are smaller than $\gamma_{(r_-)} - \epsilon_\ell$ (with $q \geqslant r_-$). Given that $(\gamma_{(k)})_k$ are in non-decreasing order, we conclude that among the set

$$\mathcal{G}_- \coloneqq \{j : \gamma_j \geq \gamma_{(r_-)}\} ,$$

then at least $q - (r_- - 1)$ elements must satisfy $\hat{\mu}_j < \gamma_{(r_-)} - \epsilon_\ell$. For each element $j \in \mathcal{G}_-$ we have $\gamma_j \geq \gamma_{(r_-)}$, hence $\{\hat{\mu}_j < \gamma_{(r_-)} - \epsilon_\ell\} \subseteq \{\hat{\mu}_j < \gamma_j - \epsilon_\ell\}$, using (21) and (22) we have that, conditionally on $\mathcal{A}_\ell$, each such downward deviation has probability at most $(u-d)/(16K)$. Therefore, we conclude that

$$\mathbb{P}(\mathcal{E}_- \mid \mathcal{A}_\ell) \leq \mathbb{P}\left(\text{Bin}(|\mathcal{G}_-|, (u-d)/(16K)) \geq q - (r_- - 1)\right) .$$

Using the definitions of q, r_- and r_i, we have

$$\begin{aligned} q - (r_- - 1) &= \lceil r_i N_\ell \rceil - \left(\left\lceil \frac{r_{i-1} + 2r_i}{3} N_\ell \right\rceil - 1 \right) \\ &\geq r_i N_\ell - \frac{r_{i-1} + 2r_i}{3} N_\ell \\ &= \frac{r_i - r_{i-1}}{3} N_\ell = \frac{u-d}{12K} N_\ell \\ &\geq \frac{u-d}{12K} |\mathcal{G}_-| . \end{aligned}$$

Applying Hoeffding's inequality for binomials yields

$$\begin{aligned} \mathbb{P}(\mathcal{E}_- \mid \mathcal{A}_\ell) &\leq \exp\left(-2|\mathcal{G}_-| \left(\frac{u-d}{12K} - \frac{u-d}{16K} \right)^2 \right) \\ &\leq \exp\left(-2|\mathcal{G}_-| \left(\frac{u-d}{48K} \right)^2 \right) . \end{aligned}$$

Finally, for $i \in \{1,2,3\}$ we have $\frac{r_{i-1}+2r_i}{3} \leq \frac{u}{K}$, hence $|\mathcal{G}_-| \geq N_\ell - r_- + 1 \geq (1 - u/K) N_\ell$. Therefore,

$$\begin{aligned} \mathbb{P}(\mathcal{E}_- \mid \mathcal{A}_\ell) &\leq \exp\left(-2 \left(1 - \frac{u}{K}\right) N_\ell \left(\frac{u-d}{48K} \right)^2 \right) \\ &\leq \exp\left(-\kappa_{d,u} \cdot N_\ell \right) . \end{aligned} \tag{23}$$

Let us show the upper tail of the claimed bound. We follow similar steps as in the lower tail proof. Consider $r_+ \coloneqq \left\lceil \frac{2r_i + r_{i+1}}{3} N_\ell \right\rceil$ and define

$$\mathcal{E}_+ \coloneqq \{\hat{\mu}_{(q)} > \gamma_{(r_+)} + \epsilon_\ell\} .$$

If $\mathcal{E}_+$ occurs, then at least $N_\ell - q + 1$ empirical means exceed $\gamma_{(r_+)} + \epsilon_\ell$. At most $N_\ell - r_+$ arms can have true mean larger than $\gamma_{(r_+)}$, therefore at least $(N_\ell - q + 1) - (N_\ell - r_+) = r_+ - q + 1$ arms from the set

$$\mathcal{G}_+ \coloneqq \{j : \gamma_j \leq \gamma_{(r_+)}\}$$

must satisfy $\hat{\mu}_j > \gamma_{(r_+)} + \epsilon_\ell$. For each $j \in \mathcal{G}_+$ we have $\gamma_j \leq \gamma_{(r_+)}$, therefore $\{\hat{\mu}_j > \gamma_{(r_+)} + \epsilon_\ell\} \subseteq \{\hat{\mu}_j \geq \gamma_j + \epsilon_\ell\}$, using (21) and (22) we have that, conditionally on $\mathcal{A}_\ell$, each such upward deviation has probability at most $(u-d)/(16K)$ conditionally on $\mathcal{A}_\ell$. Thus, conditionally on $\mathcal{A}_\ell$, we have

$$\mathbb{P}(\mathcal{E}_+ \mid \mathcal{A}_\ell) \leq \mathbb{P}\left(\text{Bin}(|\mathcal{G}_+|, (u-d)/(16K)) \geq r_+ - q + 1\right) .$$

Also, we have

$$\begin{aligned} r_+ - q + 1 &\geq \frac{2r_i + r_{i+1}}{3} N_\ell - r_i N_\ell \\ &= \frac{r_{i+1} - r_i}{3} N_\ell = \frac{u-d}{12K} N_\ell . \end{aligned}$$

Therefore, using the binomial Hoeffding bound we obtain

$$\begin{aligned} \mathbb{P}(\mathcal{E}_+ \mid \mathcal{A}_\ell) &\leq \exp\left(-2|\mathcal{G}_+| \left(\frac{u-d}{48K} \right)^2 \right) \\ &\leq \exp\left(-2r_+ \left(\frac{u-d}{48K} \right)^2 \right) . \end{aligned}$$

Moreover, for $i \in \{1,2,3\}$ we have $\frac{2r_i+r_{i+1}}{3} \geq \frac{d}{K}$, so $r_+ \geq \frac{d}{K}N_\ell$. Hence,

$$\begin{aligned}\mathbb{P}(\mathcal{E}_+ \mid \mathcal{A}_\ell) &\leq \exp\left(-2\frac{d}{K}N_\ell\left(\frac{u-d}{48K}\right)^2\right) \\ &\leq \exp\left(-\kappa_{d,u}\cdot N_\ell\right) . \end{aligned} \tag{24}$$

The conclusion follows by combining (23), and (24) which leads to the bound

$$\mathbb{P}(\hat{\mu}_{(q)} \notin [\gamma_{(r_-)} - \epsilon_\ell,\ \gamma_{(r_+)} + \epsilon_\ell]) \leq 2\exp\left(-\kappa_{d,u}N_\ell\right) .$$

□

B.3 A result on Bracketed Sequential Halving based on Zhao et al. [2023]

Consider a K-armed bandit problem with Bernoulli reward with unknown means $\mu_1, \ldots, \mu_K$. As in the previous subsection, we write $\mu_{(1)} \leq \ldots \leq \mu_{(K)}$ for its ordered values. Sequential halving is a classical elimination scheme for pure-exploration problems [Karnin et al., 2013]. It proceeds in at most $\lceil \log_2 K \rceil$ phases. Starting from the full set of K candidate arms, each phase spends approximately $\lfloor T/\log_2 K \rfloor$ samples by distributing them uniformly across the surviving arms, then ranks arms by their empirical means and discards the top half. Since our goal is to identify arms with the smallest mean, we retain the bottom-ranked half after each phase. Bracketed SH uses the bracket-opening and sampling schedule of Zhao et al. [2023], with the following recommendation convention: a bracket is eligible only after its opening phase ends (when the next bracket opens), and only if it has completed an SH run. Among eligible brackets we return the representative with the smallest final-stage empirical mean; if there is none, we return an arbitrary arm. With a single candidate arm, we simply return that arm. The following bound combines the SH guarantee of Zhao et al. [2023] with the bracketing argument.

Theorem B.5. *Consider Bracketed SH as defined above with $K \geq 2$ and an integer budget $T \geq 1$. The output I_T satisfies for any $\epsilon \in (0,1)$ and $m \in [K]$:*

$$\mathbb{P}\big(\mu_{I_T} \geq \mu_{(m)} + \epsilon\big) \leq \min\left\{1,\ c_1 \log(2eKT)\exp\left[-\frac{c_2 m\epsilon^2 T}{K\log(eT)\log^3(2K)}\right]\right\},$$

where $c_1, c_2 > 0$ are universal constants.

Proof. Write $x = m\epsilon^2 T/[K\log(eT)\log^3(2K)]$. By increasing c_1 and decreasing c_2, the claim is vacuous when x is smaller than a sufficiently large universal constant; this covers small budgets and the case with no eligible bracket. We henceforth assume that x exceeds this constant. Let $a > 0$ be a sufficiently small universal constant and set

$$z = \frac{a\epsilon^2 T}{\log(eT)\log^2(2m)\log(2K)} .$$

The b-th bracket of Bracketed SH contains $\min\{K, 2^b\}$ arms and its opening phase ends when bracket $b+1$ opens. Choose a bracket of size $n = K$ if $z \geq K$, and otherwise a bracket of size $n \in [z/2, z]$. Such a bracket exists since the sizes double until they reach K; the choice of a ensures that its opening phase has ended by time T. Indeed, the phase of a bracket of size 2^b ends by time $b2^b$, while $z\log_2(2z) \leq CaT$ for an absolute constant C. The full-K bracket is likewise mature when $z \geq K$. If $B \geq 3$ brackets are open at time T, then $T < B2^B$ and the last completed opening phase gives each of the $B-1$ mature brackets at least $2^{B-2} > T/(4B) \geq cT/\log(eT)$ samples, since $B \leq C\log(eT)$. Since $n\log_2 n \leq CaT/\log(eT)$, our chosen bracket has completed an SH run when a is small enough. In the doubling schedule of SH runs, the last completed run in every eligible bracket uses at least one quarter of that bracket's samples.

Let X be the number of positions in the chosen bracket occupied by globally best-m arms. If $n < K$, then $X \sim \mathrm{Bin}(n, m/K)$. Writing $v = nm/K$, our choice of n gives $v \geq ax/2$, and the binomial lower-tail bound gives

$$\mathbb{P}(X < v/2) \leq \exp(-v/8) \leq \exp(-cx) .$$

On the complementary event, $q = \lfloor v/2 \rfloor \geq v/4$ of the bracket's arms have means at most $\mu_{(m)}$; the initial reduction to nonvacuous x ensures $v \geq 4$. If $n = K$, the bracket contains every arm and we instead take $q = m$, with no binomial error.

Conditional on the bracket, apply the SH bound of Theorem 4 in Zhao et al. [2023] to the last completed SH run, with the reward signs reversed and accuracy $\epsilon/4$. Its budget is at least $cT/\log(eT)$, while $q \le m$ and $n \le z$ in the non-full case. Since the q-th mean in the bracket is at most $\mu_{(m)}$, choosing a small enough absorbs the offset in that bound and yields

$$\mathbb{P}\big(\mu_{J_*} \ge \mu_{(m)} + \epsilon/2,\ X \ge q\big) \le C\log(2K)\exp(-cx)\ ,$$

where J_* is the chosen bracket's representative. In the full-K case the same conclusion follows directly from its run budget and $q = m$. Here and below $C, c > 0$ denote universal constants, which may change from line to line.

It remains to compare the representatives. In every completed SH run, the final-stage candidate set has at most two arms and is determined before the fresh final-stage samples are drawn. The representative's final-stage empirical mean may therefore be controlled by a union bound over these two candidates, without treating the selected arm as fixed. Every eligible completed bracket has a last-run budget at least $cT/\log(eT)$, so each final-stage candidate receives at least $cT/[\log(eT)\log(2K)]$ samples. There are at most $C\log(eT)$ eligible brackets. Hoeffding's inequality and a union bound thus give

$$\mathbb{P}(\exists \text{ eligible completed bracket } b:\ |\hat{\mu}_{J_b} - \mu_{J_b}| \ge \epsilon/4) \le C\log(eT)\exp\left(-\frac{c\epsilon^2 T}{\log(eT)\log(2K)}\right)\ .$$

Outside these events, the chosen bracket has true mean below $\mu_{(m)} + \epsilon/2$, and its empirical mean is below $\mu_{(m)} + 3\epsilon/4$. The recommended arm has no larger empirical mean, so its true mean is below $\mu_{(m)} + \epsilon$. Combining the three bounds, using $m \le K$ and $\log^2(2m) \le \log^2(2K)$, and absorbing their constants into c_1, c_2 proves the claim. □

C Guarantees on FB CWI procedure (Algorithm 1) in the fixed budget setting

In order to structure the proofs, in this section we present guarantees about the output of Algorithm 1 when fed with input (δ, T). More precisely the output being $(\phi_1 \vee \phi_2, I)$, here we provide upper bounds on the probability of misidentification error for the arm candidate I. This corresponds to the typical kind of guarantees encountered in context of best arm identification in the fixed budget framework. In turn, we apply these results in Section D to prove the guarantees presented in Theorem 2.1.

C.1 First Upper Bound

Theorem C.1. *The output of* FB CWI *(Algorithm 1) with input $T \ge 2$ satisfies:*

$$\mathbb{P}\left(\psi_T \ne i^*\right) \le 27K\log(K)\log(T)\cdot\exp\left(-c\cdot\frac{T}{\log(T)\log(K)H_{cw}}\right),$$

where c is a numerical constant, and we recall that H_{cw} is defined by

$$H_{cw} = \sum_{i\ne i^*}\frac{1}{\Delta_{i^*,i}^2}\ ,$$

if $\Delta_{i^*,i} > 0$ for all $i \in [K]\setminus\{i^*\}$ and $H_{cw} = +\infty$ otherwise.

Notation: Let $\mathbf{\Delta}^{(k)} \in [-1/2, 1/2]^{|A_k|\times|A_k|}$ denote the sub-matrix of $\mathbf{\Delta}$ restricted to rows and columns in A_k. For $\alpha \in A_k$, let $\left(\Delta^{(k)}_{\alpha,(i)}\right)_{i\in\{1,\dots,|A_k|-1\}}$ denote the ordered gaps between α and arms in $A_k \setminus \{\alpha\}$ such that:

$$\Delta^{(k)}_{\alpha,(1)} \le \cdots \le \Delta^{(k)}_{\alpha,(|A_k|-1)}\ .$$

Since the gaps sub-matrix for arms in A_k is skew-symmetric (i.e., $\forall i,j:\ \mathbf{\Delta}^{(k)}_{i,j} = -\mathbf{\Delta}^{(k)}_{j,i}$), the number of arms such that $\Delta^{(k)}_{\alpha,(\lceil|A_k|/4\rceil)} \le 0$ is at least $\lceil|A_k|/4\rceil$ (see Lemma H.6). Let $E_k \subset A_k$ denote the last set of arms:

$$E_k := \left\{\alpha \in A_k : \Delta^{(k)}_{\alpha,(\lceil|A_k|/4\rceil)} \le 0\right\}\ .$$

Finally, we remind the reader that for any $j \in [K]$, the quantities $\Delta_{j,(1)} \leq \cdots \leq \Delta_{j,(K-1)}$ correspond to the ordered gaps between j and all arms in $[K] \setminus \{j\}$.

Conditioning convention: For each round k, let $\mathcal{F}_{k-1}$ denote the complete history before the scores at round k are computed. All roundwise estimates below are established conditionally on $\mathcal{F}_{k-1}$, for every history such that $i^* \in A_k$. In particular, A_k, B_k, E_k, F_k and the associated gap thresholds are then fixed. We omit this common conditioning from the notation within the technical lemmas and the roundwise calculations unless it is explicitly displayed. The final bounds follow by averaging these conditional estimates over the history.

Proof of Theorem C.1. Suppose that $T \geq 8K \log_{8/7}(K)$, otherwise the bound is vacuous. Assume that $\Delta_{i^*,i} > 0$ for all $i \neq i^*$. Otherwise, if $\Delta_{i^*,j} = 0$ for some $j \neq i^*$, then $H_{\mathrm{cw}} = +\infty$ and the stated bound is trivial.

We start by bounding the probability of the event $\psi_T \neq i^*$ by the probabilities that i^* gets eliminated at some step k. Since $i^* \in A_1 = [K]$, the event $\{\psi_T \neq i^*\}$ implies that there exists a round $k \in \{1, \ldots, k_{\max} - 1\}$ such that $i^* \in A_k$ but $i^* \notin A_{k+1}$. Hence,

$$\begin{aligned}
\mathbb{P}(\psi_T \neq i^*) &= \mathbb{P}\Big(\bigcup_{k=1}^{k_{\max}-1} \{i^* \in A_k,\ i^* \notin A_{k+1}\} \Big) \\
&\leq \sum_{k=1}^{k_{\max}-1} \mathbb{P}(i^* \in A_k,\ i^* \notin A_{k+1}) \\
&\leq k_{\max} \cdot \max_{k \in \{1,\ldots,k_{\max}-1\}} \mathbb{P}(i^* \notin A_{k+1} \mid i^* \in A_k)\,. \qquad (25)
\end{aligned}$$

Recall that $k_{\max} \leq \lceil \log_{8/7}(K) \rceil$.

Next, we upper-bound $\mathbb{P}(i^* \notin A_{k+1} \mid i^* \in A_k)$. Fix $k \in \{1, \ldots, k_{\max} - 1\}$ and condition on a realization of $\mathcal{F}_{k-1}$ such that $i^* \in A_k$. If $i^* \notin A_{k+1}$, then by the definition of the next set (keeping only the top fraction), the number of arms in A_k with score smaller than $S_k(i^*)$ is at most $\lceil |A_k|/8 \rceil$. Equivalently, at least $|A_k| - \lceil |A_k|/8 \rceil$ arms in A_k have score at least $S_k(i^*)$.

By Lemma H.7 applied to the skew-symmetric matrix $\mathbf{\Delta}^{(k)}$, we conclude that if $|A_k| \geq 3$ then the intersection between E_k and A_{k+1} (which have a size of $|A_k| - \lceil |A_k|/8 \rceil$) is non-empty, therefore

$$\exists \alpha \in E_k : \quad S_k(\alpha) \geq S_k(i^*).$$

Otherwise, if $|A_k| = 2$ and $i^* \in A_k$, we necessarily have $E_k = A_k \setminus \{i^*\}$. We conclude that

$$\begin{aligned}
\mathbb{P}(i^* \notin A_{k+1} \mid \mathcal{F}_{k-1}) &\leq \mathbb{P}\Big(\exists \alpha \in E_k :\ S_k(\alpha) \geq S_k(i^*) \Big) \\
&\leq \underbrace{\mathbb{P}\left(S_k(i^*) \leq \frac{1}{2}\, \Delta^{(k)}_{i^*,(\lceil |A_k|/8 \rceil)} \right)}_{\text{Term 1}} + \underbrace{\mathbb{P}\left(\exists \alpha \in E_k :\ S_k(\alpha) \geq \frac{1}{2}\, \Delta^{(k)}_{i^*,(\lceil |A_k|/8 \rceil)} \right)}_{\text{Term 2}}\,.
\end{aligned}$$

Denote $\Delta_k \coloneqq \Delta^{(k)}_{i^*,(\lceil |A_k|/8 \rceil)}$. By Lemma C.2, Terms 1 and 2 satisfy

$$\begin{aligned}
\mathbb{P}(i^* \notin A_{k+1} \mid \mathcal{F}_{k-1}) &\leq (K + 2\log(T)) \exp\left(-c\, \frac{\Delta_k^2}{\log(\lceil B_k/2 \rceil)}\, B_k \right) \\
&\quad + K \log(T) \exp\left(-c'\, \frac{\Delta_k^2}{\log(\lceil B_k/2 \rceil)}\, B_k \right) \\
&\leq 3K \log(T) \exp\left(-c_1\, \frac{\Delta_k^2}{\log(\lceil B_k/2 \rceil)}\, B_k \right)\,, \qquad (26)
\end{aligned}$$

where $c_1 = \min\{c, c'\}$.

Next, we develop a bound on $\Delta^{(k)}_{i^*,(\lceil |A_k|/8 \rceil)}$ using $H_{\mathrm{cw}} = \sum_{i \neq i^*} \frac{1}{\Delta^2_{i,i^*}}$. Recall $B_k = \left\lfloor \frac{T}{|A_k| \log_{8/7}(K)} \right\rfloor$. We have

$$\begin{aligned}
\Delta_k^2 \, B_k &= \Delta_k^2 \left\lfloor \frac{T}{|A_k| \log_{8/7}(K)} \right\rfloor \\
&\geq \Delta_k^2 \cdot \frac{T}{2|A_k| \log_{8/7}(K)} \\
&\geq \frac{\Delta_k^2}{\lceil |A_k|/8 \rceil} \cdot \frac{T}{16 \log_{8/7}(K)} \\
&\geq \frac{T}{16 \log_{8/7}(K) \cdot \sum_{i \neq i^*} \frac{1}{\Delta^2_{i^*,i}}} = \frac{T}{16 \log_{8/7}(K) \, H_{\mathrm{cw}}} \,,
\end{aligned}$$

where we used in the second line the fact that $T \geq 8K \log_{8/7}(K)$ and Lemma H.2 in the last line (using $\Delta_k \coloneqq \Delta^{(k)}_{i^*,(\lceil |A_k|/8 \rceil)}$). Plugging this into (26) yields

$$\begin{aligned}
\mathbb{P}(i^* \notin A_{k+1} \mid \mathcal{F}_{k-1}) &\leq 3K \log(T) \cdot \exp\left(-c_1 \, \frac{T}{16 \log_{8/7}(K) \, \log(\lceil B_k/2 \rceil) \, H_{\mathrm{cw}}} \right) \\
&\leq 3K \log(T) \exp\left(-c_1' \, \frac{T}{\log(K) \log(T) \, H_{\mathrm{cw}}} \right) \,, \qquad (27)
\end{aligned}$$

for numerical constants $c_1, c_1' > 0$ (using $\log_{8/7}(K) = \Theta(\log K)$ and $\log(\lceil B_k/2 \rceil) \leq \log(T)$).

The bound above is uniform over all histories satisfying $i^* \in A_k$. Therefore, using

$$\mathbb{P}(i^* \notin A_{k+1} \mid i^* \in A_k) = \mathbb{E}[\mathbb{P}(i^* \notin A_{k+1} \mid \mathcal{F}_{k-1}) \,|\, i^* \in A_k] \,,$$

the same bound holds conditionally only on $i^* \in A_k$.

Finally, we combine the bounds (25) and (27), and use $k_{\max} \leq \lceil \log_{8/7}(K) \rceil$, we get

$$\mathbb{P}(\psi_T \neq i^*) \leq \lceil \log_{8/7}(K) \rceil \cdot 3K \log(T) \exp\left(-c_1' \, \frac{T}{\log(K) \log(T) \, H_{\mathrm{cw}}} \right) \,,$$

which yields the claimed form

$$\mathbb{P}(\psi_T \neq i^*) \leq 27K \log(K) \log(T) \exp\left(-c \, \frac{T}{\log(T) \log(K) H_{\mathrm{cw}}} \right) \,,$$

for a numerical constant $c > 0$.

It remains to prove the following technical lemma.

Lemma C.2. *Consider step k in Algorithm 1 with $T \geq 8K \log_{8/7}(K)$ and assume that $i^* \in A_k$. Let $\Delta_k \coloneqq \Delta^{(k)}_{i^*,(\lceil |A_k|/8 \rceil)}$, then*

$$\begin{aligned}
\mathbb{P}\left(S_k(i^*) \leq \frac{1}{2}\Delta_k \right) &\leq (K + 2\log(T)) \exp\left(-c \frac{\Delta_k^2}{\log(\lceil B_k/2 \rceil)} \, B_k \right) , \\
\mathbb{P}\left(\exists \alpha \in E_k : \; S_k(\alpha) \geq \frac{1}{2}\Delta_k \right) &\leq K \log(T) \exp\left(-c' \, \frac{\Delta_k^2}{\log(\lceil B_k/2 \rceil)} \, B_k \right) \,,
\end{aligned}$$

for numerical constants $c, c' > 0$.

Proof. Assume $T \geq 8K \log_{8/7}(K)$, this guarantees $B_k = \left\lfloor \frac{T}{|A_k| \log_{8/7}(K)} \right\rfloor \geq 8$. Let c_1 denote the constant from Corollary B.2.

Proof of the first bound: Let i_s^* be the strong opponent chosen for i^* at step k of Algorithm 1. Recall

$$S_k(i^*) = \min\{Z_k^{(s)}(i^*), 0\} + Z_k^{(w)}(i^*)\,.$$

Therefore, we have

$$\mathbb{P}\left(S_k(i^*) \le \frac{1}{2}\Delta_k\right) \le \mathbb{P}\left(Z_k^{(s)}(i^*) + Z_k^{(w)}(i^*) \le \frac{1}{2}\Delta_k\right) + \mathbb{P}\left(Z_k^{(w)}(i^*) \le \frac{1}{2}\Delta_k\right)\,. \tag{28}$$

Recall that the event $\left\{Z_k^{(s)}(i^*) + Z_k^{(w)}(i^*) \le \frac{1}{2}\Delta_k\right\}$ implies that

$$\left\{Z_k^{(s)}(i^*) \le -\frac{1}{4}\Delta_k \text{ or } Z_k^{(w)}(i^*) \le \frac{3}{4}\Delta_k\right\}\,,$$

Combining with Inequality 28 we obtain

$$\mathbb{P}\left(S_k(i^*) \le \frac{1}{2}\Delta_k\right) \le \mathbb{P}\left(Z_k^{(s)}(i^*) \le -\frac{1}{4}\Delta_k\right) + 2\mathbb{P}\left(Z_k^{(w)}(i^*) \le \frac{3}{4}\Delta_k\right)\,. \tag{29}$$

We use Hoeffding's inequality (Lemma H.10) to bound the first term in the upper bound above. For any fixed $i \in [K] \setminus \{i^*\}$ and $\epsilon > 0$, we have

$$\mathbb{P}\left(\hat{\Delta}_{i^*,i} - \Delta_{i^*,i} \le -\epsilon\right) \le \exp\left(-\frac{\epsilon^2 B_k}{2}\right)\,,$$

where $\hat{\Delta}_{i^*,i}$ is the empirical mean of duels between (i^*, i) computed using $\lceil B_k/4 \rceil$ samples. Therefore, applying the bound above with $\epsilon = \Delta_k/4$ and a union bound over the arms, we have

$$\begin{aligned}\mathbb{P}\left(Z_k^{(s)}(i^*) \le -\frac{1}{4}\Delta_k\right) &\le \mathbb{P}\left(Z_k^{(s)}(i^*) - \Delta_{i^*,i_s^*} \le -\frac{1}{4}\Delta_k\right) \\ &\le (K-1)\exp\left(-\frac{\Delta_k^2}{32}B_k\right)\,. \end{aligned}\tag{30}$$

where we used the fact that $\Delta_{i^*,j} \ge 0$ for all $j \in [K]$. Now, using the shifted-rank version of Corollary B.2 with $N = |A_k| - 1$, which gives a guarantee on the output $Z_k^{(w)}(i^*)$, we have

$$\begin{aligned}\mathbb{P}\left(Z_k^{(w)}(i^*) \le \frac{3}{4}\Delta_k\right) &= \mathbb{P}\left(Z_k^{(w)}(i^*) \le \Delta^{(k)}_{i^*,(\lceil |A_k|/8 \rceil)} - \frac{1}{4}\Delta_k\right) \\ &\le \log\left(\left\lceil \frac{B_k}{2} \right\rceil\right) \exp\left(-c_1 \frac{\Delta_k^2}{32\log(\lceil B_k/2 \rceil)} B_k\right)\,. \end{aligned}\tag{31}$$

We conclude by combining (30), (31) and (29) that

$$\mathbb{P}\left(S_k(i^*) \le \frac{1}{2}\Delta_k\right) \le (K - 1 + 2\log(T)) \exp\left(-c\frac{\Delta_k^2}{\log(\lceil B_k/2 \rceil)} B_k\right)\,,$$

where c is a numerical constant.

Proof of the second bound: Fix $\alpha \in E_k$. By definition of E_k, we have

$$\Delta^{(k)}_{\alpha,(\lceil |A_k|/4 \rceil)} \le 0\,.$$

Moreover, $\min\{Z_k^{(s)}(\alpha), 0\} \le 0$, hence

$$\begin{aligned}\mathbb{P}\left(S_k(\alpha) \ge \frac{1}{2}\Delta_k\right) &= \mathbb{P}\left(\min\{Z_k^{(s)}(\alpha), 0\} + Z_k^{(w)}(\alpha) \ge \frac{1}{2}\Delta_k\right) \\ &\le \mathbb{P}\left(Z_k^{(w)}(\alpha) \ge \frac{1}{2}\Delta_k\right) \\ &\le \mathbb{P}\left(Z_k^{(w)}(\alpha) \ge \Delta^{(k)}_{\alpha,(\lceil |A_k|/4 \rceil)} + \frac{1}{2}\Delta_k\right)\,, \end{aligned}$$

where the last step uses $\Delta^{(k)}_{\alpha,(\lceil |A_k|/4 \rceil)} \leq 0$. Applying the shifted-rank version of Corollary B.2 with $N = |A_k| - 1$ to $Z^{(w)}_k(\alpha)$ then yields

$$\mathbb{P}\left(S_k(\alpha) \geq \frac{1}{2}\Delta_k\right) \leq \log\left(\left\lceil \frac{B_k}{2} \right\rceil\right) \exp\left(-c_1 \frac{\Delta_k^2}{8\log(\lceil B_k/2 \rceil)} B_k\right) . \tag{32}$$

Finally, union bound over $\alpha \in E_k$ gives

$$\mathbb{P}\left(\exists \alpha \in E_k : \ S_k(\alpha) \geq \frac{1}{2}\Delta_k\right) \leq |E_k| \log\left(\left\lceil \frac{B_k}{2} \right\rceil\right) \exp\left(-c' \frac{\Delta_k^2}{\log(\lceil B_k/2 \rceil)} B_k\right) ,$$

We may bound $|E_k|$ by K; hence the above yields the stated form

$$\mathbb{P}\left(\exists \alpha \in E_k : \ S_k(\alpha) \geq \frac{1}{2}\Delta_k\right) \leq K \log(T) \exp\left(-c' \frac{\Delta_k^2}{\log(\lceil B_k/2 \rceil)} B_k\right) ,$$

for a numerical constant $c' > 0$. This proves the second inequality and concludes the lemma. □

C.2 Second Upper Bound

For each $i \neq i^*$, let $\Delta_{i,(k)}$ denote the ordered gaps $(\Delta_{i,j})_{i \neq j}$ as

$$\Delta_{i,(1)} \leq \cdots \leq \Delta_{i,(K-1)} ,$$

Denote by $K_{i;<0}$ the number of j such that $\Delta_{i,j} < 0$. For each $i \in [K]$, let $s_i \leq K_{i;<0}$, and $\boldsymbol{s} = (s_1, \ldots, s_K)$. Here, we take the convention $K_{i^*;<0} = 0$. We recall the expressions of the quantities $H_{\text{certify}}(\boldsymbol{s})$, $H^{(0)}_{\text{explore}}(\boldsymbol{s})$ and $H^{(1)}_{\text{explore}}(\boldsymbol{s})$

$$H_{\text{certify}}(\boldsymbol{s}) = \sum_{i \neq i^*} \frac{1}{\Delta^2_{i,(s_i)}}, \quad H^{(1)}_{\text{explore}}(\boldsymbol{s}) = \max_{i \neq i^*} \frac{K}{s_i \Delta^2_{i,(s_i)}} \text{ and } H^{(0)}_{\text{explore}}(\boldsymbol{s}) = \sum_{i \neq i^*} \frac{K}{s_i \Delta^2_{i,(s_i)}} .$$

Theorem C.3. *For $T \geq 2$ and any $\boldsymbol{s}$ such that, for every $i \neq i^*$, $1 \leq s_i \leq K_{i;<0}$, it holds that*

$$\mathbb{P}\left(\psi_T \neq i^*\right) \leq 47K \log(K) \log(T) \times \exp\left(-\frac{c_1}{\log^4(K)\log(T)} \frac{T - c_2 \log^6(H^{(0)}_{\textit{explore}}(\boldsymbol{s})) \cdot H^{(0)}_{\textit{explore}}(\boldsymbol{s})}{H^{(1)}_{\textit{explore}}(\boldsymbol{s}) + H_{\textit{certify}}(\boldsymbol{s})}\right),$$

where c_1 and c_2 are numerical constants.

We restate and extend the notation introduced in the last section.

Notation: Let $\boldsymbol{\Delta}^{(k)} \in [-1/2, 1/2]^{|A_k| \times |A_k|}$ denote the sub-matrix of $\boldsymbol{\Delta}$ restricted to lines and rows in A_k. For $\alpha \in A_k$, let $\left(\Delta^{(k)}_{\alpha,(i)}\right)_{i \in \{1,\ldots,|A_k|-1\}}$ denote the ordered gaps between α and arms in A_k such that:

$$\Delta^{(k)}_{\alpha,(1)} \leq \cdots \leq \Delta^{(k)}_{\alpha,(|A_k|-1)}.$$

Recall that since the gaps sub-matrix for arms in A_k is skew-symmetric (i.e., $\forall i, j : \boldsymbol{\Delta}^{(k)}_{i,j} = -\boldsymbol{\Delta}^{(k)}_{j,i}$), the number of arms such that $\Delta^{(k)}_{\alpha,(\lceil |A_k|/4 \rceil)} \leq 0$ is at least $\lceil |A_k|/4 \rceil$ (see Lemma H.6). Let $E_k \subset A_k$ denote the last set of arms

$$E_k := \left\{\alpha \in A_k : \Delta^{(k)}_{\alpha,(\lceil |A_k|/4 \rceil)} \leq 0\right\} .$$

We rank the quantities $(\Delta_{\alpha,(s_\alpha)})_{\alpha \in E_k}$. We denote the ranked sequence with ties broken arbitrarily $(\Delta_{E_k:i})_{i \in [|E_k|]}$

$$\Delta_{E_k:1} \leq \cdots \leq \Delta_{E_k:|E_k|} .$$

Define the quantity $\bar{\Delta}_k$ by

$$\bar{\Delta}_k := \Delta_{E_k:\lceil \frac{7}{8}|E_k| \rceil} \leq 0 . \tag{33}$$

Observe that when $\bar{\Delta}_k = 0$, we necessarily have $\Delta_{i,(s_i)} = 0$ for some $i \in [K] \setminus \{i^*\}$, this implies in particular that $H_{\text{certify}} = \infty$ and the bound becomes loose. Therefore, in the remainder of this proof, we assume that $\bar{\Delta}_k < 0$.

Let F_k denote the subset of arms in E_k such that $\Delta_{\alpha,(s_\alpha)} \leq \bar{\Delta}_k$.

$$F_k := \left\{\alpha \in E_k : \Delta_{\alpha,(s_\alpha)} \leq \bar{\Delta}_k\right\} . \tag{34}$$

Finally, we denote for each $i \neq i^* :\ \Gamma_i := s_i \Delta^2_{i,(s_i)}$, and let $(\Gamma_{(i)})_{i \neq i^*}$ correspond to the ranked quantities $\Gamma_{(2)} \leq \cdots \leq \Gamma_{(K)}$, with ties broken arbitrarily.

Proof of Theorem C.3. Fix $\boldsymbol{s}$ such that $1 \leq s_i \leq K_{i;<0}$ for all $i \neq i^*$ (and $K_{i^*;<0} = 0$ by convention). Note that by the assumption of the uniqueness of the Condorcet winner we have $K_{i;<0} \geq 1$ for any $i \neq i^*$. Let $c > 0$ be a numerical constant (chosen smaller than the constants appearing in Corollary B.2 and Theorem B.5). We assume that $T \geq 8K \log_{8/7}(K)$, otherwise the bound of the theorem is vacuous.

Similar to the proof of Theorem C.1, we start by bounding the probability of the event $\psi_T \neq i^*$ by the probabilities that i^* gets eliminated at some step k. We have

$$\mathbb{P}(\psi_T \neq i^*) \leq \sum_{k=1}^{k_{\max}-1} \mathbb{P}(i^* \notin A_{k+1},\ i^* \in A_k) \leq k_{\max} \cdot \max_{k \leq k_{\max}-1} \mathbb{P}(i^* \notin A_{k+1} \mid i^* \in A_k) . \tag{35}$$

Recall $k_{\max} \leq \lceil \log_{8/7}(K) \rceil$. Fix $k \in \{1, \ldots, k_{\max} - 1\}$ and condition on a realization of $\mathcal{F}_{k-1}$ such that $i^* \in A_k$. We will first consider the case where $|A_k| \geq 3$. The case where $|A_k| = 2$ is simple and is left to the end of this proof.

Next, we build the argument of our proof on the observation that given $i^* \in A_k$, the event $i^* \notin A_{k+1}$ implies in particular that the number of arms $\alpha \in A_k$ with a score $S_k(\alpha)$ at least as large as $S_k(i^*)$ is at least $|A_k| - \lceil |A_k|/8 \rceil$. Therefore, the event $i^* \notin A_{k+1}$ implies that the number of arms in F_k with a score $S_k(\cdot)$ at least as large as $S_k(i^*)$ is at least $|A_{k+1} \cap F_k| \geq \lceil |F_k|/3 \rceil$ as stated in the following lemma

Lemma C.4. *Let $k \in \{1, \ldots, k_{\max} - 1\}$, recall the definition of F_k given in* (34)*. We have, if $|A_k| \geq 3$ then*

$$|A_{k+1} \cap F_k| \geq \left\lceil \frac{1}{3}|F_k| \right\rceil .$$

This lemma implies that if i^* is eliminated at step k (i.e., $i^* \notin A_{k+1}$), then many "bad" arms in F_k beat i^*. More precisely, since A_{k+1} consists of the top-scoring arms at step k, every $\alpha \in A_{k+1}$ satisfies $S_k(\alpha) \geq S_k(i^*)$ whenever $i^* \notin A_{k+1}$. Therefore,

$$\{i^* \notin A_{k+1}\} \ \subseteq\ \left\{ \left|\{\alpha \in F_k :\ S_k(\alpha) \geq S_k(i^*)\}\right| \ \geq\ \left\lceil \frac{1}{3}|F_k| \right\rceil \right\} .$$

Introduce the threshold $\frac{1}{2}\bar{\Delta}_k$ defined by (33) and split

$$\begin{aligned}
\mathbb{P}(i^* \notin A_{k+1} \mid \mathcal{F}_{k-1}) &\leq \mathbb{P}\left(\left|\{\alpha \in F_k :\ S_k(\alpha) \geq S_k(i^*)\}\right| \geq \left\lceil \frac{1}{3}|F_k| \right\rceil \right) \\
&\leq \underbrace{\mathbb{P}\left(S_k(i^*) \leq \frac{1}{2}\bar{\Delta}_k \right)}_{\text{Term 1}} + \underbrace{\mathbb{P}\left(\left|\{\alpha \in F_k :\ S_k(\alpha) \geq \frac{1}{2}\bar{\Delta}_k\}\right| \geq \left\lceil \frac{1}{3}|F_k| \right\rceil \right)}_{\text{Term 2}} .
\end{aligned}$$

The following lemma is a key step in the proof, we postponed its proof to the next subsection.

Lemma C.5. *Under the assumptions of Theorem C.3, with $T \geq 8K \log_{8/7}(K)$, consider step k in Algorithm 1 and assume that $i^* \in A_k$. Then, we have*

$$\mathbb{P}\left(S_k(i^*) \leq \frac{1}{2}\bar{\Delta}_k \mid \mathcal{F}_{k-1} \right) \leq (K + \log(T)) \exp\left(-c \frac{\bar{\Delta}_k^2}{\log(B_k)} B_k \right)$$

$$\mathbb{P}\left(|\{\alpha \in F_k : \frac{1}{2}\bar{\Delta}_k \leq S_k(\alpha)\}| \geq \left\lceil \frac{1}{3}|F_k| \right\rceil \mid \mathcal{F}_{k-1} \right) \leq \exp\left(-\frac{c}{\log^4(K)\log(T)} \cdot \frac{T - c' \log^6(H^{(0)}_{\textit{explore}}(\boldsymbol{s})) \cdot H^{(0)}_{\textit{explore}}(\boldsymbol{s})}{H^{(1)}_{\textit{explore}}(\boldsymbol{s})} \right) ,$$

where c and c' are positive numerical constants.

A direct application of the lemma above gives (for numerical constants $c, c' > 0$)

$$\mathbb{P}(i^* \notin A_{k+1} \mid \mathcal{F}_{k-1}) \leq (K + \log(T)) \exp\left(-c \frac{\bar{\Delta}_k^2}{\log(B_k)} B_k\right) + \exp\left(-\frac{c}{\log^4(K)\log(T)} \cdot \frac{T - c' \log^6(H^{(0)}_{\text{explore}}(\boldsymbol{s}))\, H^{(0)}_{\text{explore}}(\boldsymbol{s})}{H^{(1)}_{\text{explore}}(\boldsymbol{s})}\right). \tag{36}$$

Next, we will convert the dependence of the bound on $\bar{\Delta}_k$ into $H_{\text{certify}}(\boldsymbol{s})$.

Recall $|E_k| \geq \lceil |A_k|/4 \rceil$. Since $\bar{\Delta}_k = \Delta_{E_k:\lceil \frac{7}{8}|E_k| \rceil}$, the gaps with ranks at least $\lceil 7|E_k|/8 \rceil$ lie between $\bar{\Delta}_k$ and zero. Hence $\sum_{\alpha \in E_k} 1/\Delta^2_{\alpha,(s_\alpha)} \geq (|E_k| - \lceil 7|E_k|/8 \rceil + 1)/\bar{\Delta}_k^2 \geq \lceil |E_k|/8 \rceil / \bar{\Delta}_k^2$, which yields

$$|A_k| \cdot \frac{1}{\bar{\Delta}_k^2} \leq 4|E_k| \cdot \frac{1}{\bar{\Delta}_k^2} \leq 32 \cdot \left\lceil \frac{|E_k|}{8} \right\rceil \cdot \frac{1}{\Delta^2_{E_k:\lceil \frac{7}{8}|E_k| \rceil}} \leq 32 \sum_{\alpha \in E_k} \frac{1}{\Delta^2_{\alpha,(s_\alpha)}} \leq 32 H_{\text{certify}}(\boldsymbol{s}). \tag{37}$$

Using $B_k = \left\lfloor \frac{T}{|A_k| \log_{8/7}(K)} \right\rfloor \geq \frac{T}{2|A_k| \log_{8/7}(K)}$ (and $\log(B_k) \leq \log(T)$), we obtain

$$(K + \log(T)) \exp\left(-c \frac{\bar{\Delta}_k^2}{\log(B_k)} B_k\right) \leq 2(K + \log(T)) \exp\left(-c' \frac{T}{H_{\text{certify}}(\boldsymbol{s})\, \log(T)\, \log(K)}\right), \tag{38}$$

for a numerical constant $c' > 0$ (absorbing $\log_{8/7}(K) = \Theta(\log K)$ into constants).

Next combine (36) and (38), and using $\exp(-a) + \exp(-b) \leq 2\exp(-\min\{a, b\})$, we get that if $|A_k| \geq 3$, we have

$$\mathbb{P}(i^* \notin A_{k+1} \mid \mathcal{F}_{k-1}) \leq 3(K + \log(T)) \times \exp\left(-\frac{c}{\log^4(K)\log(T)} \cdot \frac{T - c_2 \log^6(H^{(0)}_{\text{explore}}(\boldsymbol{s}))\, H^{(0)}_{\text{explore}}(\boldsymbol{s})}{H^{(1)}_{\text{explore}}(\boldsymbol{s}) + H_{\text{certify}}(\boldsymbol{s})}\right),$$

for a numerical constant $c_2 > 0$ (renaming constants). To conclude we need to consider the edge case where $|A_k| = 2$ (last iteration). In this case we have $E_k = F_k = \{\alpha\}$. Therefore,

$$\begin{aligned}\mathbb{P}(i^* \notin A_{k+1} \mid \mathcal{F}_{k-1}) &\leq \mathbb{P}(S_k(\alpha) \geq S_k(i^*)) \\ &\leq \mathbb{P}\left(S_k(i^*) \leq \frac{1}{2}\bar{\Delta}_k\right) + \mathbb{P}\left(S_k(\alpha) \geq \frac{1}{2}\bar{\Delta}_k\right).\end{aligned}$$

The first term in the upper bound can be bounded using (C.5), the second term can be bounded using Lemma C.6. The resulting bound is smaller than the one obtained when $|A_k| \geq 3$.

The bound above is uniform over all histories satisfying $i^* \in A_k$. Therefore, using

$$\mathbb{P}(i^* \notin A_{k+1} \mid i^* \in A_k) = \mathbb{E}[\mathbb{P}(i^* \notin A_{k+1} \mid \mathcal{F}_{k-1}) \,|\, i^* \in A_k],$$

the same bound holds conditionally only on $i^* \in A_k$.

Finally, using (35) and $k_{\max} \leq \lceil \log_{8/7}(K) \rceil$, and absorbing $\lceil \log_{8/7}(K) \rceil$ and additive logarithms into the prefactor, we obtain

$$\mathbb{P}(\psi_T \neq i^*) \leq 47K \log(K) \log(T) \times \exp\left(-\frac{c_1}{\log^4(K)\log(T)} \cdot \frac{T - c_2 \log^6(H^{(0)}_{\text{explore}}(\boldsymbol{s}))\, H^{(0)}_{\text{explore}}(\boldsymbol{s})}{H^{(1)}_{\text{explore}}(\boldsymbol{s}) + H_{\text{certify}}(\boldsymbol{s})}\right),$$

which is the claim of Theorem C.3.

C.3 Proofs of Technical Lemmas

C.3.1 Proof of Lemma C.4

Proof. Recall $\bar{\Delta}_k = \Delta_{E_k:\lceil \frac{7}{8}|E_k| \rceil}$ and $F_k = \{\alpha \in E_k : \Delta_{\alpha,(s_\alpha)} \le \bar{\Delta}_k\}$, hence

$$|F_k| \;\ge\; \left\lceil \frac{7}{8}|E_k| \right\rceil. \tag{39}$$

Algorithm 1 keeps $|A_{k+1}| = |A_k| - \lceil |A_k|/8 \rceil$ arms, so for any $A_{k+1} \subseteq A_k$ and $F_k \subseteq A_k$,

$$|A_{k+1} \cap F_k| \;\ge |A_{k+1}| + |F_k| - |A_k| = |F_k| - \left\lceil \frac{|A_k|}{8} \right\rceil . \tag{40}$$

Case 1: $|A_k| \ge 5$**.** By Lemma H.6, $|E_k| \ge \lceil |A_k|/4 \rceil \ge 2$, hence $\lceil |A_k|/8 \rceil \le \lceil |E_k|/2 \rceil$. Moreover, for every integer $m \ge 2$,

$$2\left\lceil \frac{7m}{8} \right\rceil \ge 3\left\lceil \frac{m}{2} \right\rceil , \tag{41}$$

(which follows by writing $m = 8q + r$ and checking $r \in \{0, \dots, 7\}$; the only delicate residue $r = 1$ is harmless since then $q \ge 1$). Applying (41) with $m = |E_k|$ and using (39) gives $\lfloor \frac{2}{3}|F_k| \rfloor \ge \lceil |E_k|/2 \rceil \ge \lceil |A_k|/8 \rceil$. Plugging into (40) yields

$$|A_{k+1} \cap F_k| \ge |F_k| - \left\lfloor \frac{2}{3}|F_k| \right\rfloor = \left\lceil \frac{1}{3}|F_k| \right\rceil .$$

Case 2: $|A_k| \in \{3, 4\}$**.** Here $\lceil |A_k|/8 \rceil = 1$. By skew-symmetry of $\Delta^{(k)}$, at most one row can have all off-diagonal entries > 0, hence at least $|A_k| - 1$ rows have $\Delta^{(k)}_{\alpha,(1)} \le 0$, so $|E_k| \ge |A_k| - 1 \in \{2, 3\}$. Then $\lceil \frac{7}{8}|E_k| \rceil = |E_k|$, and since $F_k \subseteq E_k$, (39) implies $|F_k| = |E_k| \ge |A_k| - 1$. Using (40),

$$|A_{k+1} \cap F_k| \;\ge\; |F_k| - 1 ,$$

and for $|F_k| \in \{2, 3\}$ this satisfies $|F_k| - 1 \ge \lceil |F_k|/3 \rceil$.

Combining both cases proves that for every $|A_k| \ge 3$,

$$|A_{k+1} \cap F_k| \;\ge\; \left\lceil \frac{1}{3}|F_k| \right\rceil .$$

□

C.3.2 Proof of Lemma C.5

Proof. Fix a round $k \in \{1, \dots, k_{\max} - 1\}$ and recall $\bar{\Delta}_k \le 0$ by construction.

Proof of the first bound

$$\mathbb{P}\left(S_k(i^*) \le \frac{1}{2}\bar{\Delta}_k \right) \le (K + \log(T)) \exp\left(-c \frac{\bar{\Delta}_k^2}{\log(B_k)} B_k \right) .$$

If $\bar{\Delta}_k = 0$ the bound is immediate. Assume $\bar{\Delta}_k < 0$. Recall $S_k(i^*) = \min\{Z_k^{(s)}(i^*), 0\} + Z_k^{(w)}(i^*)$. Then

$$\begin{aligned}
\mathbb{P}\left(S_k(i^*) \le \frac{1}{2}\bar{\Delta}_k \right) &\le \mathbb{P}\left(Z_k^{(s)}(i^*) + Z_k^{(w)}(i^*) \le \frac{1}{2}\bar{\Delta}_k \right) + \mathbb{P}\left(Z_k^{(w)}(i^*) \le \frac{1}{2}\bar{\Delta}_k \right) \\
&\le \mathbb{P}\left(Z_k^{(s)}(i^*) \le \frac{1}{4}\bar{\Delta}_k \right) + 2\,\mathbb{P}\left(Z_k^{(w)}(i^*) \le \frac{1}{4}\bar{\Delta}_k \right) ,
\end{aligned}$$

where the last step uses $\bar{\Delta}_k < 0$, hence $\{Z_k^{(w)}(i^*) \le \frac{1}{2}\bar{\Delta}_k\} \subseteq \{Z_k^{(w)}(i^*) \le \frac{1}{4}\bar{\Delta}_k\}$.

The first term in the last upper-bound is bounded using Hoeffding and a union bound over the opponent choice,

$$\mathbb{P}\left(Z_k^{(s)}(i^*) \le \frac{1}{4}\bar{\Delta}_k\right) \le (K-1)\exp\left(-\frac{\bar{\Delta}_k^2}{32}B_k\right).$$

The second term is bounded by the shifted-rank version of Corollary B.2 with $N = |A_k| - 1$,

$$\begin{aligned}\mathbb{P}\left(Z_k^{(w)}(i^*) \le \frac{1}{4}\bar{\Delta}_k\right) &\le \mathbb{P}\left(Z_k^{(w)}(i^*) \le \Delta^{(k)}_{i^*,(\lceil |A_k|/8\rceil)} + \frac{1}{4}\bar{\Delta}_k\right)\\ &\le \log(T)\exp\left(-c\,\frac{\bar{\Delta}_k^2}{\log(\lceil B_k/2\rceil)}\,B_k\right).\end{aligned}$$

Finally, absorbing $\log(\lceil B_k/2\rceil)$ into $\log(B_k)$ and constants yields

$$\mathbb{P}\left(S_k(i^*) \le \frac{1}{2}\bar{\Delta}_k\right) \le (K+\log(T))\exp\left(-c\,\frac{\bar{\Delta}_k^2}{\log(B_k)}\,B_k\right).$$

Proof of the second bound

$$\mathbb{P}\left(|\{\alpha \in F_k :\ S_k(\alpha) \ge \frac{1}{2}\bar{\Delta}_k\}| \ge \lceil |F_k|/3\rceil\right) \le \exp\left(-\frac{c}{\log^4(K)\log(T)} \cdot \frac{T - c'\log^6(H^{(0)}_{\text{explore}}(\boldsymbol{s}))\cdot H^{(0)}_{\text{explore}}(\boldsymbol{s})}{H^{(1)}_{\text{explore}}(\boldsymbol{s})}\right).$$

We start from the indicator-sum form

$$\mathbb{P}\left(|\{\alpha \in F_k :\ S_k(\alpha) \ge \tfrac{1}{2}\bar{\Delta}_k\}| \ge \left\lceil \frac{|F_k|}{3}\right\rceil\right) = \mathbb{P}\left(\sum_{\alpha\in F_k} \mathbb{1}\big(S_k(\alpha) \ge \tfrac{1}{2}\bar{\Delta}_k\big) \ge \left\lceil \frac{|F_k|}{3}\right\rceil\right). \quad (42)$$

Next, we keep only the hardest $3/4$ of F_k. More formally, we rank $\Gamma_\alpha = s_\alpha \Delta^2_{\alpha,(s_\alpha)}$ over $\alpha \in F_k$ as $\Gamma_{F_k:1} \le \cdots \le \Gamma_{F_k:|F_k|}$, and let $F_k^{(3/4)}$ be the subset containing the top $\lceil 3|F_k|/4\rceil$ arms with largest Γ_α. Then $|F_k \setminus F_k^{(3/4)}| = \lfloor |F_k|/4\rfloor$, so

$$\left\{\sum_{\alpha\in F_k} \mathbb{1}(S_k(\alpha) \ge \tfrac{1}{2}\bar{\Delta}_k) \ge \left\lceil \frac{|F_k|}{3}\right\rceil\right\} \subseteq \left\{\sum_{\alpha\in F_k^{(3/4)}} \mathbb{1}(S_k(\alpha) \ge \tfrac{1}{2}\bar{\Delta}_k) \ge \left\lceil \frac{|F_k|}{12}\right\rceil\right\}.$$

Let us develop a uniform per-arm bound on $F_k^{(3/4)}$. Lemma C.6 below gives such a bound

Lemma C.6. *Let $\alpha \in F_k^{(3/4)}$ and $T \ge 8K\log_{8/7}(K)$. We have*

$$\mathbb{P}\left(S_k(\alpha) \ge \frac{1}{2}\bar{\Delta}_k\right) \le (\log(T)+K)\exp\left(-c"\,\frac{T}{\log^4(K)\log(T)\ \sum_{i\in E_k}\frac{K}{s_i\Delta^2_{i,(s_i)}}}\right), \quad (43)$$

where $c"$ is a positive numerical constant. Moreover, there exists a sufficiently large universal numerical constant $c_2 > 0$, depending only on $c"$, such that if $T \ge c_2\, H^{(0)}_{\text{explore}}(\boldsymbol{s})\ \log^6\big(H^{(0)}_{\text{explore}}(\boldsymbol{s})\big)$, we have

$$\mathbb{P}\left(S_k(\alpha) \ge \frac{1}{2}\bar{\Delta}_k\right) \le \frac{1}{18}.$$

In the remainder of this proof we assume that the condition $T \ge c_2\, H^{(0)}_{\text{explore}}(\boldsymbol{s})\ \log^6\big(H^{(0)}_{\text{explore}}(\boldsymbol{s})\big)$ is satisfied, otherwise the upper bound stated by the lemma is greater than 1 and is thus vacuous. Denote by p_k the bound given by the lemma above

$$p_k \coloneqq \frac{1}{18} \wedge\ (\log(T)+K)\exp\left(-c"\,\frac{T}{\log^4(K)\log(T)\ \sum_{i\in E_k}\frac{K}{s_i\Delta^2_{i,(s_i)}}}\right).$$

We use Lemma H.9, which is purely technical and deferred to Section H, to obtain the following upper bound

$$p_k \leq \exp\left(-\bar{c}_1 \cdot \frac{T}{\log^4(K)\log(T)\sum_{i\in E_k}\frac{K}{s_i\Delta^2_{i,(s_i)}}}\right) , \tag{44}$$

where $\bar{c}_1$ is a numerical constant depending only on c". Conditionally on $\mathcal{F}_{k-1}$, the scores $(S_k(\alpha))_{\alpha\in F_k^{(3/4)}}$ are independent, since the algorithm uses independent fresh samples and randomization for each arm. Therefore, the following stochastic domination holds conditionally on $\mathcal{F}_{k-1}$:

$$\sum_{\alpha\in F_k^{(3/4)}} \mathbb{1}\big(S_k(\alpha) \geq \tfrac{1}{2}\bar{\Delta}_k\big) \text{ is stochastically dominated by } M_k \sim \mathrm{Bin}\left(\left\lceil\frac{3}{4}|F_k|\right\rceil, p_k\right) .$$

Consequently, using the fact that $p_k \leq \frac{1}{18}$ implies $\left\lceil\frac{|F_k|}{12}\right\rceil - p_k\left\lceil\frac{3}{4}|F_k|\right\rceil \geq \frac{|F_k|}{24}$ we have

$$\begin{aligned}\mathbb{P}\left(\sum_{\alpha\in F_k^{(3/4)}} \mathbb{1}\big(S_k(\alpha) \geq \tfrac{1}{2}\bar{\Delta}_k\big) \geq \left\lceil\frac{|F_k|}{12}\right\rceil\right) &\leq \mathbb{P}\left(M_k \geq \left\lceil\frac{|F_k|}{12}\right\rceil\right)\\ &\leq \mathbb{P}\left(M_k - \mathbb{E}[M_k] \geq \frac{|F_k|}{24}\right) .\end{aligned} \tag{45}$$

Next, we use Lemma H.12 which provides a deviation bound for binomial variables in regimes where the parameters can be small. Recall that M_k is a binomial distribution with parameters $(p_k, \lceil 3|F_k|/4\rceil)$. We have

$$\mathbb{P}\left(M_k - \mathbb{E}[M_k] \geq \frac{|F_k|}{24}\right) \leq \exp\left(-\frac{|F_k|^2}{1152\lceil 3|F_k|/4\rceil\,\phi(p_k)}\right) \leq \exp\left(-\frac{|F_k|}{1152\,\phi(p_k)}\right) , \tag{46}$$

where $\lceil 3|F_k|/4\rceil \leq |F_k|$ and ϕ is the function defined in Lemma H.11. Since we have proved that $p_k \leq \frac{1}{18}$, the expression of $\phi(p_k)$ is therefore given by

$$\phi(p_k) = \frac{\frac{1}{2} - p_k}{\log(1-p_k) - \log(p_k)} .$$

Since the function ϕ is increasing on $(0, 1/2)$ and Lemma H.5 gives, for any $y > 0$,

$$0 < \frac{\frac{1}{2} - \exp(-y)}{\log(1-\exp(-y)) - \log(\exp(-y))} \leq \frac{1}{2y} ,$$

with the middle expression defined by continuity to equal $1/4$ at $y = \log 2$, we conclude using the bound (44) that

$$\frac{1}{\phi(p_k)} \geq 2\frac{\bar{c}_1}{320\log(T)\log^4(K)} \cdot \frac{T}{\sum_{i\in E_k}\frac{K}{s_i\Delta^2_{i,(s_i)}}} .$$

Using the bound above with (46) and $|F_k| \geq \frac{7}{8}|E_k|$, we have

$$\begin{aligned}\mathbb{P}\left(M_k - \mathbb{E}[M_k] \geq \frac{|F_k|}{24}\right) &\leq \exp\left(-\frac{|F_k|}{1152}\cdot\frac{\bar{c}_1}{320\log(T)\log^4(K)}\cdot\frac{T}{\sum_{i\in E_k}\frac{K}{s_i\Delta^2_{i,(s_i)}}}\right)\\ &\leq \exp\left(-\bar{c}_2\cdot\frac{T|E_k|}{\log^4(K)\log(T)\sum_{i\in E_k}\frac{K}{s_i\Delta^2_{i,(s_i)}}}\right) ,\end{aligned}$$

where $\bar{c}_2$ is a numerical constant. We then use the fact that

$$\frac{|E_k|}{\sum_{i\in E_k}\frac{K}{s_i\Delta^2_{i,(s_i)}}} \;\geq\; \frac{1}{\max_{i\neq i^*}\frac{K}{s_i\Delta^2_{i,(s_i)}}} \;=\; \frac{1}{H^{(1)}_{\mathrm{explore}}(\boldsymbol{s})} .$$

Plugging these two relations into (45) then (42) yields for a numerical constant $\bar{c}_3$

$$\mathbb{P}\left(|\{\alpha \in F_k :\ S_k(\alpha) \geq \tfrac{1}{2}\bar{\Delta}_k\}| \geq \left\lceil \frac{|F_k|}{3} \right\rceil\right) \leq \exp\left(-\bar{c}_3 \cdot \frac{T}{\log^4(K)\log(T)\, H^{(1)}_{\mathrm{explore}}(\boldsymbol{s})}\right),$$

as soon as $T \geq c_2 H^{(0)}_{\mathrm{explore}}(\boldsymbol{s})\log^6(H^{(0)}_{\mathrm{explore}}(\boldsymbol{s}))$. Reintroducing the shift (to cover smaller T) gives the stated bound in Lemma C.5. □

Proof. **of Lemma C.6**. Fix $\alpha \in F_k^{(3/4)}$, let $\alpha^{(s)}$ denote the strong opponent chosen for α. We have

$$\begin{aligned}
\mathbb{P}\left(S_k(\alpha) \geq \frac{1}{2}\bar{\Delta}_k\right) &= \mathbb{P}\left(\min\{Z_k^{(s)}(\alpha),\ 0\} + Z_k^{(w)}(\alpha) \geq \frac{1}{2}\bar{\Delta}_k\right) \\
&\leq \mathbb{P}\left(Z_k^{(s)}(\alpha) + Z_k^{(w)}(\alpha) \geq \frac{1}{2}\bar{\Delta}_k\right) \\
&\leq \mathbb{P}\left(Z_k^{(s)}(\alpha) - \Delta_{\alpha,\alpha^{(s)}} \geq -\frac{1}{4}\bar{\Delta}_k\right) \\
&\quad + \mathbb{P}\left(\Delta_{\alpha,\alpha^{(s)}} \geq \frac{7}{8}\bar{\Delta}_k\right) + \mathbb{P}\left(Z_k^{(w)}(\alpha) \geq -\frac{1}{8}\bar{\Delta}_k\right). \qquad (47)
\end{aligned}$$

Using Hoeffding's concentration inequality with a union bound over the possible choices of $\alpha^{(s)}$, we have

$$\mathbb{P}\left(Z_k^{(s)}(\alpha) - \Delta_{\alpha,\alpha^{(s)}} \geq -\frac{1}{4}\bar{\Delta}_k\right) \leq (K-1)\exp\left(-\frac{\bar{\Delta}_k^2}{32}B_k\right). \qquad (48)$$

Since $\alpha \in F_k \subset E_k$, we have $\Delta^{(k)}_{\alpha,(\lceil |A_k|/4 \rceil)} \leq 0$. Therefore, by the shifted-rank version of Corollary B.2 with $N = |A_k| - 1$, we get

$$\begin{aligned}
\mathbb{P}\left(Z_k^{(w)}(\alpha) \geq -\frac{1}{8}\bar{\Delta}_k\right) &\leq \mathbb{P}\left(Z_k^{(w)}(\alpha) \geq \Delta^{(k)}_{\alpha,(\lceil |A_k|/4 \rceil)} - \frac{1}{8}\bar{\Delta}_k\right) \\
&\leq \log(T)\exp\left(-c \cdot \frac{\bar{\Delta}_k^2}{64\log(\lceil B_k/2 \rceil)}\left\lceil \frac{B_k}{2} \right\rceil\right) \\
&\leq \log(T)\exp\left(-c \cdot \frac{\bar{\Delta}_k^2}{128\log(B_k)}B_k\right). \qquad (49)
\end{aligned}$$

Since $\alpha \in F_k$, by definition of F_k given in (34), we have $\Delta_{\alpha,(s_\alpha)} \leq \bar{\Delta}_k \leq 0$. Therefore,

$$\mathbb{P}\left(\Delta_{\alpha,\alpha^{(s)}} \geq \frac{7}{8}\bar{\Delta}_k\right) \leq \mathbb{P}\left(\Delta_{\alpha,\alpha^{(s)}} \geq \frac{7}{8}\Delta_{\alpha,(s_\alpha)}\right).$$

If $K = 2$, the sole candidate opponent is returned and the following probability is zero. Otherwise, using Theorem B.5 and absorbing numerical constants into $C, c > 0$, we have

$$\begin{aligned}
\mathbb{P}\left(\Delta_{\alpha,\alpha^{(s)}} \geq \frac{7}{8}\bar{\Delta}_k\right) &\leq \mathbb{P}\left(\Delta_{\alpha,\alpha^{(s)}} \geq \frac{7}{8}\Delta_{\alpha,(s_\alpha)}\right) \\
&= \mathbb{P}\left(\Delta_{\alpha,\alpha^{(s)}} \geq \Delta_{\alpha,(s_\alpha)} - \frac{1}{8}\Delta_{\alpha,(s_\alpha)}\right) \\
&\leq C\log(2eKT)\exp\left(-c \cdot \frac{s_\alpha \Delta^2_{\alpha,(s_\alpha)}}{K\log(eT)\log^3(2K)}B_k\right). \qquad (50)
\end{aligned}$$

We conclude by plugging the bounds (48), (49) and (50) into (47)

$$\mathbb{P}\left(S_k(\alpha) \geq \frac{1}{2}\bar{\Delta}_k\right) \leq C(K + \log(T))\exp\left(-c'\ \min\left\{\frac{\Gamma_\alpha}{K\log(eT)\log^3(2K)}, \frac{\bar{\Delta}_k^2}{\log(B_k)}\right\}B_k\right),$$

where $c' := \frac{c}{1024}$. Now it remains to prove that

$$\min\left\{\frac{\Gamma_{F_k:\lceil |F_k|/4 \rceil}}{K\log^3(K)\log(T)},\ \frac{\bar{\Delta}_k^2}{\log(B_k)}\right\}B_k\ \geq c'\frac{T}{\log^4(K)\log(T)\ \sum_{i\in E_k}\frac{K}{\Gamma_i}}.$$

Recall Lemma H.2 gives

$$\frac{\left\lceil \frac{|F_k|}{4} \right\rceil}{\Gamma_{F_k:\lceil \frac{1}{4}|F_k| \rceil}} \leq \sum_{i \in F_k} \frac{1}{\Gamma_i} \leq \sum_{i \in E_k} \frac{1}{\Gamma_i} \,.$$

Therefore,

$$\frac{|F_k|}{4 \sum_{i \in E_k} \frac{1}{\Gamma_i}} \leq \Gamma_{F_k:\lceil \frac{1}{4}|F_k| \rceil} \,.$$

Hence, using the bound above and the definition of B_k we obtain

$$\begin{aligned}
\frac{\Gamma_{F_k:\lceil \frac{1}{4}|F_k| \rceil}}{K \log^3(K) \log(T)} B_k &\geq \frac{|F_k|}{4 \sum_{i \in E_k} \frac{1}{\Gamma_i}} \cdot \frac{1}{K \log^3(K) \log(T)} \cdot \frac{T}{2|A_k| \log_{8/7}(K)} \\
&= \frac{T}{8 \log^3(K) \log(T) \log_{8/7}(K)} \cdot \frac{1}{\sum_{i \in E_k} \frac{K}{\Gamma_i}} \cdot \frac{|F_k|}{|A_k|} \\
&\geq \frac{T}{138 \log^4(K) \log(T) \sum_{i \in E_k} \frac{K}{\Gamma_i}} \cdot \frac{|F_k|}{|A_k|} \,,
\end{aligned}$$

Recall that $|F_k| \geq \left\lceil \frac{7}{8}|E_k| \right\rceil \geq \left\lceil \frac{3}{16}|A_k| \right\rceil$. Therefore, the bound above gives

$$\frac{\Gamma_{F_k:\lceil \frac{1}{4}|F_k| \rceil}}{K \log^3(K) \log(T)} B_k \geq \frac{T}{736 \log^4(K) \log(T) \sum_{i \in E_k} \frac{K}{\Gamma_i}} \,. \tag{51}$$

Moreover, we have

$$\begin{aligned}
|A_k| \frac{1}{\bar{\Delta}_k^2} \leq 4|E_k| \cdot \frac{1}{\bar{\Delta}_k^2} &\leq 32 \cdot \left\lceil \frac{1}{8}|E_k| \right\rceil \frac{1}{\Delta^2_{E_k:\lceil \frac{7}{8}|E_k| \rceil}} \\
&\leq 32 \cdot \sum_{\alpha \in E_k} \frac{1}{\Delta^2_{\alpha,(s_\alpha)}} \,,
\end{aligned} \tag{52}$$

where the last inequality follows directly because the gaps with ranks at least $\lceil 7|E_k|/8 \rceil$ lie between $\bar{\Delta}_k$ and zero, so $\sum_{\alpha \in E_k} 1/\Delta^2_{\alpha,(s_\alpha)} \geq (|E_k| - \lceil 7|E_k|/8 \rceil + 1)/\bar{\Delta}_k^2 \geq \lceil |E_k|/8 \rceil / \bar{\Delta}_k^2$. Therefore, we have

$$\begin{aligned}
\frac{\bar{\Delta}_k^2}{\log(B_k)} B_k &\geq \frac{\bar{\Delta}_k^2}{\log(B_k)} \frac{T}{2|A_k| \log_{8/7}(K)} \\
&\geq \frac{1}{\sum_{\alpha \in E_k} \frac{1}{\Delta^2_{\alpha,(s_\alpha)}}} \cdot \frac{T}{64 \log(B_k) \log_{8/7}(K)} \,.
\end{aligned} \tag{53}$$

Therefore, combining (51) and (53), we get,

$$\begin{aligned}
&\mathbb{P}\left(S_k(\alpha) \geq \frac{1}{2} \bar{\Delta}_k \right) \\
&\qquad \leqslant \min\left\{ 1, C(\log(T) + K) \exp\left(-c' \, \min\left\{ \frac{\Gamma_{F_k:\lceil \frac{1}{4}|F_k| \rceil}}{K \log^3(K) \log(T)}, \, \frac{\bar{\Delta}_k^2}{\log(B_k)} \right\} B_k \right) \right\} \\
&\qquad \leq (\log(T) + K) \exp\left(-c_1 \, \frac{T}{\log^4(K) \log(T) \sum_{i \in E_k} \frac{K}{\Gamma_i}} \right) \,,
\end{aligned} \tag{54}$$

where we used $\log(T) \geq \log(K)$ and $\sum_{i \in E_k} K/\Gamma_i \geq \sum_{\alpha \in E_k} 1/\Delta^2_{\alpha,(s_\alpha)}$ in the last line. The numerical factor C is absorbed by decreasing the exponent constant: for $(\log(T) + K) \geq 2$, $\min\{1, C(\log(T) + K) e^{-bx}\} \leq l e^{-b'x}$ with a sufficiently small numerical $b' > 0$.

For the remainder of this proof, let $H \coloneqq H^{(0)}_{\text{explore}}(\boldsymbol{s})$ and $T_0 \coloneqq c_2 H \log^6(H)$. We have $H \geq 4K \geq 8$ and

$$\log(T_0) \leq \left(7 + \frac{\log(c_2)}{\log 8} \right) \log(H) \,.$$

Choose c_2 large enough that $T_0 \geq e^6$ and $T_0/(\log T_0)^6 \geq 2H/c_1$; such a universal choice exists since $c_2/(7+\log(c_2)/\log 8)^6 \to \infty$. Since $t/(\log t)^6$ is non-decreasing for $t \geq e^6$, for every $T \geq T_0$ we have $T/(\log T)^6 \geq 2H/c_1$. Using $\sum_{i \in E_k} K/\Gamma_i \leq H$ and $K \leq T$, it follows that

$$\frac{c_1 T}{\log^4(K)\log(T)\sum_{i \in E_k} K/\Gamma_i} \geq \frac{c_1 T}{H(\log T)^5} \geq 2\log T \, .$$

For sufficiently large c_2, we also have $K + \log T \leq T$ and $T \geq 18$. Hence (54) gives

$$\mathbb{P}\left(S_k(\alpha) \geq \frac{1}{2}\bar{\Delta}_k\right) \leq T e^{-2\log T} = \frac{1}{T} \leq \frac{1}{18} \, .$$

□

D Proof of Theorem 2.1

This routine, presented in Algorithm 4, serves as one of the two certification sub-procedures in the fixed-confidence algorithm. Given a confidence level δ, a query budget T, and a candidate Condorcet winner I, it sequentially tests whether one can certify—using at most T comparisons—that all pairwise gaps $(\Delta_{I,i})_{i \neq I}$ are positive with probability at least $1-\delta$. The budget is allocated uniformly across these gaps, and the procedure terminates as soon as either (i) a negative gap is detected, (ii) all gaps are certified positive, or (iii) the budget T is exhausted.

Algorithm 4 Test-CW

Input: $I \in [K], \delta, T$.
Initialize: $C = [K] \setminus \{I\}$, empirical means $\tilde{\Delta}_{I,j} = 0$ for $j \in C$, count variable $t \leftarrow 1$.
if $T < 8K \log_{8/7}(K)$ **then**
  Return **False**.
end if
Let $n \leftarrow 1 + \left\lceil \log_2\left(\frac{T}{8K\log(K)}\right)\right\rceil$.
$N_{I,j} \leftarrow 0$ for all $j \in [K]$ //Query count
while $C \neq \emptyset$ and $t \leq T$ **do**
  Sample duel (I, j) for $j \in \text{argmin}_{j \in C} N_{I,j}$ and update corresponding empirical means.
  $N_{I,j} \leftarrow N_{I,j} + 1$, $t \leftarrow t + 1$.
  `/* Check the sign of the gaps using concentration */`
  for $j \in C$ **do**
    if $N_{I,j} > 0$ **then**
      if $\tilde{\Delta}_{I,j} \geq \sqrt{\frac{\log\left(KN_{I,j}^2 \frac{n(n+1)}{\delta}\right)}{N_{I,j}}}$ **then**
        $C \leftarrow C \setminus \{j\}$.
      else if $\tilde{\Delta}_{I,j} \leq -\sqrt{\frac{\log\left(KN_{I,j}^2 \frac{n(n+1)}{\delta}\right)}{N_{I,j}}}$ **then**
        Return **False**.
      end if
    end if
  end for
end while
if $C = \emptyset$ **then**
  Return **True**
else
  Return **False**
end if

Both Algorithms 1 and 4 use the same counter

$$n = 1 + \left\lceil \log_2\left(\frac{T}{8K\log(K)}\right)\right\rceil .$$

On the doubling schedule $T^{(n)} = 8K\log(K)\,2^{n-1}$ of Algorithm 2, this is exactly the call number. Both sub-procedures return a false certification flag when $T < 8K\log_{8/7}(K)$. All samples used by a new call, by each score computation, and by Test-CW are fresh and independent of the preceding history. As in Section C, roundwise concentration bounds are proved conditionally on the complete pre-round history and then averaged; this conditioning is omitted from the notation.

D.1 Proof of δ-correctness

Let c_0 denote the absolute numerical constant corresponding to the one appearing in the upper bound of Corollary B.2. Theorem 2.1 states that Algorithm 2, with input $\delta \in (0, 1/6)$ and $c \geq 2/c_0$, outputs an arm different from the CW with probability at most δ. Let ψ_δ denote the output of Algorithm 2 when the input is δ. We will prove that

$$\mathbb{P}(\psi_\delta \neq i^*) \leq \delta\,.$$

To prove this claim, we introduce the following notation. In the n-th iteration (i.e., the n-th call to Algorithm 1), denote by $\bar{\alpha}^{(n)}, \phi_1^{(n)}, \phi_2^{(n)}, I^{(n)}$ and $T^{(n)}$ the corresponding values of $\bar{\alpha}, \phi_1, \phi_2, I$ and T, and let $\varphi^{(n)} := \phi_1^{(n)} \vee \phi_2^{(n)}$. For convenience define, for all $n \geq 1$,

$$\delta_n := \frac{\delta}{8K^2\log_{8/7}(K)\,\log(T^{(n)})\,n(n+1)}\,.$$

Let $S_k^{(n)}(\cdot)$ denote the score used at round k within the n-th call to Algorithm 1, and let $Z_k^{(w,n)}(\cdot)$ denote its weak component. Recall that each call to Algorithm 1 has at most $k_{\max} \leq \lceil \log_{8/7}(K) \rceil$ rounds. Finally, in the n-th call to TEST-CW (Algorithm 4) with inputs $(I^{(n)}, \delta, T^{(n)})$, let $\tilde{\Delta}_{i,j}(N)$ denote the empirical estimate based on the first N fresh samples of pair (i,j) in that call. For the concentration arguments, these independent sample sequences may be generated in advance, including their unused prefixes. Calls with $T^{(n)} < 8K\log_{8/7}(K)$ return `False` and do not contribute to the error event; the following bounds concern the remaining calls.

If $\psi_\delta \neq i^*$, then for some $n \geq 1$ the algorithm must have certified an incorrect candidate, namely $\{I^{(n)} \neq i^*\}$ and $\{\varphi^{(n)} = \texttt{True}\}$. Hence, by a union bound,

$$\begin{aligned}\mathbb{P}(\psi_\delta \neq i^*) &\leq \mathbb{P}\Big(\exists n \geq 1 :\ I^{(n)} \neq i^*,\ \varphi^{(n)} = \texttt{True}\Big) \\ &\leq \sum_{n=1}^{\infty} \mathbb{P}\Big(I^{(n)} \neq i^*,\ \phi_1^{(n)} = \texttt{True}\Big) + \sum_{n=1}^{\infty} \mathbb{P}\Big(I^{(n)} \neq i^*,\ \phi_2^{(n)} = \texttt{True}\Big)\,. \end{aligned} \tag{55}$$

We first bound the contribution of ϕ_1. On the event $\{I^{(n)} \neq i^*,\ \phi_1^{(n)} = \texttt{True}\}$, during the n-th run of Algorithm 1 the true Condorcet winner i^* must have been eliminated at some round $k < k_{\max}$ (otherwise the procedure would return $I^{(n)} = i^*$). By the definition of the selection $\bar{\alpha}^{(n)}$ and the condition $\phi_1^{(n)} = \texttt{True}$, this implies that for some $k < k_{\max}$,

$$S_k^{(n)}(i^*) < -\sqrt{\frac{2c\log(T^{(n)})\log(1/\delta_n)}{\lceil B_k^{(n)}/4 \rceil}}.$$

Using $S_k^{(n)}(i^*) = \min\{\hat{\Delta}_{i^*,u}^{(k,n)}, 0\} + Z_k^{(w,n)}(i^*)$ (for the opponent u queried at that round) and the fact that $\min\{x,0\} + y < -\eta$ implies $(x < -\eta/2)$ or $(y < -\eta/2)$, we get

$$\begin{aligned}\mathbb{P}\Big(I^{(n)} \neq i^*,\ \phi_1^{(n)} = \texttt{True}\Big) &\leq \sum_{k<k_{\max}} \mathbb{P}\left(i^* \in A_k^{(n)},\ S_k^{(n)}(i^*) < -\sqrt{\frac{2c\log(T^{(n)})\log(1/\delta_n)}{\lceil B_k^{(n)}/4 \rceil}}\right) \\ &\leq \sum_{k<k_{\max}} \sum_{u \neq i^*} \mathbb{P}\left(i^* \in A_k^{(n)},\ \hat{\Delta}_{i^*,u}^{(k,n)} < -\sqrt{\frac{c\log(T^{(n)})\log(1/\delta_n)}{2\lceil B_k^{(n)}/4 \rceil}}\right) \\ &\quad + \sum_{k<k_{\max}} \mathbb{P}\left(i^* \in A_k^{(n)},\ Z_k^{(w,n)}(i^*) < -\sqrt{\frac{c\log(T^{(n)})\log(1/\delta_n)}{2\lceil B_k^{(n)}/4 \rceil}}\right)\,. \end{aligned} \tag{56}$$

For both terms, condition on a pre-round history such that $i^* \in A_k^{(n)}$. For the first term, since $\Delta_{i^*,u} > 0$ for all $u \neq i^*$, we can center and apply Hoeffding's inequality: for $N = \lceil B_k^{(n)}/4 \rceil$,

$$\mathbb{P}\left(\hat{\Delta}_{i^*,u}^{(k,n)} < -\sqrt{\frac{c \log(T^{(n)}) \log(1/\delta_n)}{2N}} \right) \leq \mathbb{P}\left(\hat{\Delta}_{i^*,u}^{(k,n)} - \Delta_{i^*,u} < -\sqrt{\frac{\log(1/\delta_n)}{2N}} \right)$$
$$\leq \exp\left(-2N \cdot \frac{\log(1/\delta_n)}{2N} \right) \leq \delta_n \, . \tag{57}$$

For the second term in (56), the shifted-rank version of Corollary B.2 with $N = |A_k| - 1$ (with constant c_0) gives

$$\mathbb{P}\left(Z_k^{(w,n)}(i^*) < -\sqrt{\frac{c \log(T^{(n)}) \log(1/\delta_n)}{2\lceil B_k^{(n)}/4 \rceil}} \right) \leq \mathbb{P}\left(Z_k^{(w,n)}(i^*) < \Delta^{(k)}_{i^*,(\lceil |A_k|/8 \rceil)} - \sqrt{\frac{c \log(T^{(n)}) \log(1/\delta_n)}{2\lceil B_k^{(n)}/4 \rceil}} \right)$$
$$\leq \log\left(\left\lceil \frac{B_k^{(n)}}{2} \right\rceil \right) \exp\left(-c_0 \cdot \frac{c \log(T^{(n)}) \log(1/\delta_n)}{2 \log(\lceil B_k^{(n)}/2 \rceil)} \right)$$
$$\leq \log\left(\left\lceil \frac{B_k^{(n)}}{2} \right\rceil \right) \delta_n, \tag{58}$$

where, without loss of generality, $c_0 \leq 1$, and we used $c \geq 2/c_0$ and $B_k^{(n)} \leqslant T^{(n)}$. Averaging over the pre-round history and plugging (57) and (58) into (56), summing over $k < k_{\max}$ and using $\log(\lceil B_k^{(n)}/2 \rceil) \leq \log(T^{(n)})$ and $k_{\max} \leq \lceil \log_{8/7}(K) \rceil$, we obtain

$$\sum_{n=1}^{\infty} \mathbb{P}\Big(I^{(n)} \neq i^*, \ \phi_1^{(n)} = \texttt{True} \Big) \leq \sum_{n=1}^{\infty} \Big((k_{\max}(K-1))\delta_n + k_{\max} \log(T^{(n)}) \delta_n \Big)$$
$$\leq \sum_{n=1}^{\infty} \frac{\delta}{2n(n+1)} \leq \frac{\delta}{2}. \tag{59}$$

We now bound the contribution of ϕ_2 in (55). On the event $\{I^{(n)} \neq i^*, \ \phi_2^{(n)} = \texttt{True}\}$, the n-th call to TEST-CW returns `True` although $I^{(n)}$ is not the Condorcet winner. In particular, for some $N \geq 1$ the test must have accepted the comparison against i^*, meaning that

$$\tilde{\Delta}_{I^{(n)},i^*}(N) \geq \sqrt{\frac{\log\Big(K N^2 \frac{n(n+1)}{\delta} \Big)}{N}}.$$

Hence, by a union bound over $n \geq 1$, $N \geq 1$ and all $i \neq i^*$, and since $\Delta_{i,i^*} \leq 0$ when $i \neq i^*$,

$$\sum_{n=1}^{\infty} \mathbb{P}\Big(I^{(n)} \neq i^*, \ \phi_2^{(n)} = \texttt{True} \Big) \leq \sum_{n=1}^{\infty} \sum_{N=1}^{\infty} \sum_{i \neq i^*} \mathbb{P}\left(\tilde{\Delta}_{i,i^*}(N) \geq \sqrt{\frac{\log\Big(K N^2 \frac{n(n+1)}{\delta} \Big)}{N}} \right)$$
$$\leq \sum_{n=1}^{\infty} \sum_{N=1}^{\infty} \sum_{i \neq i^*} \mathbb{P}\left(\tilde{\Delta}_{i,i^*}(N) - \Delta_{i,i^*} \geq \sqrt{\frac{\log\Big(K N^2 \frac{n(n+1)}{\delta} \Big)}{N}} \right)$$
$$\leq \sum_{n=1}^{\infty} \sum_{N=1}^{\infty} \sum_{i \neq i^*} \frac{\delta}{4K \, N^2 \, n(n+1)} \ \leq \ \frac{\delta}{2} \, , \tag{60}$$

where the last inequality follows from Hoeffding's inequality and $\sum_{N \geq 1} 1/N^2 \leq 2$.

Finally, combining (55) with (59) and (60) yields

$$\mathbb{P}(\psi_\delta \neq i^*) \leq \frac{\delta}{2} + \frac{\delta}{2} = \delta \, ,$$

which concludes the proof.

D.2 Proof of Theorem 2.1 (sample complexity statement)

We build on the guarantees established for Algorithm 1 (fixed-budget elimination with certification) to prove the second statement of Theorem 2.1. We use the notation: for each $i \neq i^*$, $\Delta_{i,(1)} \leq \cdots \leq \Delta_{i,(K-1)}$ denotes the ordered list of gaps $(\Delta_{i,j})_{j\neq i}$, and $K_{i;<0} := \big|\{j : \Delta_{i,j} < 0\}\big|$. Fix any vector $\boldsymbol{s} = (s_1, \ldots, s_K)$ such that $1 \leq s_i \leq K_{i;<0}$, with s_i integer, for all $i \neq i^*$ (and $K_{i^*;<0} = 0$ by convention), and recall

$$H_{\text{certify}}(\boldsymbol{s}) = \sum_{i\neq i^*} \frac{1}{\Delta^2_{i,(s_i)}}, \qquad H^{(1)}_{\text{explore}}(\boldsymbol{s}) = \max_{i\neq i^*} \frac{K}{s_i \Delta^2_{i,(s_i)}}, \qquad H^{(0)}_{\text{explore}}(\boldsymbol{s}) = \sum_{i\neq i^*} \frac{K}{s_i \Delta^2_{i,(s_i)}} .$$

Let c_1 be the numerical constant in Theorem C.1 and let c_2, c_3 be the numerical constants in Lemma C.5. Let c_0 be the numerical constant in Corollary B.2, and assume $c \geq 2/c_0$ as in the statement of Theorem 2.1. Without loss of generality, take $c_0 \leq 1$, $0 < c_1 \leq 1/16$, $0 < c_2 \leq \log(8/7)(\log 2)^2/32$, and $c_3 \geq 1$; decreasing the constants in the exponents and increasing the warm-up constant preserves the preceding bounds. For concision, define

$$\begin{aligned}
G_{1,\delta} &:= \frac{32}{c_1} H_{\text{cw}} \log(K) \log\Big(\frac{32}{c_1}\frac{K H_{\text{cw}}}{\delta}\Big) \log\Big(c_1^{-1} H_{\text{cw}} \log(K/\delta)\Big) , \\
G_{2,\delta} &:= \frac{512c}{c_2} H_{\text{certify}}(\boldsymbol{s}) \, \log^3(K) \, \log\Big(\frac{32c}{c_2}\frac{K H^{(0)}_{\text{explore}}(\boldsymbol{s})}{\delta}\Big) \\
&\qquad \times \log\left(\frac{512c}{c_2} H_{\text{certify}}(\boldsymbol{s}) \, \log^3(K) \, \log\Big(\frac{32c}{c_2}\frac{K H^{(0)}_{\text{explore}}(\boldsymbol{s})}{\delta}\Big)\right) , \\
G_{3,\delta} &:= \frac{32c}{c_2} H^{(1)}_{\text{explore}}(\boldsymbol{s}) \, \log^4(K) \, \log\left(\frac{2Kk_{\max}}{\delta}\right) \log\left(\frac{32c}{c_2} H^{(1)}_{\text{explore}}(\boldsymbol{s}) \log^4(K) \log\left(\frac{2Kk_{\max}}{\delta}\right)\right) , \\
G_0 &:= \frac{2c_3}{c_2} \, H^{(0)}_{\text{explore}}(\boldsymbol{s}) \, \log^6\Big(\frac{2c_3}{c_2} H^{(0)}_{\text{explore}}(\boldsymbol{s})\Big) .
\end{aligned}$$

Algorithm 2 doubles the budget parameter T at each unsuccessful iteration; therefore, if one can show that whenever

$$T \in \Big[M_\delta, \, 2M_\delta\Big], \qquad \text{where } M_\delta := \min\Big\{G_{1,\delta}, \, G_{2,\delta} + G_{3,\delta} + G_0\Big\} , \tag{61}$$

the call to Algorithm 1 with inputs (δ, T, c) returns $\phi_1 \vee \phi_2 = \texttt{True}$ with probability at least $1-\delta$, then it follows that the total number of queries N_δ used by Algorithm 2 is at most a universal constant multiple of M_δ with probability at least $1-\delta$ (since the sum of a doubling schedule up to the first successful budget is at most 2 times that budget). We now verify this success probability for any T satisfying (61), distinguishing two regimes depending on which term attains the minimum.

Regime 1: $G_{1,\delta} \leq G_{2,\delta} + G_{3,\delta} + G_0$. Then, we focus on the regime $T \in [G_{1,\delta}, 2G_{1,\delta}]$. In this regime we certify correctness through the fixed-budget guarantee of Theorem C.1. Indeed, for $T \in [G_{1,\delta}, \, 2G_{1,\delta}]$, we have

$$\begin{aligned}
\mathbb{P}(I \neq i^*) &\leq 27K \log(K) \log(T) \exp\Big(-c_1 \frac{T}{\log(T)\log(K) H_{\text{cw}}}\Big) \\
&\leq 27K \log(K) \log(2G_{1,\delta}) \exp\Big(-c_1 \frac{G_{1,\delta}}{\log(2G_{1,\delta})\log(K) H_{\text{cw}}}\Big) .
\end{aligned} \tag{62}$$

We now use the explicit definition of $G_{1,\delta}$ and the crude upper bound

$$\log(2G_{1,\delta}) \leq 16 \log\Big(c_1^{-1} H_{\text{cw}} \log(K/\delta)\Big) , \tag{63}$$

which results from the expression of $G_{1,\delta}$, $K, H_{\text{cw}} \geq 2$ and $\delta \in (0, 1/6)$. Using the bound (63) and plugging it back in (62), we obtain

$$\begin{aligned}
\mathbb{P}\left(I \neq i^*\right) &\leq 432K \log(K) \log\Big(c_1^{-1} H_{\text{cw}} \log(K/\delta)\Big) \exp\Big(-2\log(32 c_1^{-1} K H_{\text{cw}}/\delta)\Big) \\
&\leq \frac{27 c_1 \delta^2 \log(K)\log(K/\delta)}{64 K H_{\text{cw}}} \leq \frac{\delta}{4} .
\end{aligned}$$

Here we used $\log x \le x$, $H_{\mathrm{cw}} \ge 4(K-1)$, $\log K \le K-1$, and $\delta \log(K/\delta) \le K/2$ for $K \ge 2$ and $\delta \in (0, 1/6)$. It remains to argue that, conditional on $I = i^*$, the auxiliary certification Test-CW (Algorithm 4) returns `True` with probability at least $1 - \delta/2$, hence overall $\mathbb{P}(\phi_1 \vee \phi_2 = \texttt{True}) \ge 1 - 3\delta/4 \ge 1 - \delta$ in this regime. Run Test-CW with inputs (i^*, δ, T) and let T_i be the number of comparisons allocated to pair (i^*, i). By construction, $\sum_{i \neq i^*} T_i \le T$. If Test-CW returns `False`, then either it exhausted the budget without eliminating all opponents, or it triggered a negative-deviation stopping rule. Formally, define

$$\mathcal{E}_1 := \Big\{ \sum_{i \neq i^*} T_i = \lfloor T \rfloor \Big\}, \qquad \mathcal{E}_2 := \Big\{ \exists j \neq i^*,\ \exists N \ge 1 :\ \tilde{\Delta}_{i^*,j}(N) \le -\sqrt{\tfrac{\log(KN^2\, \frac{n(n+1)}{\delta})}{N}} \Big\},$$

so that $\{\phi_2 = \texttt{False}\} \subseteq \mathcal{E}_1 \cup \mathcal{E}_2$. Since $\Delta_{i^*,j} \ge 0$ for all $j \neq i^*$, Hoeffding's inequality and a union bound give

$$\begin{aligned} \mathbb{P}(\mathcal{E}_2) &\le \sum_{j \neq i^*} \sum_{N \ge 1} \mathbb{P}\left(\tilde{\Delta}_{i^*,j}(N) - \Delta_{i^*,j} \le -\sqrt{\tfrac{\log(KN^2\, \frac{n(n+1)}{\delta})}{N}} \right) \\ &\le \sum_{j \neq i^*} \sum_{N \ge 1} \frac{\delta}{K\, n(n+1)\, N^2} \le \frac{2\delta}{n(n+1)} \le \frac{\delta}{4}\,, \end{aligned} \tag{64}$$

where we used $\sum_{N \ge 1} N^{-2} \le 2$ and $n \ge 4$, which follows from $T \ge 8K \log_{8/7}(K)$ and the common definition of n.

Next, for each $i \neq i^*$ define

$$\bar{T}_i := \left\lceil \frac{16}{\Delta_{i^*,i}^2} \log\left(\frac{32Kn(n+1)}{\delta \Delta_{i^*,i}^2} \right) \right\rceil .$$

Lemma D.1 below ensures that $\sum_{i \neq i^*} \bar{T}_i < \lfloor G_{1,\delta} \rfloor \le \lfloor T \rfloor$, hence

$$\begin{aligned} \mathbb{P}(\mathcal{E}_1) = \mathbb{P}\left(\sum_{i \neq i^*} T_i = \lfloor T \rfloor \right) &\le \mathbb{P}\left(\sum_{i \neq i^*} T_i \ge \lfloor G_{1,\delta} \rfloor \right) \le \mathbb{P}\big(\exists i \neq i^* : T_i > \bar{T}_i \big) \\ &\le \sum_{i \neq i^*} \mathbb{P}(T_i > \bar{T}_i)\,. \end{aligned} \tag{65}$$

If $T_i > \bar{T}_i$, then at time $N = \bar{T}_i$ the arm i was not eliminated, meaning

$$\tilde{\Delta}_{i^*,i}(\bar{T}_i) \ <\ \sqrt{\frac{\log\big(K\bar{T}_i^2\, \frac{n(n+1)}{\delta}\big)}{\bar{T}_i}}\,.$$

By Lemma D.1, the RHS is at most $\Delta_{i^*,i} - \sqrt{\frac{\log(K\bar{T}_i^2\, \frac{n(n+1)}{\delta})}{2\bar{T}_i}}$, hence

$$\begin{aligned} \mathbb{P}(T_i > \bar{T}_i) &\le \mathbb{P}\left(\tilde{\Delta}_{i^*,i}(\bar{T}_i) - \Delta_{i^*,i} < -\sqrt{\frac{\log(K\bar{T}_i^2\, \frac{n(n+1)}{\delta})}{2\bar{T}_i}} \right) \\ &\le \sum_{N \ge 1} \frac{\delta}{K\, n(n+1)\, N^2} \le \frac{2\delta}{K\, n(n+1)} \le \frac{\delta}{4K}\,, \end{aligned} \tag{66}$$

where the last line uses Hoeffding and a union bound over $N \ge 1$. Combining (65) and (66) yields $\mathbb{P}(\mathcal{E}_1) \le \delta/4$. Together with (64), we obtain $\mathbb{P}(\phi_2 = \texttt{False}) \le \delta/2$, hence $\mathbb{P}(\phi_2 = \texttt{True}) \ge 1 - \delta/2$ when $I = i^*$. This completes Regime 1.

Regime 2: $G_{1,\delta} > G_{2,\delta} + G_{3,\delta} + G_0$. Then $T \in [G_{2,\delta} + G_{3,\delta} + G_0,\ 2(G_{2,\delta} + G_{3,\delta} + G_0)]$. We show that, for such T, the certification variable ϕ_1 in Algorithm 1 remains `True` with probability at least $1 - \delta$. In this regime, we condition on an arbitrary pre-round history, including histories for which $i^* \notin A_k$. The second bound in Lemma C.5 remains valid under this conditioning: its

proof uses only $E_k \subseteq [K] \setminus \{i^*\}$, skew-symmetry, and the fresh independent samples for each arm, and does not require $i^* \in A_k$. Likewise, Lemma C.4 remains valid; in its case $|A_k| \in \{3, 4\}$, one simply allows $|F_k| \in \{2, 3, 4\}$, for which $|F_k| - 1 \geq \lceil |F_k|/3 \rceil$ still holds. Let $\bar{\alpha}_k$ be the arm ranked $|A_k| - \lceil |A_k|/8 \rceil + 1$ at round k according to scores $S_k(\cdot)$. Define

$$L_{k,\delta} := \sqrt{\frac{2c \log(T)}{\lceil B_k/4 \rceil} \log\Big(\frac{1}{\delta_{n,K}}\Big)}, \qquad \delta_{n,K} := \frac{\delta}{8K^2 \log_{8/7}(K) \log(T)\, n(n+1)} .$$

By the update rule for ϕ_1, the event $\{\phi_1 = \texttt{False}\}$ implies that for some $k < k_{\max}$, $S_k(\bar{\alpha}_k) \geq -L_{k,\delta}$. The definition of $\bar{\alpha}_k$ entails that at most $\lceil |A_k|/8 \rceil$ arms have score strictly smaller than $-L_{k,\delta}$, i.e.

$$\{S_k(\bar{\alpha}_k) \geq -L_{k,\delta}\} \subseteq \left\{ \sum_{\alpha \in A_k} \mathbb{1}\big(S_k(\alpha) < -L_{k,\delta}\big) \leq \left\lceil \frac{|A_k|}{8} \right\rceil \right\} . \tag{67}$$

We now relate $-L_{k,\delta}$ to the threshold $\frac{1}{2}\bar{\Delta}_k$ used in Lemma C.5. Recall the definitions (as in the proof of Theorem C.3): let

$$E_k := \Big\{ \alpha \in A_k : \ \Delta^{(k)}_{\alpha,(\lceil |A_k|/4 \rceil)} \leq 0 \Big\}, \qquad |E_k| \geq \lceil |A_k|/4 \rceil ,$$

and define the $7/8$-quantile $\bar{\Delta}_k := \Delta_{E_k:\lceil (7/8)|E_k| \rceil} \leq 0$ and the subset

$$F_k := \Big\{ \alpha \in E_k : \ \Delta_{\alpha,(s_\alpha)} \leq \bar{\Delta}_k \Big\} .$$

Since $1 \leq s_\alpha \leq K_{\alpha;<0}$ for every $\alpha \in E_k$, we have $\bar{\Delta}_k < 0$. Lemma D.2 ensures that for all $k < k_{\max}$ and all T in the present regime,

$$-L_{k,\delta} \ \geq \ \frac{1}{2}\, \bar{\Delta}_k . \tag{68}$$

We assume that $|A_k| \geq 3$, the case $|A_k| = 2$ is treated in the end. Using (68) inside (67) gives

$$\begin{aligned} \mathbb{P}(S_k(\bar{\alpha}_k) \geq -L_{k,\delta}) &\leq \mathbb{P}\left(\sum_{\alpha \in A_k} \mathbb{1}\Big(S_k(\alpha) < \tfrac{1}{2}\bar{\Delta}_k\Big) \leq \left\lceil \frac{|A_k|}{8} \right\rceil \right) \\ &= \mathbb{P}\left(\sum_{\alpha \in A_k} \mathbb{1}\Big(S_k(\alpha) \geq \tfrac{1}{2}\bar{\Delta}_k\Big) \geq |A_k| - \left\lceil \frac{|A_k|}{8} \right\rceil \right) . \end{aligned} \tag{69}$$

Since $|A_k| - \lceil |A_k|/8 \rceil = |A_{k+1}|$ and Lemma C.4 gives $|A_{k+1} \cap F_k| \geq \lceil |F_k|/3 \rceil$, the RHS of (69) is upper bounded by

$$\mathbb{P}\left(\sum_{\alpha \in F_k} \mathbb{1}\Big(S_k(\alpha) \geq \tfrac{1}{2}\bar{\Delta}_k\Big) \geq \left\lceil \frac{|F_k|}{3} \right\rceil \right) .$$

Lemma C.5 then yields, for a numerical constant $c_2 > 0$,

$$\mathbb{P}(S_k(\bar{\alpha}_k) \geq -L_{k,\delta}) \leq \exp\left(-c_2 \frac{T - c_3 H^{(0)}_{\text{explore}}(\boldsymbol{s}) \log^6(H^{(0)}_{\text{explore}}(\boldsymbol{s}))}{\log^4(K) \log(T)} \cdot \frac{1}{H^{(1)}_{\text{explore}}(\boldsymbol{s})} \right) . \tag{70}$$

Next, we use $T \geq G_{3,\delta} + G_0$ with Lemma D.3, which turns the last inequality into a bound on the exponent term of (70) leading to

$$\mathbb{P}(S_k(\bar{\alpha}_k) \geq -L_{k,\delta}) \leq \frac{\delta}{2K k_{\max}} . \tag{71}$$

which is the desired per-round bound.

Suppose that $|A_k| = 2$. Then $\bar{\alpha}_k$ is an arm with the smaller score. Thus $S_k(\bar{\alpha}_k) \geq -L_{k,\delta}$ implies that every arm in F_k has score at least $\frac{1}{2}\bar{\Delta}_k$. Consequently,

$$\begin{aligned}\mathbb{P}(S_k(\bar{\alpha}_k) \geq -L_{k,\delta}) &\leq \mathbb{P}\left(\sum_{\alpha \in F_k} \mathbb{1}\big(S_k(\alpha) \geq \tfrac{1}{2}\bar{\Delta}_k\big) \geq \left\lceil \frac{|F_k|}{3} \right\rceil\right) \\ &\leq \exp\left(-c_2 \frac{T - c_3 H^{(0)}_{\text{explore}}(\boldsymbol{s}) \log^6(H^{(0)}_{\text{explore}}(\boldsymbol{s}))}{\log^4(K)\log(T) H^{(1)}_{\text{explore}}(\boldsymbol{s})}\right) \\ &\leq \frac{\delta}{2K k_{\max}}\,,\end{aligned}$$

where we used the second bound in Lemma C.5, whose proof also applies when $|A_k| = 2$, and Lemma D.3.

Finally, since $\{\phi_1 = \texttt{False}\} \subseteq \bigcup_{k<k_{\max}} \{S_k(\bar{\alpha}_k) \geq -L_{k,\delta}\}$, a union bound and (71) yield

$$\mathbb{P}(\phi_1 = \texttt{False}) \;\leq\; \sum_{k=1}^{k_{\max}-1} \frac{\delta}{2K\,k_{\max}} \;\leq\; \frac{\delta}{2K} \leq \frac{\delta}{4}\,,$$

which completes Regime 2.

Conclusion. In either regime, for any T satisfying (61) the call to Algorithm 1 returns $\phi_1 \vee \phi_2 = \texttt{True}$ with probability at least $1-\delta$. Since Algorithm 2 doubles T until this event occurs, its total number of queries N_δ is at most a universal constant multiple of $M_\delta = \min\{G_{1,\delta}, G_{2,\delta} + G_{3,\delta} + G_0\}$ with probability at least $1-\delta$. Absorbing numerical constants into $\bar{c}_1, \bar{c}_2$ yields the stated bounds of Theorem 2.1. The Regime 2 argument also applies to every $T \geq G_{2,\delta} + G_{3,\delta} + G_0$, and its failure bound tends to zero as $T \to \infty$. Hence the doubling procedure terminates almost surely.

The lemmas below are technical.

Lemma D.1. *Consider the notation introduced in Regime 1 of the proof of Theorem 2.1, with $T \leq 2G_{1,\delta}$. Then we have*

$$\sum_{i \neq i^*} \bar{T}_i < \lfloor G_{1,\delta} \rfloor\,.$$

Moreover, for all $i \neq i^*$

$$2\sqrt{\frac{\log\left(K\bar{T}_i^2 \frac{n(n+1)}{\delta}\right)}{\bar{T}_i}} < \Delta_{i^*,i}\,.$$

Proof. Using $\lceil x \rceil \leq x+1$ and $\Delta_{i^*,i}^{-2} \leq H_{\text{cw}}$, we have

$$\sum_{i \neq i^*} \bar{T}_i \leq K - 1 + 16 H_{\text{cw}}\Big(\log(32KH_{\text{cw}}/\delta) + \log(n(n+1))\Big)\,. \tag{72}$$

The common counter definition gives $n+1 \leq \log_2 T$ for $T \geq 8K\log_{8/7}(K)$. Since $T \leq 2G_{1,\delta}$, (63) therefore implies

$$\log(n(n+1)) \leq 2\log\Big(\log_2(2G_{1,\delta})\Big) \leq 4\log\Big(c_1^{-1} H_{\text{cw}} \log(K/\delta)\Big)\,, \tag{73}$$

where we used $c_1 \leq 1/16$, $H_{\text{cw}} \geq 4(K-1)$ and $\delta < 1/6$ in the last inequality. Since $K - 1 \leq H_{\text{cw}}/4$ and $\log(c_1^{-1} H_{\text{cw}} \log(K/\delta)) \leq \log(32 c_1^{-1} K H_{\text{cw}}/\delta)$, we obtain

$$\begin{aligned}\sum_{i \neq i^*} \bar{T}_i &\leq \frac{H_{\text{cw}}}{4} + 80 H_{\text{cw}} \log\Big(32 c_1^{-1} K H_{\text{cw}}/\delta\Big) \\ &\leq 81 H_{\text{cw}} \log\Big(32 c_1^{-1} K H_{\text{cw}}/\delta\Big) < G_{1,\delta} - 1 \leq \lfloor G_{1,\delta} \rfloor\,.\end{aligned}$$

This proves the first claim.

For the second claim, the definition of $\bar{T}_i$ and $|\Delta_{i^*,i}| \le 1/2$ give

$$\bar{T}_i \le \frac{17}{\Delta_{i^*,i}^2} \log\left(\frac{32Kn(n+1)}{\delta\Delta_{i^*,i}^2}\right).$$

Consequently,

$$\log\left(K\bar{T}_i^2 \frac{n(n+1)}{\delta}\right) < 4\log\left(\frac{32Kn(n+1)}{\delta\Delta_{i^*,i}^2}\right) \le \frac{\Delta_{i^*,i}^2 \bar{T}_i}{4}.$$

Indeed, the first strict inequality follows by substituting the preceding upper bound on $\bar{T}_i$ and using $\log^2 x < x$ for $x \ge 256$; here $32Kn(n+1)/(\delta\Delta_{i^*,i}^2) > 256$. This is equivalent to the second claim. □

Lemma D.2. *Consider the notation introduced in the proof of Theorem 2.1. If $\bar{\Delta}_k < 0$ and*

$$T \ge G_{2,\delta} + G_0 ,$$

then

$$-L_{k,\delta} \ge \frac{1}{2}\bar{\Delta}_k .$$

Proof. Let

$$\delta_{n,K} = \frac{\delta}{8K^2 \log_{8/7}(K) \log(T)\, n(n+1)}.$$

Assume $\bar{\Delta}_k < 0$. Since $L_{k,\delta} \ge 0$, the inequality $-L_{k,\delta} \ge \frac{1}{2}\bar{\Delta}_k$ is equivalent to $L_{k,\delta} \le -\frac{1}{2}\bar{\Delta}_k$, i.e.

$$L_{k,\delta}^2 \;\le\; \frac{\bar{\Delta}_k^2}{4}. \tag{74}$$

Recalling

$$L_{k,\delta}^2 = \frac{2c\log(T)}{\lceil B_k/4 \rceil} \log\Big(\frac{1}{\delta_{n,K}}\Big), \quad B_k = \left\lfloor \frac{T}{|A_k| \log_{8/7}(K)} \right\rfloor ,$$

and using $\lceil B_k/4 \rceil \ge B_k/4 \ge T/(8|A_k|\log_{8/7}(K))$, we see that (74) is implied by

$$\log(T)\, \log\Big(\frac{1}{\delta_{n,K}}\Big) \;\le\; \frac{\bar{\Delta}_k^2}{64c} \frac{T}{|A_k| \log_{8/7}(K)} . \tag{75}$$

Next, using the inequality (52)

$$\frac{|A_k|}{\bar{\Delta}_k^2} \;\le\; 32 \sum_{\alpha \in E_k} \frac{1}{\Delta_{\alpha,(s_\alpha)}^2} ,$$

and the fact that $i^* \notin E_k$ (if $i^* \in A_k$, its restricted gaps are all positive, and otherwise it is absent from $E_k \subseteq A_k$), we have $\sum_{\alpha \in E_k} \frac{1}{\Delta_{\alpha,(s_\alpha)}^2} \le H_{\mathrm{certify}}(\boldsymbol{s})$. Hence

$$\frac{\bar{\Delta}_k^2}{|A_k|} \;\ge\; \frac{1}{32\, H_{\mathrm{certify}}(\boldsymbol{s})} . \tag{76}$$

Moreover, the common counter definition gives $n+1 \le \log_2 T$. Hence

$$\begin{aligned}\log\Big(\frac{1}{\delta_{n,K}}\Big) &= \log\Big(\frac{8K^2 \log_{8/7}(K)}{\delta}\Big) + \log\log(T) + \log(n(n+1)) \\ &\le \log\Big(\frac{8K^2 \log_{8/7}(K)}{\delta}\Big) + 3\log\log(T) + 4 . \end{aligned} \tag{77}$$

Combining (76), (77) and (75), it suffices to show

$$\log(T) \left[\log\Big(\frac{8K^2 \log_{8/7}(K)}{\delta}\Big) + 3\log\log(T) + 4\right] \le \frac{T}{2048c \log_{8/7}(K) H_{\mathrm{certify}}(\boldsymbol{s})} . \tag{78}$$

Since $T \ge G_{2,\delta} + G_0$, this follows from Lemma D.4, which concludes the proof. □

Lemma D.3. *Suppose that $T \geq G_0 + G_{3,\delta}$. Then we have*

$$\exp\left(-c_2 \frac{T - c_3 H^{(0)}_{\textit{explore}}(\boldsymbol{s}) \log^6(H^{(0)}_{\textit{explore}}(\boldsymbol{s}))}{\log^4(K)\log(T)} \frac{1}{H^{(1)}_{\textit{explore}}(\boldsymbol{s})}\right) \leq \frac{\delta}{2Kk_{\max}} .$$

Proof. The desired inequality is equivalent to

$$\frac{T - c_3 H^{(0)}_{\text{explore}}(\boldsymbol{s}) \log^6(H^{(0)}_{\text{explore}}(\boldsymbol{s}))}{\log(T)} \;\geq\; \frac{1}{c_2} H^{(1)}_{\text{explore}}(\boldsymbol{s}) \log^4(K) \log\left(\frac{2Kk_{\max}}{\delta}\right) . \tag{79}$$

Define

$$A \coloneqq 32 \frac{c}{c_2} H^{(1)}_{\text{explore}}(\boldsymbol{s}) \log^4(K) \log\left(\frac{2Kk_{\max}}{\delta}\right) .$$

By assumption,

$$T \;\geq\; 2\frac{c_3}{c_2} H^{(0)}_{\text{explore}}(\boldsymbol{s}) \log^6(H^{(0)}_{\text{explore}}(\boldsymbol{s})) \;+\; A\log(A) .$$

Since $c_2 \leq 1$ (as is the case for the numerical constant c_2 coming from the preceding bounds), the first term implies $T \geq 2c_3 H^{(0)}_{\text{explore}}(\boldsymbol{s}) \log^6(H^{(0)}_{\text{explore}}(\boldsymbol{s}))$, hence

$$T - c_3 H^{(0)}_{\text{explore}}(\boldsymbol{s}) \log^6(H^{(0)}_{\text{explore}}(\boldsymbol{s})) \geq \frac{T}{2}. \tag{80}$$

Next, the function $f(x) := x/\log x$ for $x > e$ is increasing. Moreover, we have $\log(2Kk_{\max}/\delta) > 1$ and $H^{(1)}_{\text{explore}}(\boldsymbol{s}) \geq 4$ (since $|\Delta|_{i,(s_i)} \leq \frac{1}{2}$ and $s_i \leq K$), and with $c \geq 1$ and $c_2 \leq 1$ this yields $A \geq 32 \cdot 4 \cdot (\log 2)^4 \cdot \log 4 > e$. Therefore, $A \log A > e$ and in particular $\log(A \log A) > 0$.

Since $T \geq A \log A$ and f is increasing on (e, ∞), we obtain

$$\frac{T}{\log T} \;=\; f(T) \;\geq\; f(A\log A) \;=\; \frac{A\log A}{\log(A\log A)}. \tag{81}$$

Finally, because $A > e$, we have $\log(A\log A) = \log A + \log\log A \leq \log A + \log A = 2\log A$, and therefore

$$\frac{A\log A}{\log(A\log A)} \;\geq\; \frac{A\log A}{2\log A} \;=\; \frac{A}{2}. \tag{82}$$

Combining (80), (81), and (82) gives

$$\frac{T - c_3 H^{(0)}_{\text{explore}}(\boldsymbol{s}) \log^6(H^{(0)}_{\text{explore}}(\boldsymbol{s}))}{\log T} \;\geq\; \frac{1}{2} \cdot \frac{T}{\log T} \;\geq\; \frac{1}{2} \cdot \frac{A}{2} \;=\; \frac{A}{4}.$$

By the definition of A,

$$\frac{A}{4} \;=\; 8\,\frac{c}{c_2}\, H^{(1)}_{\text{explore}}(\boldsymbol{s}) \log^4(K)\; \log(2Kk_{\max}/\delta) \;\geq\; \frac{1}{c_2}\, H^{(1)}_{\text{explore}}(\boldsymbol{s}) \log^4(K)\; \log(2Kk_{\max}/\delta),$$

where we used $c \geq 1$. This proves (79), and hence

$$\exp\left(-c_2\, \frac{T - c_3 H^{(0)}_{\text{explore}}(\boldsymbol{s}) \log^6(H^{(0)}_{\text{explore}}(\boldsymbol{s}))}{\log^4(K)\log(T)\, H^{(1)}_{\text{explore}}(\boldsymbol{s})}\right) \leq e^{-\log(2Kk_{\max}/\delta)} = \frac{\delta}{2Kk_{\max}} .$$

□

Lemma D.4. *Let $K \geq 2$, $\delta \in (0,1)$ and $T \geq G_0 + G_{2,\delta}$. Then*

$$\log(T)\left[\log\Big(\frac{8K^2 \log_{8/7}(K)}{\delta}\Big) + 3\log\log(T) + 4\right] \leq \frac{T}{2048c \log_{8/7}(K) H_{\textit{certify}}(\boldsymbol{s})} .$$

Proof. Let $L := \log_{8/7}(K)$, $H := H_{\mathrm{certify}}(\boldsymbol{s})$, and set

$$M := 2048cLH, \qquad B := \log\Big(\frac{32c}{c_2}\frac{KH^{(0)}_{\mathrm{explore}}(\boldsymbol{s})}{\delta}\Big) .$$

Since $H \le H^{(0)}_{\mathrm{explore}}(\boldsymbol{s})$, $H^{(0)}_{\mathrm{explore}}(\boldsymbol{s}) \ge 4K$, $c \ge 1$ and $c_2 \le \log(8/7)(\log 2)^2/32$, we have

$$B \ge 8, \qquad \log M \le B, \qquad \log\Big(\frac{8K^2L}{\delta}\Big) \le 2B .$$

Moreover,

$$\frac{512c}{c_2} H \log^3(K) \ge 8M ,$$

so the definition of $G_{2,\delta}$ gives

$$T \ge G_{2,\delta} \ge T_0 := 8MB\log(8MB) \ge e^2 .$$

Set

$$u(T) := \log(T)\left[\log\Big(\frac{8K^2L}{\delta}\Big) + 3\log\log T + 4\right] .$$

A direct differentiation shows that $u(T)/T$ is decreasing on $[e^2, \infty)$. Hence, for $T \ge T_0$,

$$u(T) \le \frac{T}{T_0} u(T_0) .$$

Since $B \ge 8$ and $\log M \le B$, we have

$$\log(8MB) \le 2B, \qquad \log T_0 \le 2\log(8MB) \le 4B, \qquad \log\log T_0 \le \log(4B) \le B/2 .$$

Therefore,

$$\begin{aligned} u(T_0) &\le 2\log(8MB)\big(2B + 3B/2 + 4\big) \\ &\le 8B\log(8MB) = \frac{T_0}{M} . \end{aligned}$$

Combining these inequalities yields $u(T) \le T/M$, which is the desired bound. □

E Minimax Fixed-Budget Lower Bounds

In this section, we establish lower bounds for the fixed-budget CW identification. In some way, these are the counterparts of the results of Section 3, only that we only derive minimax bounds, similar to Corollary 3.3.

We specify a class of distributions, parametrized by the row of the Condorcet Winner. Let $\underline{\Delta} = (\Delta_1, \Delta_2, \dots, \Delta_K)$, with $\Delta_1 = 0$ and $\Delta_i \in (0, 1/4)$ for $i \ne 1$. Define $\mathbb{D}^{(1)}(\underline{\Delta})$ as the set of dueling feedback distributions whose gap matrix $\boldsymbol{\Delta} \in \mathbb{D}_{\mathrm{cw}}$ is such that the row of the Condorcet winner i^*, namely $\Delta_{i^*,\cdot}$, is equal to $\underline{\Delta}$ up to a permutation σ. Formally,

$$\mathbb{D}^{(1)}(\underline{\Delta}) := \Big\{\boldsymbol{\Delta} \in \mathbb{D}_{\mathrm{cw}} : \exists\, \sigma \in \mathfrak{S}_K \text{ s.t. } i^*_{\boldsymbol{\Delta}} = \sigma^{-1}(1), \text{ and } \Delta_{i^*,\cdot} = \sigma(\underline{\Delta})\Big\}, \tag{83}$$

where $\mathfrak{S}_K$ is the set of permutations on $[K]$, and for any $x \in \mathbb{R}^K$, $\sigma(x) = (x_{\sigma(i)})_{i\in[K]}$. The following result is the counterpart of Theorem 3.1.

Theorem E.1. *Let $K \ge 2$ and $T \in \mathbb{N}^*$ and consider any vector $\underline{\Delta}$ with $\Delta_1 = 0$, and $\Delta_i \in (0, 1/4)$ for $i \ne 1$. For any algorithm A with a fixed budget T, one has*

$$\max_{\boldsymbol{\Delta}\in\mathbb{D}^{(1)}(\underline{\Delta})} \mathbb{P}_{\boldsymbol{\Delta},A}\left(\hat{i}_T \ne i^*\right) \ge \frac{1}{4}\exp\left(-22\frac{T}{H_{\mathrm{cw}}}\right) ,$$

where $H_{\mathrm{cw}} = \sum_{i=2}^K \frac{1}{\Delta_i^2}$.

Remarks. *This theorem reveals that one can exhibit a matrix Δ for which the the exponential error decay scaling as $\exp(-T/H_{\mathrm{cw}})$ in Theorem C.1 is tight. The proof can be found in Appendix G.1.*

We now turn to the counterpart of Theorem 3.2. Consider the following class of environments, for which the quantities $(\boldsymbol{s}^*_{\boldsymbol{\Delta}}, \Delta_{(\boldsymbol{s}^*)})$, defined in Section 3 are equal to a given couple $(\underline{\Delta}, \underline{s})$ up to a permutation of the arms. By convention, we extend $\boldsymbol{s}^*_{\boldsymbol{\Delta}} = (s^*_i)_{i \neq i^*}$ into a K-dimensional vector by fixing as 0 the i^*-th entry, that is $s^*_{i^*} := 0$. We proceed similarly for $\Delta_{(s^*)}$.

$$\mathbb{D}^{(2)}(\underline{\Delta}, \underline{s}) := \left\{ \boldsymbol{\Delta} \in \mathbb{D}_{\mathrm{cw}} : \exists \sigma \in \mathfrak{S}_K \text{ s.t.} \boldsymbol{s}^*_{\boldsymbol{\Delta}} = \sigma(\underline{s}) \text{ and } \boldsymbol{\Delta}_{(s^*)} = -\sigma(\underline{\Delta}) \right\}. \tag{84}$$

Theorem E.2. *Let $T \in \mathbb{N}^*$, K a multiple of 8, $\underline{\Delta} = (\Delta_1, \dots, \Delta_K)$ with $\Delta_1 = 0$ and $\Delta_i \in (0, 1/4)$ for $i \neq 1$, and $\underline{s} = (s_1, \dots, s_K)$ with $s_1 = 0$, $1 \leqslant s_i \leqslant K/8$ for $i = 2, \dots, K$. Then, any fixed-budget algorithm A satisfies*

$$\sup_{\boldsymbol{\Delta} \in \mathbb{D}^{(2)}(\underline{\Delta}, \underline{s})} \mathbb{P}_{A, \boldsymbol{\Delta}}(\hat{i}_T \neq i^*(\boldsymbol{\Delta})) \geqslant \frac{1}{4} \exp\left(-5 \frac{T}{\max_{i=2}^{K/2} \frac{K}{s_i \Delta_i^2}} \right). \tag{85}$$

If, additionally $\Delta_i = \Delta > 0$ and $s_i = s \in [K/8]$ for all $i \in \{2, \dots, K/2\}$, then

$$\sup_{\boldsymbol{\Delta} \in \mathbb{D}^{(2)}(\underline{\Delta}, \underline{s})} \mathbb{P}_{\boldsymbol{\Delta}, A}(\hat{i}_T \neq i^*(\boldsymbol{\Delta})) \geqslant \frac{1}{4} \exp\left(-10 \frac{T\Delta^2}{K} \right), \tag{86}$$

$$\sup_{\boldsymbol{\Delta} \in \mathbb{D}^{(2)}(\underline{\Delta}, \underline{s})} \mathbb{P}_{\boldsymbol{\Delta}, A}(\hat{i}_T \neq i^*(\boldsymbol{\Delta})) \geqslant \frac{1}{2} - \sqrt{37 \frac{s\Delta^2}{K^2} T}. \tag{87}$$

Remarks. *Equation* (87) *shows that for matrices where half the rows equal (up to permutation) a vector with s entries at $-\Delta$ and the rest near zero, any algorithm must pay $\frac{K^2}{s\Delta^2}$ to reach a constant success probability. This is aligned with the probability-independent cost $H^{(0)}_{\mathrm{explore}}(\underline{s})$) that suffers Algorithm 1 –see Theorem C.1.*

Equation (85) *establishes probability of error is no smaller than $\exp(-T/H^{(1)}_{\mathrm{explore}}(\underline{s}))$, while* (86) *implies that the quantity $\exp(-T/H_{\mathrm{certify}}(\boldsymbol{s}))$ is required, at least for some specific highly structured matrices.*

Comparison with Fixed Confidence lower bounds. The proofs for fixed-confidence and fixed-budget settings are remarkably similar. This reflects the strong connection between fixed-budget algorithms and fixed-confidence algorithms with high-probability budget bounds. However, fixed-budget minimax lower bounds cannot be directly deduced from the instance-dependent fixed confidence lower bounds derived earlier. We conjecture that obtaining instance-dependent lower bounds in the fixed-budget setting is considerably more challenging—if not outright impossible.

The fundamental reason lies in the nature of the algorithms themselves. Any δ-correct fixed-confidence algorithm must incorporate an internal stopping criterion that checks that the identified arm $\hat{i}$ is indeed the CW, under the assumption that such a unique CW exists in $\mathbb{D}_{\mathrm{cw}}$. Consider now what happens when applying this algorithm to a modified environment with two "weak" Condorcet winner candidates i^*_1 and i^*_2, where both rows satisfy $\Delta_{i^*_1,\cdot} \geqslant \mathbf{0}$ and $\Delta_{i^*_2,\cdot} \geqslant \mathbf{0}$. The algorithm cannot distinguish between these candidates with probability $1 - \delta$, forcing an infinite expected stopping time as the verification step never confidently resolves the ambiguity.

Fixed-budget algorithms fundamentally lack such stopping rules, rendering this infinite-budget argument inapplicable. Instead, our lower bounds rely on symmetry arguments: for matrices where multiple suboptimal arms have identical "difficulty profiles" (i.e., identical multisets of negative gaps), any reasonable algorithm must select uniformly at random among the ambiguous best arms. Deriving such lower bounds requires constructing highly symmetric matrices that exploit this randomization, a significantly more restrictive condition than the instance-dependent constructions used in fixed confidence.

In Sections F and G, we propose two complete constructions respectively for each setting. The proofs for fixed-budget results are similar to the fixed-confidence ones, except that they are applied to specific highly symmetric matrices for the reasons described above.

F Proofs of Section 3

In this section, we provide all proofs for the instance-dependent fixed-confidence lower bounds.

We begin with a short roadmap in Subsection F.1 describing the classical change-of-measure arguments underlying all our constructions (along the way, we fix some notation). In Subsection F.2, we prove Proposition 3.1, separating the expected budget bound (Subsection F.2.1) from the high-probability quantile bound (Subsection F.2.2).

In Subsection F.3, we explain the construction leading to Theorem 3.2 and state a more precise formulation in Theorem F.2, proved in Subsection F.4. Corollary 3.3 follows in Subsection F.5.

Finally, Subsection F.6 discusses lower bounds preserving CW row structure.

F.1 Roadmap on change-of-measure lower bounds

All our proofs follow a common three-step structure.

First step: reference and alternative instances. We fix a *reference instance* $\mathbf{\Delta} \in \mathbb{D}_{\mathrm{cw}}$, that is, a gap matrix admitting a (unique) Condorcet winner $i^*(\mathbf{\Delta})$. In the fixed-confidence regime, our lower bounds are instance-dependent, so all constructions are built directly from the given $K \times K$ matrix

$$\mathbf{\Delta} = (\Delta_{i,j})_{i,j\in[K]}.$$

For some results (e.g., Theorem 3.2), we also consider a local class of instances obtained from $\mathbf{\Delta}$ by permuting the negative entries of each row, while preserving prescribed structural features (CW, sign structure, multiset of negative entries, effective sparsity, and so on). This leads to families $\{\mathbf{\Delta}^{(\pi)}\}_\pi$ indexed by permutations π, but the reference object remains $\mathbf{\Delta}$.

For each suboptimal arm $k \neq i^*$, we construct an *alternative instance*

$$\mathbf{\Delta}^{(k)}$$

in such a way that i^* is no longer the (strong) Condorcet winner, while k becomes the CW or at least a *weak* CW, in the sense that the k-th row of $\mathbf{\Delta}^{(k)}$ contains only non-negative entries. A typical construction consists in modifying only the k-th row of $\mathbf{\Delta}$, for example by setting all its negative entries to a small constant $\epsilon \geq 0$, and updating the k-th column so as to preserve symmetry. In more refined arguments, we first permute the negative entries within each row.

Second step: total variation. We then exploit the properties of a given algorithm A in order to exhibit, for each $k \neq i^*$, a separating event B_k on which the two laws assign very different probabilities,

$$\mathbb{P}_{\mathbf{\Delta},A}(B_k) \text{ is large (typically } \geq 1-\delta), \qquad \mathbb{P}_{\mathbf{\Delta}^{(k)},A}(B_k) \text{ is small (typically } \leq \delta).$$

Here $\mathbb{P}_{\mathbf{\Delta},A}$ denotes the law of all observations (and internal randomness) when algorithm A interacts with environment $\mathbf{\Delta}$. A natural choice, when A is δ-correct for CW identification, is the event $\{\hat{i} = k\}$, where $\hat{i}$ is the recommendation output by A. For quantile (high-probability) lower bounds, we additionally introduce the $(1-\delta)$-quantile χ of the budget N_δ under $\mathbb{P}_{\mathbf{\Delta},A}$, and we consider events such as $\{N_\delta \leq \chi\}$ or their intersections with identification events. The precise choice of B_k varies from theorem to theorem, but the goal is always to produce a set on which the two laws $\mathbb{P}_{\mathbf{\Delta},A}$ and $\mathbb{P}_{\mathbf{\Delta}^{(k)},A}$ have very different probabilities, thereby enforcing a large total-variation distance:

$$\mathrm{TV}\big(\mathbb{P}_{\mathbf{\Delta},A}, \mathbb{P}_{\mathbf{\Delta}^{(k)},A}\big) \;\geq\; \big|\mathbb{P}_{\mathbf{\Delta},A}(B_k) - \mathbb{P}_{\mathbf{\Delta}^{(k)},A}(B_k)\big|.$$

The total-variation distance is then controlled from above through a standard data-processing inequality: we use either Pinsker's inequality, the Bretagnolle–Huber inequality, or a Fano-type inequality (see Lemma H.4) to relate TV to the Kullback–Leibler divergence. For example,

$$\mathrm{TV}(P,Q) \;\leq\; \sqrt{\tfrac{1}{2}\,\mathrm{KL}(P,Q)}, \qquad \text{or} \qquad 1-\mathrm{TV}(P,Q) \;\geq\; \tfrac{1}{2}\exp\big(-\mathrm{KL}(P,Q)\big).$$

The choice depends on the error-probability regime: Bretagnolle–Huber is convenient in the very small-δ regime, while Pinsker is often sharper in moderate-error regimes.

Third step: decomposition and control of the KL divergence. The last step is to decompose the Kullback–Leibler divergence between the laws induced by A under the two environments. Let $N_{i,j}$ denote the total number of observed duels of the ordered pair (i, j) between time 1 and the stopping time N_δ, and let $N_{\{i,j\}} = N_{i,j} + N_{j,i}$ be the total number of observations of the unordered pair $\{i, j\}$. A standard KL-decomposition for adaptive bandit algorithms (see, e.g., Lattimore and Szepesvári, 2020, Lemma 15.1) yields

$$\mathrm{KL}\big(\mathbb{P}_{\boldsymbol{\Delta},A}, \mathbb{P}_{\boldsymbol{\Delta}^{(k)},A}\big) \;=\; \sum_{1\le i<j\le K} \mathbb{E}_{\boldsymbol{\Delta},A}\big[N_{\{i,j\}}\big]\, \mathrm{KL}\big(\nu_{\Delta_{i,j}}, \nu_{\Delta^{(k)}_{i,j}}\big),$$

where $\nu_{\Delta_{i,j}}$ denotes the Bernoulli distribution with parameter $1/2 + \Delta_{i,j}$. In all our constructions, $\boldsymbol{\Delta}$ and $\boldsymbol{\Delta}^{(k)}$ differ only on a small set of pairs (typically those involving arm k), so the sum reduces to those indices. The Bernoulli KL-divergence admits the classical upper bound, for $p, q \in (0, 1)$,

$$\mathrm{kl}(p, q) \;\le\; \frac{(p-q)^2}{q(1-q)},$$

which, in our setting, leads to bounds of the form

$$\mathrm{KL}\big(\mathbb{P}_{\boldsymbol{\Delta},A}, \mathbb{P}_{\boldsymbol{\Delta}^{(k)},A}\big) \;\lesssim\; \sum_{\{i,j\}\text{ modified}} \mathbb{E}_{\boldsymbol{\Delta},A}[N_{\{i,j\}}]\, \big(\Delta_{i,j} - \Delta^{(k)}_{i,j}\big)^2.$$

Combining this upper bound with the lower bound on total variation from the third step yields a constraint that any δ-correct algorithm A must satisfy. Rearranging these inequalities then produces a lower bound on the sample complexity, typically expressed in terms of the expected budget $\mathbb{E}_{\boldsymbol{\Delta},A}[N_\delta]$ or on the $(1-\delta)$-quantile of N_δ, and involving the instance-dependent hardness parameters (such as $\Delta_{i,(1)}$ or $\|\Delta_i^-\|_2^2/K_{i;<0}$).

F.2 Proof of Proposition 3.1

We prove separately the bound in expectation (paragraph F.2.1) and the bound on the $(1-\delta)$-quantile of N_δ (paragraph F.2.2). The two arguments share the same change-of-measure structure (see Section F.1) and differ only in the way we exploit either the identification rule or the stopping rule to construct a separating event. The expectation bound 88 already appears as Theorem 5.2 in Haddenhorst et al. [2021], but we provide here a proof for completeness.

F.2.1 Bound in expectation from Proposition 3.1

Fix $K \geqslant 2$ and $\delta \in (0, 1)$. Consider an algorithm A that is δ-correct over the entire class $\mathbb{D}_{\mathrm{cw}}$. We fix any matrix $\boldsymbol{\Delta} \in \mathbb{D}_{\mathrm{cw}}$ such that $i^* = 1$ and prove the following expected lower bound that holds for this specific instance:

$$\mathbb{E}_{\boldsymbol{\Delta},A}[N_\delta] \;\geqslant\; \frac{1}{4} \sum_{i\neq i^*} \frac{\log\big(\frac{1}{4\delta}\big)}{\Delta^2_{i,(1)}}\,. \tag{88}$$

Sketch of proof. We follow the three-step roadmap of Section F.1. *(i) Reference and alternative instances.* Fix any $\boldsymbol{\Delta} \in \mathbb{D}_{\mathrm{cw}}$ with CW $i^* = 1$; this is our reference instance. For each suboptimal arm $k \neq 1$, we construct an alternative instance $\boldsymbol{\Delta}^{(k)}$ by lifting all non-positive entries in row k to a small constant $\epsilon > 0$ (and adjusting the k-th column to preserve symmetry), so that k becomes the CW in $\boldsymbol{\Delta}^{(k)}$. *(ii) Separating event and total variation.* Since the algorithm is δ-correct for CW identification, it must distinguish $\boldsymbol{\Delta}$ from each $\boldsymbol{\Delta}^{(k)}$ with error at most δ when deciding between i^* and k as CW. Using the test event $B_k = \{\hat{i} = k\}$, we obtain that the total-variation distance between $\mathbb{P}_{\boldsymbol{\Delta},A}$ and $\mathbb{P}_{\boldsymbol{\Delta}^{(k)},A}$ is at least $1 - 2\delta$, which, via the Bretagnolle–Huber inequality, yields a lower bound on $\mathrm{KL}(\mathbb{P}_{\boldsymbol{\Delta},A}, \mathbb{P}_{\boldsymbol{\Delta}^{(k)},A})$. *(iii) KL decomposition.* We then decompose this KL along unordered pairs. The two instances $\boldsymbol{\Delta}$ and $\boldsymbol{\Delta}^{(k)}$ differ only on duels involving arm k, and for each such pair the Bernoulli parameters differ by at most a constant of order $|\Delta_{k,(1)}|$. A Bernoulli KL upper bound, combined with the decomposition, forces $\mathbb{E}_{\boldsymbol{\Delta},A}[N_k] \gtrsim \Delta^{-2}_{k,(1)} \log(1/\delta)$, where N_k counts duels involving k. Summing this constraint over all $k \neq 1$ yields the desired lower bound (88).

Proof. Step 1: reference and perturbed instances.

The argument is fully instance-dependent: we fix an arbitrary gap matrix $\mathbf{\Delta} \in \mathbb{D}_{\mathrm{cw}}$ with Condorcet winner $i^* = 1$, and work throughout with this specific instance. We denote $\mathbb{P}$ the probability induced by the interaction between $\mathbf{\Delta}$ and A.

Let $\epsilon > 0$ be a constant, arbitrary small. Let $k \neq 1$ be an arm that is not the CW under $\mathbf{\Delta}$. A simple way to modify $\mathbf{\Delta}$ so that k becomes the CW, is to make all non-positive entries in the k-th row of $\mathbf{\Delta}$ equal to ϵ.

Construct the gap matrix $\mathbf{\Delta}^{(k)}$ as follows. For all $i, j \notin \{1, k\}$, set $\Delta^{(k)}_{i,j} = \Delta_{i,j}$. Set $\Delta^{(k)}_{k,1} = \epsilon$ and $\Delta^{(k)}_{1,k} = -\epsilon$. Finally, for each $j \notin \{1, k\}$, define

$$\Delta^{(k)}_{k,j} = \begin{cases} \Delta_{k,j}, & \text{if } \Delta_{k,j} > 0, \\ \epsilon, & \text{if } \Delta_{k,j} \leqslant 0, \end{cases} \qquad \Delta^{(k)}_{j,k} = -\Delta^{(k)}_{k,j}. \tag{89}$$

For ϵ small enough, the modified matrix $\mathbf{\Delta}^{(k)}$ can be represented as

$$\Delta^{(k)} = \begin{pmatrix} 0 & \Delta_{1,2} & \cdots & \Delta_{1,k-1} & -\epsilon & \Delta_{1,k+1} & \cdots & \Delta_{1,K} \\ \Delta_{2,1} & 0 & \cdots & \Delta_{2,k-1} & -(\epsilon \vee \Delta_{k,2}) & \Delta_{2,k+1} & \cdots & \Delta_{2,K} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ \Delta_{k-1,1} & \Delta_{k-1,2} & \cdots & 0 & -(\epsilon \vee \Delta_{k,k-1}) & \Delta_{k-1,k+1} & \cdots & \Delta_{k-1,K} \\ \epsilon & \epsilon \vee \Delta_{k,2} & \cdots & \epsilon \vee \Delta_{k,k-1} & 0 & \epsilon \vee \Delta_{k,k+1} & \cdots & \epsilon \vee \Delta_{k,K} \\ \Delta_{k+1,1} & \Delta_{k+1,2} & \cdots & \Delta_{k+1,k-1} & -(\epsilon \vee \Delta_{k,k+1}) & 0 & \cdots & \Delta_{k+1,K} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\ \Delta_{K,1} & \Delta_{K,2} & \cdots & \Delta_{K,k-1} & -(\epsilon \vee \Delta_{k,K}) & \Delta_{K,k+1} & \cdots & 0 \end{pmatrix}.$$

where the blue entries indicate the differences with respect to the reference $\mathbf{\Delta}$. In fact, only non-positive entries of row k are modified between $\mathbf{\Delta}$ and $\mathbf{\Delta}^{(k)}$.

Since $\epsilon > 0$, the k-th row $\Delta^{(k)}_{k,\cdot}$ is positive (aside from $\Delta^{(k)}_{k,k}$). Hence, the CW of $\mathbf{\Delta}^{(k)}$ is k. As standard in this line of work, our construction is motivated by the fact that the instance $\mathbf{\Delta}^{(k)}$ is hard to distinguish from the reference gap matrix $\mathbf{\Delta}$.

For $k \geq 2$, denote by $\mathbb{P}^{(k)}$ the distribution of the data when the underlying gap matrix is $\mathbf{\Delta}^{(k)}$. Formally, when the algorithm queries a pair (i, j) with $i < j$, it receives a sample $X_{i,j} \sim \mathcal{B}\big(\Delta^{(k)}_{i,j} + \frac{1}{2}\big)$, where $\mathcal{B}(p)$ denotes the Bernoulli distribution with parameter p.

Step 2: information-theoretic arguments.

Let $\hat{i}$ denote the output of algorithm A, which is assumed to be δ-correct over $\mathbb{D}_{\mathrm{cw}}$. When the gap matrix is $\mathbf{\Delta}^{(k)}$ the CW is k, so δ-correctness implies

$$\forall k \neq i^*, \ \mathbb{P}^{(k)}(\hat{i} = k) \geq 1 - \delta \quad \text{and} \quad \mathbb{P}(\hat{i} = k) \leqslant \delta \, .$$

By the definition of the total variation distance,

$$\mathrm{TV}\big(\mathbb{P}, \mathbb{P}^{(k)}\big) \geq \big|\mathbb{P}(\hat{i} = k) - \mathbb{P}^{(k)}(\hat{i} = k)\big| \geq 1 - 2\delta \, . \tag{90}$$

From (90), the Bretagnolle–Huber inequality (Theorem 14.2 in Lattimore and Szepesvári, 2020) yields

$$1 - 2\delta \leqslant \mathrm{TV}\big(\mathbb{P}, \mathbb{P}^{(k)}\big) \leqslant 1 - \frac{1}{2} \exp\Big\{-\mathrm{KL}\big(\mathbb{P}, \mathbb{P}^{(k)}\big)\Big\} \, . \tag{91}$$

Step 3: computing the KL divergence and concluding on the budget.

For any unordered pair $\{i, j\}$ with $i \neq j$, denote by $N_{\{i,j\}}$ the number of duels involving either (i, j) or (j, i), that is,

$$N_{\{i,j\}} \coloneqq N_{i,j} + N_{j,i},$$

where

$$N_{i,j} \coloneqq |\{t \in [N_\delta] : (I_t, J_t) = (i, j)\}|,$$

and N_δ is the stopping time of the algorithm.

Using the divergence decomposition lemma (Lemma 15.1 in Lattimore and Szepesvári, 2020) and the fact that the two instances $\mathbf{\Delta}$ and $\mathbf{\Delta}^{(k)}$ differ only on pairs involving arm k, we obtain

$$\begin{aligned}\mathrm{KL}\big(\mathbb{P}, \mathbb{P}^{(k)}\big) &= \sum_{1 \leqslant i < j \leqslant K} \mathbb{E}\big[N_{\{i,j\}}\big]\ \mathrm{KL}\big(\mathbb{P}_{i,j}, \mathbb{P}^{(k)}_{i,j}\big) \\ &= \sum_{\substack{i=1 \\ i \neq k}}^{K} \mathbb{E}\big[N_{\{k,i\}}\big]\ \mathrm{KL}\big(\mathbb{P}_{k,i}, \mathbb{P}^{(k)}_{k,i}\big). \end{aligned} \tag{92}$$

We now upper bound $\mathrm{KL}\big(\mathbb{P}_{k,i}, \mathbb{P}^{(k)}_{k,i}\big)$. Let $i \in [K]$ with $i \neq k$. If $\Delta_{k,i} > 0$, then by construction $\mathbb{P}_{k,i} = \mathbb{P}^{(k)}_{k,i}$. Otherwise, if $\Delta_{k,i} < 0$, the corresponding Bernoulli feedback distributions satisfy

$$\mathbb{P}_{k,i} = \mathcal{B}\big(\tfrac{1}{2} + \Delta_{k,i}\big)\,, \qquad \mathbb{P}^{(k)}_{k,i} = \mathcal{B}\big(\tfrac{1}{2} + \epsilon\big)\,,$$

so that

$$\begin{aligned}\mathrm{KL}\big(\mathbb{P}_{k,i}, \mathbb{P}^{(k)}_{k,i}\big) &= \mathrm{kl}\big(\Delta_{k,i} + \tfrac{1}{2},\ \epsilon + \tfrac{1}{2}\big) \\ &\leqslant \frac{\big(\Delta_{k,i} + \frac{1}{2} - (\epsilon + \frac{1}{2})\big)^2}{\big(-\epsilon + \frac{1}{2}\big)\big(\epsilon + \frac{1}{2}\big)} \qquad (93) \\ &\leqslant \frac{(\epsilon - \Delta_{k,(1)})^2}{\big(-\epsilon + \frac{1}{2}\big)\big(\epsilon + \frac{1}{2}\big)}\ . \qquad (94)\end{aligned}$$

Here, (93) follows from the standard upper bound $\mathrm{kl}(p, q) \leqslant (p - q)^2/[q(1 - q)]$ for $p, q \in (0, 1)$, while (94) uses that $\Delta^2_{k,i} \leqslant \Delta^2_{k,(1)}$ by definition of $\Delta_{k,(1)}$ as the smallest negative entry in row k. Similarly, if $\Delta_{k,j} = 0$,

$$\mathbb{P}_{k,j} = \mathcal{B}\big(\tfrac{1}{2}\big)\,, \qquad \mathbb{P}^{(k)}_{k,j} = \mathcal{B}\big(\tfrac{1}{2} + \epsilon\big)\,,$$

so that the same computation gives

$$\mathrm{KL}\big(\mathbb{P}_{k,j}, \mathbb{P}^{(k)}_{k,j}\big) \leqslant \frac{\epsilon^2}{\big(-\epsilon + \frac{1}{2}\big)\big(\epsilon + \frac{1}{2}\big)}\ ,$$

which vanishes to 0 with $\epsilon \to 0$.

Combining these bounds with (92), and taking the limit $\epsilon \to 0$, we obtain

$$\mathrm{KL}\big(\mathbb{P}, \mathbb{P}^{(k)}\big) \leqslant 4 \left(\sum_{\substack{i=1 \\ i \neq k}}^{K} \mathbb{1}_{\{\Delta_{k,i} < 0\}}\, \mathbb{E}\big[N_{\{k,i\}}\big] \right) \Delta^2_{k,(1)}.$$

Using the Bretagnolle–Huber inequality (91) from Step 3 , we then get

$$\sum_{\substack{i=1 \\ i \neq k}}^{K} \mathbb{1}_{\{\Delta_{k,i} < 0\}}\, \mathbb{E}\big[N_{\{k,i\}}\big] \geq \frac{1}{4\,\Delta^2_{k,(1)}} \log \frac{1}{4\delta}\ .$$

Summing over $k \geq 2$, we conclude that the total number of queries N_δ satisfies

$$\begin{aligned}\mathbb{E}\left[N_\delta\right] &\geq \sum_{k \neq 1}^{K} \left(\sum_{\substack{i=1 \\ i \neq k}}^{K} \mathbb{1}_{\{\Delta_{k,i} < 0\}}\, \mathbb{E}\big[N_{\{k,i\}}\big] \right) \\ &\geq \frac{1}{4} \sum_{k=2}^{K} \frac{1}{\Delta^2_{k,(1)}} \log \frac{1}{4\delta}\ ,\end{aligned}$$

which is exactly the claimed lower bound (88). □

F.2.2 Bound in quantile in Proposition 3.1

In this paragraph, we prove the quantile bound from Proposition 3.1, namely

$$\mathbb{P}_{\boldsymbol{\Delta},A}\left(N_\delta \geqslant \frac{1}{3}\sum_{i\neq i^*}\frac{\left(\frac{1}{6\delta}\right)}{\Delta_{i,(1)}^2}\right) \geqslant \delta\,. \tag{95}$$

Sketch of proof. Even though expectation lower bounds are naturally weaker than quantile bounds, deducing a quantile lower bound from its expectation counterpart is nontrivial. Still, the arguments largely mirror the expectation proof (paragraph F.2.1), we follow the same three-step roadmap. *(i) Reference instance and alternative instances*: we fix $\boldsymbol{\Delta} \in \mathbb{D}_{\mathrm{cw}}$ with CW i^* and, for each $k \neq i^*$, construct $\boldsymbol{\Delta}^{(k)}$ by zeroing the negative entries in row/column k. In particular, $\boldsymbol{\Delta}^{(k)}$ has two nonnegative rows (those of 1 and k) and therefore does not belong to $\mathbb{D}_{\mathrm{cw}}$. *(ii) Separating event and total variation*: we now use the stopping rule instead of the recommendation and consider the event $B = \{N_\delta > Q\}$, where Q is the $(1-\delta)$-quantile of N_δ under $\boldsymbol{\Delta}$. This event has small probability under $\mathbb{P}$, but (after a continuity argument using perturbations of $\boldsymbol{\Delta}^{(k)}$) it has large probability under $\mathbb{P}^{(k)}$, since A cannot quickly decide between $\hat{i}=1$ and $\hat{i}=k$. *(iii) KL decomposition*: as before, we decompose the KL along pairs and use that $\boldsymbol{\Delta}$ and $\boldsymbol{\Delta}^{(k)}$ differ only on duels involving k. The only extra ingredient is that, since B depends only on the first Q observations, we introduce a truncated algorithm $\tilde{A}$ that stops at time Q, apply the KL decomposition to $\tilde{A}$, and then reinterpret the resulting inequality as a lower bound on Q, i.e., on the $(1-\delta)$-quantile of N_δ.

Proof. Let A be any δ-correct algorithm on $\mathbb{D}_{\mathrm{cw}}$.

Step 1: Reference and alternative instances.

As for the expectation bound, we fix any $\boldsymbol{\Delta}$ as the reference instance, with $i^* = 1$. Denote $\mathbb{P}_A$ for the probability induced by the interaction between A and $\boldsymbol{\Delta}$. Then, for each suboptimal arm $k \in \{2,\dots,K\}$, construct $\boldsymbol{\Delta}^{(k)}$ by setting to zero all entries (k,j) with $\Delta_{k,j} < 0$, that is, $\boldsymbol{\Delta}^{(k)}$ is defined by Equation (89) with $\epsilon = 0$.

The matrices $\boldsymbol{\Delta}$ and $\boldsymbol{\Delta}^{(k)}$ differ only in row/column k. By construction, rows 1 and k of $\boldsymbol{\Delta}^{(k)}$ contain only nonnegative entries, so $\boldsymbol{\Delta}^{(k)}$ contains two weak CW, and 1 and k are tied. In particular, $\boldsymbol{\Delta}^{(k)} \notin \mathbb{D}_{\mathrm{cw}}$. Denote $\mathbb{P}_A^{(k)}$ the distribution induced by A interacting with $\boldsymbol{\Delta}^{(k)}$.

Step 2: bound in total variation and reduction to fixed budget.

Consider the recommendation rule $\hat{i}$ and the budget N_δ of A. Define Q as the $(1-\delta)$-quantile of the budget under $\mathbb{P}_A$:

$$Q = \inf\{x > 0 \text{ s.t. } \mathbb{P}_A(N_\delta \geqslant x) \leqslant \delta\}. \tag{96}$$

Consider the event $B := \{N_\delta \leqslant Q\}$ where the budget is at most Q.

We define a truncated version $\tilde{A}$ of A with budget at most Q as follows: run A until it stops or Q queries have been made; if A stops by time Q, return $\tilde{i} = \hat{i}$; otherwise, return $\tilde{i} = 0$. Let $\tilde{N}_\delta$, $\tilde{i}$ be $\tilde{A}$'s budget/recommendation, $\mathbb{P}_{\tilde{A}}$ (resp. $\mathbb{P}_{\tilde{A}}^{(k)}$) its law under $\boldsymbol{\Delta}$ (resp. $\boldsymbol{\Delta}^{(k)}$). By construction,

$$B = \{\tilde{i} \neq 0\}\,,$$

so this event is measurable with respect to the observations of algorithm $\tilde{A}$. Moreover, it has the same probability under A and $\tilde{A}$.

We now lower bound the total variation distance between $\mathbb{P}_{\tilde{A}}$ and $\mathbb{P}_{\tilde{A}}^{(k)}$.

By the definition of Q in (96), we have

$$\mathbb{P}_{\tilde{A}}(\tilde{i}=0) = \mathbb{P}_A(N_\delta > Q) \;\leqslant\; \delta. \tag{97}$$

Under $\boldsymbol{\Delta}^{(k)}$, the instance does not belong to $\mathbb{D}_{\mathrm{cw}}$, so we cannot directly invoke δ-correctness. We therefore approximate $\boldsymbol{\Delta}^{(k)}$ by nearby instances in $\mathbb{D}_{\mathrm{cw}}$. More precisely, define $\boldsymbol{\Delta}^{(k,\epsilon)}$ as (89), i.e.,

by lifting all off-diagonal zero entries in row k of $\mathbf{\Delta}^{(k)}$ to $\epsilon > 0$ (and adjusting the k-th column to preserve antisymmetry). For $0 < \epsilon \leqslant 1/4$, the matrix $\mathbf{\Delta}^{(k,\epsilon)}$ lies in $\mathbb{D}_{\mathrm{cw}}$ and admits k as its CW. Similarly, define $\mathbf{\Delta}^{(k,-\epsilon)}$ by subtracting ϵ from all off-diagonal zero entries in the k-th row of $\mathbf{\Delta}^{(k)}$ (and adjusting the k-th column to preserve antisymmetry), so that $\mathbf{\Delta}^{(k,-\epsilon)} \in \mathbb{D}_{\mathrm{cw}}$ and admits 1 as its CW.

Let $\mathbb{P}_A^{(k,\epsilon)}$ and $\mathbb{P}_A^{(k,-\epsilon)}$ denote the laws of A under $\mathbf{\Delta}^{(k,\epsilon)}$ and $\mathbf{\Delta}^{(k,-\epsilon)}$, respectively. Since A is δ-correct on $\mathbb{D}_{\mathrm{cw}}$, we have

$$\mathbb{P}_A^{(k,\epsilon)}(\hat{i} \neq k) \leqslant \delta, \qquad \mathbb{P}_A^{(k,-\epsilon)}(\hat{i} \neq 1) \leqslant \delta\,. \tag{98}$$

Moreover, for either sign, the law of the transcript of $\tilde{A}$ under $\mathbf{\Delta}^{(k,\pm\epsilon)}$ converges in total variation to $\mathbb{P}_{\tilde{A}}^{(k)}$ as $\epsilon \to 0$, since $\tilde{A}$ uses at most Q observations. In particular, this convergence applies to the events $\{\hat{i} = 1, N_\delta \leqslant Q\}$ and $\{\hat{i} \neq 1, N_\delta \leqslant Q\}$, which are measurable with respect to this truncated transcript. Using these facts and letting $\epsilon \to 0$, we obtain

$$\begin{aligned}
\mathbb{P}_A^{(k)}(N_\delta \leqslant Q) &= \mathbb{P}_A^{(k)}(\hat{i} = 1, N_\delta \leqslant Q) + \mathbb{P}_A^{(k)}(\hat{i} \neq 1, N_\delta \leqslant Q) \\
&= \lim_{\epsilon \to 0} \Big[\mathbb{P}_A^{(k,\epsilon)}(\hat{i} = 1, N_\delta \leqslant Q) + \mathbb{P}_A^{(k,-\epsilon)}(\hat{i} \neq 1, N_\delta \leqslant Q)\Big] \\
&\leqslant 2\delta\ ,
\end{aligned} \tag{99}$$

where the last inequality follows from (98).

Since $\tilde{A}$ and A coincide up to time Q, we also have $\mathbb{P}_{\tilde{A}}^{(k)}(\tilde{i} \neq 0) = \mathbb{P}_A^{(k)}(N_\delta \leqslant Q)$, so (99) implies

$$\mathbb{P}_{\tilde{A}}^{(k)}(\tilde{i} \neq 0) \leqslant 2\delta.$$

Combining (97) with this inequality, and using the event $\{\tilde{i} = 0\}$, we obtain

$$\mathrm{TV}\big(\mathbb{P}_{\tilde{A}}, \mathbb{P}_{\tilde{A}}^{(k)}\big) \geqslant \mathbb{P}_{\tilde{A}}^{(k)}(\tilde{i} = 0) - \mathbb{P}_{\tilde{A}}(\tilde{i} = 0) \geqslant 1 - 3\delta\ . \tag{100}$$

Remarks. *Intuitively, one may think of $\tilde{A}$ as a fixed-budget algorithm with budget Q. This can be viewed as a reduction: from any δ-correct algorithm A that enjoys a high-probability control on its budget (namely, $\mathbb{P}_A(N_\delta \leqslant Q) \geqslant 1 - \delta$), we construct a fixed-budget algorithm $\tilde{A}$ with budget Q that inherits the same distinguishing power between the reference instance and its perturbations.*

Step 3: computing the KL divergence.

By the Bretagnolle–Huber inequality (see, e.g., Lattimore and Szepesvári, 2020), we have

$$1 - 3\delta \leqslant \mathrm{TV}\big(\mathbb{P}_{\tilde{A}}, \mathbb{P}_{\tilde{A}}^{(k)}\big) \leqslant 1 - \frac{1}{2}\exp\Big\{-\mathrm{KL}\big(\mathbb{P}_{\tilde{A}}, \mathbb{P}_{\tilde{A}}^{(k)}\big)\Big\}\ .$$

In particular,

$$\mathrm{KL}\big(\mathbb{P}_{\tilde{A}}, \mathbb{P}_{\tilde{A}}^{(k)}\big)\ \geqslant\ \log\left(\frac{1}{6\delta}\right). \tag{101}$$

We now decompose $\mathrm{KL}\big(\mathbb{P}_{\tilde{A}}, \mathbb{P}_{\tilde{A}}^{(k)}\big)$. For $i \neq j$ in $[K]$, let $N_{i,j}$ denote the number of duels (i, j) made by $\tilde{A}$, while $N_{\{i,j\}} = N_{i,j} + N_{j,i}$ is the number of duels with unordered pair $\{i, j\}$. By the standard KL decomposition for adaptive procedures [Lattimore and Szepesvári, 2020],

$$\begin{aligned}
\mathrm{KL}\big(\mathbb{P}_{\tilde{A}}, \mathbb{P}_{\tilde{A}}^{(k)}\big) &= \sum_{i \neq j} \mathbb{E}_{\tilde{A}}\big[N_{i,j}\big]\, \mathrm{KL}\big(\mathbb{P}_{i,j}, \mathbb{P}_{i,j}^{(k)}\big) \\
&= \sum_{j:\Delta_{k,j}<0} \mathbb{E}_{\tilde{A}}\big[N_{\{k,j\}}\big]\, \mathrm{KL}\big(\mathbb{P}_{k,j}, \mathbb{P}_{k,j}^{(k)}\big)\ ,
\end{aligned} \tag{102}$$

since $\mathbf{\Delta}$ and $\mathbf{\Delta}^{(k)}$ differ only on duels $\{k, j\}$ with $\Delta_{k,j} < 0$.

For j with $\Delta_{k,j} < 0$, $\mathbb{P}_{k,j}^{(k)} = \mathcal{B}(1/2)$, $\mathbb{P}_{k,j} = \mathcal{B}(1/2 + \Delta_{k,j})$ with $\Delta_{k,j} \in [-1/4, 0]$. Thus,

$$\begin{aligned}\mathrm{KL}\big(\mathbb{P}_{k,j}, \mathbb{P}_{k,j}^{(k)}\big) &\leqslant \frac{1}{2}\log\left(\frac{\frac{1}{2}}{\frac{1}{2}+\Delta_{k,j}}\right) + \frac{1}{2}\log\left(\frac{\frac{1}{2}}{\frac{1}{2}-\Delta_{k,j}}\right) \\ &= -\frac{1}{2}\log\big(1-4\Delta_{k,j}^2\big) \;\leqslant\; 8\log\left(\frac{4}{3}\right)\Delta_{k,j}^2 \,, \end{aligned} \tag{103}$$

where (103) follows from $\sup_{x\in[0,1/4]} \frac{-\log(1-x)}{x} \leqslant 4\log\big(\frac{4}{3}\big)$, applied with $x = 4\Delta_{k,j}^2 \in [0, 1/4]$. Plugging these bounds into (102), we obtain

$$\begin{aligned}\mathrm{KL}\big(\mathbb{P}_{\tilde{A}}, \mathbb{P}_{\tilde{A}}^{(k)}\big) &\leqslant 8\log\left(\frac{4}{3}\right) \sum_{j:\Delta_{k,j}<0} \mathbb{E}_{\tilde{A}}[N_{\{k,j\}}]\,\Delta_{k,j}^2 \\ &\leqslant 8\log\left(\frac{4}{3}\right)\left(\sum_{j:\Delta_{k,j}<0} \mathbb{E}_{\tilde{A}}[N_{\{k,j\}}]\right)\Delta_{k,(1)}^2 \,, \end{aligned} \tag{104}$$

since $\Delta_{k,j}^2 \leqslant \Delta_{k,(1)}^2$ whenever $\Delta_{k,j} < 0$. Combining (101)–(104), we obtain

$$\frac{1}{8\log(4/3)}\frac{1}{\Delta_{k,(1)}^2}\log\left(\frac{1}{6\delta}\right) \leqslant \sum_{j:\Delta_{k,j}<0} \mathbb{E}_{\tilde{A}}[N_{\{k,j\}}] \,. \tag{105}$$

Since $\tilde{A}$ uses at most Q duels and each duel is counted at most once, we have almost surely, under $\mathbb{P}_{\tilde{A}}$,

$$\sum_{k\neq i^*}\sum_{j:\Delta_{k,j}<0} N_{\{k,j\}} \;\leqslant\; \tilde{N}_\delta \leqslant Q \,,$$

in particular, the same bound also holds in expectation $\mathbb{E}_{\tilde{A}}$.

Summing in (105) over $k \neq i^*$ yields

$$\frac{1}{8\log(4/3)}\sum_{k\neq i^*}\frac{1}{\Delta_{k,(1)}^2}\log\left(\frac{1}{6\delta}\right) \leqslant \sum_{k\neq i^*}\sum_{j:\Delta_{k,j}<0} \mathbb{E}_{\tilde{A}}[N_{\{k,j\}}] \leqslant Q \,.$$

Using the numerical bound $8\log(4/3) < 3$ and the definition of Q as the $(1-\delta)$-quantile of N_δ then gives the high-probability lower bound (95).

□

F.3 Proof of Theorem 3.2

In this subsection, we prove the high-probability lower bound from Theorem 3.2. We start with a sketch of proof.

Proof Sketch of (9)**.** The key idea is to reduce CW identification to multiple active signal detection problems [Castro, 2014]: for each $i \neq i^*$, we need to certify that the row $\Delta_{k,\cdot}$ has at least a negative entry, this with a probability of error at most δ. Along the way, we have to improve state-of-the-art lower bounds for such detection problems.

Consider any δ-correct algorithm A. We introduce a collection $\boldsymbol{\Delta}^{(\pi)}$ of gap matrices that differ from $\boldsymbol{\Delta}$ as we permute, on each row $i \neq i^*$, the position of the negative entries by some collection $\pi = (\pi_i)_{i\neq i^*}$ of permutations while preserving the skew symmetry. Then, we fix a specific arm $i \neq i^*$ and construct the gap matrices $\boldsymbol{\Delta}^{(\pi,i)}$, by setting to 0 the negative entries of the i-th row of $\boldsymbol{\Delta}^{(\pi)}$ while preserving skew symmetry. As $\boldsymbol{\Delta}^{(\pi,i)}$ contains two rows with non-negative entries, one easily deduces that the budget of A under $\boldsymbol{\Delta}^{(\pi,i)}$ is arbitrarily large with probability $1-\delta$. In contrast, under $\boldsymbol{\Delta}$, A finishes before χ—the $(1-\delta)$-quantile of N_δ under $\mathbb{P}_{\boldsymbol{\Delta},A}$. Hence, we reduce A to an active testing problem with budget χ for any unknown gap matrix $\tilde{\boldsymbol{\Delta}}$ of the hypotheses

$$H_0^{(i)} : \tilde{\boldsymbol{\Delta}} = \boldsymbol{\Delta}^{(\pi,i)} \text{ for some perm. } \pi \quad \text{vs.} \quad H_1^{(i)} : \tilde{\boldsymbol{\Delta}} = \boldsymbol{\Delta}^{(\pi)} \text{ for some perm. } \pi \;\; .$$

Since the permutation π is unknown to the learner, this allows to improve over (7) by accounting for the fact that the algorithm must explore all possible positions of the negative entries in row i.

Then, by a convexity argument, we deduce $\chi \gtrsim \frac{K_{i;<0}}{\|\Delta_i^-\|^2}\log(1/\delta)$. Optimizing over $i \neq i^*$ yields the first part of the bound $\chi \gtrsim \max_{i\neq i^*} \frac{K_{i;<0}}{\|\Delta_i^-\|^2}\log(1/\delta)$.

The second term $\sum_{i\neq i^*} \frac{K_{i;<0}}{\|\Delta_i^-\|^2}$ interprets as the total cost for testing all hypotheses $H_0^{(i)}$ against $H_1^{(i)}$ with a constant error probability. We develop new arguments for this multiple-hypotheses problem. Unlike the involved technique in Simchowitz et al. [2017], we reduce w.l.o.g. to the case where all tests have similar complexity $\frac{\|\Delta_i^-\|^2}{K_{i;<0}}$ and we write β^2 for this common value. Then, we build upon the symmetry of our problem to reduce to the case where each row receives the same sampling effort χ/K. Relying again on lower bounds on active signal detection, we get $\chi/K \gtrsim \beta^{-2}$, which leads to the desired result. We believe that these arguments can generalize to other multiple active testing problems. The full proof is given in Appendix F.3.

We first introduce additional notation and state a more precise version of the result, in particular, we consider here the case where $\mathbf{\Delta}$ might contains some ties.

Notation and precise formulation Consider a matrix $\mathbf{\Delta}$ with entries in $\left[-\frac{1}{4}, \frac{1}{4}\right]$ that admits a unique CW $i^*(\mathbf{\Delta})$, that is, $\mathbf{\Delta} \in \mathbb{D}_{\mathrm{cw}}$. Without loss of generality, we assume that $i^*(\mathbf{\Delta}) = 1$, and we keep this matrix fixed throughout this section. We now define a family of environments obtained from $\mathbf{\Delta}$ by permuting its entries in a specific way. This will lead to a more precise formulation of Theorem 3.2.

Fix an antisymmetric matrix $\mathbf{\Sigma} = (\sigma_{i,j})_{1\leqslant i,j\leqslant K}$ defined, for any $1 \leqslant i \neq j \leqslant K$, by

$$\sigma_{i,j} = \begin{cases} \mathrm{sign}(\Delta_{i,j}) \,, & \text{if } \Delta_{i,j} \neq 0 \,, \\ f_{\mathrm{tb}}(i,j) \,, & \text{if } \Delta_{i,j} = 0 \,, \end{cases} \tag{106}$$

where $f_{\mathrm{tb}} : [K]^2 \to \{-1, 0, 1\}$ is an antisymmetric function used as a tie-breaking convention. In other words, $\mathbf{\Sigma}$ records the sign pattern of the gap matrix $\mathbf{\Delta}$, with a fixed rule when $\Delta_{i,j} = 0$. Note that $\mathbf{\Sigma}$ is antisymmetric, since $\mathbf{\Delta}$ is antisymmetric and f_{tb} is antisymmetric by assumption. For now, we do not specify how we fix this convention, as the precise choice will arise naturally at the very end of our proofs.

For a suboptimal arm $i \in \{2, \dots, K\}$, denote by Σ_i^- the set of arms that beat i (with ties broken according to f_{tb}),

$$\Sigma_i^- \coloneqq \left\{ j \in [K] \setminus \{i\} : \sigma_{i,j} = -1 \right\} \,.$$

For any $i \in \{2, \dots, K\}$, define $\Pi_i(\mathbf{\Sigma})$ as the set of permutations of Σ_i^-:

$$\Pi_i(\mathbf{\Sigma}) \coloneqq \left\{ \pi_i : \Sigma_i^- \to \Sigma_i^- \,\middle|\, \pi_i \text{ is a bijection} \right\} \,,$$

and set

$$\Pi(\mathbf{\Sigma}) \coloneqq \Pi_2(\mathbf{\Sigma}) \times \cdots \times \Pi_K(\mathbf{\Sigma}) \,. \tag{107}$$

Fix a permutation $\pi \coloneqq (\pi_i)_{i=2}^K \in \Pi(\mathbf{\Sigma})$. For each row $i \neq 1$, we permute the entries of $\mathbf{\Delta}$ indexed by Σ_i^- according to π_i. To preserve antisymmetry, we apply the same permutation π_i to the corresponding entries in column i. We thus define the matrix $\mathbf{\Delta}^{(\pi)}$ as follows: for all $(i,j) \in [K]^2$,

$$\Delta_{i,j}^{(\pi)} \coloneqq \begin{cases} \Delta_{i,\pi_i(j)}, & \text{if } \sigma_{i,j} = -1 \,, \\ -\Delta_{j,\pi_j(i)}, & \text{if } \sigma_{i,j} = 1 \,, \\ \Delta_{i,j}, & \text{else} \,, \end{cases} \tag{108}$$

the last condition might happen when $\sigma_{i,j} = 0$, which only happens for ties.

The permutations in $\Pi(\mathbf{\Sigma})$ destroy any exploitable ordering structure between arms while preserving, in each row, the multiset of non-positive entries and hence the row-wise hardness parameters $(K_{i;<0}, \|\Delta_i^-\|_2^2)$.

We now list properties that are preserved by the permutation π. In the following lemma, a row j is said to be a *weak CW* for a given gap matrix if all entries in its row are nonnegative.

Lemma F.1. *For any $\mathbf{\Sigma}$ satisfying* (106) *and any $\pi \in \Pi(\mathbf{\Sigma})$, we have*

$$1.\ \mathbf{\Delta}^{(\pi)} = -(\mathbf{\Delta}^{(\pi)})^T \quad \textit{(antisymmetry)}$$

$$2.\ \forall j \neq 1, \Delta^{(\pi)}_{1,j} \geqslant 0 \quad \textit{(row 1 is a weak CW)}$$

$$3.\ \forall i \neq 1,\ K_{i,<0}(\mathbf{\Delta}) = K_{i,<0}(\mathbf{\Delta}^{(\pi)})$$

$$4.\ \forall i \neq 1, \forall s \leqslant K_{i,<0},\ \Delta^{(\pi)}_{i,(s)} = \Delta_{i,(s)} \quad \textit{(same ordered nonpositive entries)}$$

In particular, if $\mathbf{\Delta}$ has no ties (that is, $\forall i \neq j,\ \Delta_{i,j} \neq 0$), then $\mathbf{\Delta}^{(\pi)} \in \mathbb{D}_{\mathrm{cw}}$. Moreover, for any $\boldsymbol{s}$, $H_{\textit{explore}}(\boldsymbol{s}, \delta)$ and $H_{\textit{certify}}(\boldsymbol{s}, \delta)$ remains unchained under any permutation $\pi \in \Pi(\mathbf{\Sigma})$.

Proof of Lemma F.1. 1. Anti-symmetry Let $(i, j) \in [K]^2$ with $i \neq j$.
Assume that $(\sigma_{i,j} = -1)$ which is equivalent to $j \in \Sigma_i^-$. By the first line in Equation (108), one has $\Delta^{(\pi)}_{i,j} = \Delta^{(\pi)}_{i,\pi_i(j)}$. Then, by the second line applied with (j, i) $(\sigma_{j,i} = -1)$, one has $\Delta^{(\pi)}_{j,i} = -\Delta_{i,\pi_i(j)} = -\Delta^{(\pi)}_{i,j}$.
The case $\sigma_{i,j} = 1$ is treated similarly.
For $(\sigma_{i,j} = 0)$, one also have $\sigma_{j,i} = 0$ and $\Delta^{\pi}_{i,j} = \Delta_{i,j} = -\Delta_{j,i} = -\Delta^{\pi}_{j,i}$. It proves the anti-symmetry of $\mathbf{\Delta}^{(\pi)}$.

2. CW row By definition, $i^* \in \Sigma_i^-$ for any $i \neq i^*$ (and $i^* = 1$ by assumption). In particular, $\Delta^{(\pi)}_{1,i} = \Delta_{\pi_i(1),i} \geqslant 0$, the non-negativity comes from the fact that $\pi_i(1) \in \Sigma_i^-$, that is $\pi_i(1)$ also beats i (or is even with i in the case where Δ contains ties).

3. Negative entries. For every non-CW arm $i \neq 1$, the set of indices Σ_i^- contains all indices j such that $\Delta_{i,j} < 0$, and only indices j such that $\Delta_{i,j} \leqslant 0$. By (108), the entries indexed by Σ_i^- are permuted, while all entries outside this set remain nonnegative. Hence the multiset of strictly negative entries in row i is preserved, and so is the count $K_{i,<0}$.

4. Order statistics. Since the multiset of strictly negative entries in row i is preserved up to permutation, the ordered sequence $(\Delta_{i,(s)})_{1 \leqslant s \leqslant K_{i,<0}}$ is unchanged, which gives $\Delta^{(\pi)}_{i,(s)} = \Delta_{i,(s)}$ for all $1 \leqslant s \leqslant K_{i,<0}$. In particular, for any vector $\boldsymbol{s}$ with $1 \leqslant s_i \leqslant K_{i;<0}$ (a condition independent of π), the gaps $\Delta^{(\pi)}_{i,(s_i)}$ coincide with $\Delta_{i,(s_i)}$, so that $H_{\mathrm{explore}}(\boldsymbol{s}, \delta)$ and $H_{\mathrm{certify}}(\boldsymbol{s}, \delta)$ remain unchanged under any $\pi \in \Pi(\mathbf{\Sigma})$. □

Now we are ready to state Theorem F.2, which directly implies Theorem 3.2 and provides a more constructive formulation.

Theorem F.2. *Let A be a δ-correct algorithm over the class $\mathbb{D}_{\mathrm{cw}}$, with $\delta \leqslant 1/12$. Let $\mathbf{\Delta} \in \mathbb{D}_{\mathrm{cw}}$ be such that $i^*(\mathbf{\Delta}) = 1$.*

Define χ as the smallest positive number such that, for any sign convention $\mathbf{\Sigma}$ (satisfying (106)*) and any $\pi \in \Pi(\mathbf{\Sigma})$, one has $\mathbb{P}_{\mathbf{\Delta}^{(\pi)},A}(N_\delta > \chi) \leqslant \delta$. Equivalently,*

$$\chi \coloneqq \inf \left\{ x > 0 : \sup_{\mathbf{\Sigma}} \sup_{\pi \in \Pi(\mathbf{\Sigma})} \mathbb{P}_{\mathbf{\Delta}^{(\pi)},A}(N_\delta > x) \leqslant \delta \right\} \ . \tag{109}$$

In other words, χ is a uniform $(1-\delta)$-quantile of the budget, taken in the worst case over all admissible sign conventions and permutations.

Then, χ satisfies

$$\chi \geqslant \frac{1}{8\log(4/3)} \max_{i=2}^{K} \frac{K_{i;\leqslant 0}}{\|\Delta_i^-\|^2} \log\left(\frac{1}{6\delta}\right), \tag{110}$$

$$\chi \geqslant \frac{1}{128\log(4/3)} \cdot \frac{1}{\log(2K)} \sum_{i=2}^{K} \frac{K_{i;\leqslant 0}}{\|\Delta_i^-\|^2} \ . \tag{111}$$

We postpone the proof of Theorem F.2 to the following Subsection F.4, which is divided between the proof of Equation (110) and Equation (111). We first explain how this result directly implies Theorem 3.2.

Proof of Theorem 3.2. Let A be a δ-correct algorithm over the class $\mathbb{D}_{\mathrm{cw}}$, with $\delta \leqslant 1/12$. Fix any matrix $\mathbf{\Delta} \in \mathbb{D}_{\mathrm{cw}}$ with CW $i^* = 1$. Consider χ as defined in Equation (109).

Combining the numerical bounds $8\log(4/3) < 3$ and $128\log(4/3) < 37$ with the definition of χ and the lower bounds (110) and (111), we obtain that there exist a sign convention $\mathbf{\Sigma}$ and a permutation $\pi \in \Pi(\mathbf{\Sigma})$ such that

$$\mathbb{P}_{\mathbf{\Delta}^{(\pi)},A}\left(N_\delta \geqslant \frac{1}{3}\cdot\max_{i\neq i^*}\frac{K_{i;\leqslant 0}}{\|\Delta_i^-\|^2}\log\left(\frac{1}{6\delta}\right) \vee \frac{1}{37\log(2K)}\sum_{i\neq i^*}\frac{K_{i;\leqslant 0}}{\|\Delta_i^-\|^2}\right) \geqslant \delta\,.$$

By Lemma F.1, if $\mathbf{\Delta}$ contains no ties, then the matrix $\mathbf{\Delta}^\pi$ admits a Condorcet winner and verifies all properties required of the matrix $\tilde{\mathbf{\Delta}}$ from Theorem 3.2. This proves Theorem 3.2 in the no-ties case. □

Remarks. *Observe that, by the second point of Lemma F.1, it is possible that $\mathbf{\Delta}^{(\pi)}$ admits no CW, when $\mathbf{\Delta}$ admit some ties. Yet it still admits a weak Condorcet winner, in the same that $\Delta^\pi_{1,\cdot}$ admits only non-negative entries, while all the other rows admit at least one negative entry. Then, one can construct a matrix $\mathbf{\Delta}^{\pi,\epsilon}$ as close as we need to $\mathbf{\Delta}^{(\pi)}$ and such that $\mathbf{\Delta}^{(\pi,\epsilon)}$ admit 1 as CW. We can for instance add a small constant $\epsilon > 0$ to the first row of $\mathbf{\Delta}^{(\pi,\epsilon)}$ [and $-\epsilon$ to the first column]. For ϵ small enough, the given matrix $\mathbf{\Delta}^{(\pi,\epsilon)}$ shares most properties of Δ.*

F.4 Proof of Theorem F.2

F.4.1 Proof of Equation (110) in Theorem F.2

Let $\delta \leqslant 1/12$. Let A be any δ-correct algorithm over the entire class $\mathbb{D}_{\mathrm{cw}}$. Fix $\mathbf{\Delta} \in \mathbb{D}_{\mathrm{cw}}$ with CW $i^*(\mathbf{\Delta}) = 1$. Recall the definition

$$\chi \coloneqq \inf\left\{x > 0 : \sup_{\Sigma}\sup_{\pi\in\Pi(\mathbf{\Sigma})}\mathbb{P}_{\mathbf{\Delta}^{(\pi)},A}(N_\delta > x) \leqslant \delta\right\}\,.$$

We prove Equation (110), namely

$$\chi \geqslant \frac{1}{8\log(4/3)}\max_{i=2}^{K}\frac{K_{i;\leqslant 0}}{\|\Delta_i^-\|^2}\log\left(\frac{1}{6\delta}\right).$$

Sketch of proof. We follow the three-step roadmap of Section F.1. The arguments are very similar to the proof of the quantile proof of Proposition 3.1 in paragraph F.2.2, except that we work with a local class of permuted instances and average over all permutations of $\mathbf{\Delta}$.

(i) Reference and alternative instances. We fix $\mathbf{\Delta} \in \mathbb{D}_{\mathrm{cw}}$ with $i^*(\mathbf{\Delta}) = 1$ and consider the *local class of reference* $(\Delta^{(\pi)})_{\pi\in\Pi}$ obtained by permuting, within each non-CW row, the positions of its negative entries according to π. For each suboptimal arm k, we construct the *alternative instance* $\mathbf{\Delta}^{(\pi,k)}$ by lifting to 0 all negative entries in row k, so that rows 1 and k become nonnegative and the instance has two weak CWs (hence lies outside $\mathbb{D}_{\mathrm{cw}}$). *(ii) Separating event and total variation.* As in Subsection F.2.2, we truncate A at the worst-case $(1-\delta)$-quantile χ to obtain a fixed-budget algorithm $\tilde{A}$, and we use the event $\{N_\delta > \chi\}$ as a separating event. This yields a total-variation lower bound $\mathrm{TV}(\mathbb{P}^{(\pi)}_{\tilde{A}}, \mathbb{P}^{(\pi,k)}_{\tilde{A}}) \geq 1 - 3\delta$, hence a KL lower bound of order $\log(1/(6\delta))$ for each k. *(iii) KL decomposition and extraction of $K_{k;\leqslant 0}/\|\Delta_k^-\|_2^2$.* We decompose the KL along pairs and use that $\mathbf{\Delta}^{(\pi)}$ and $\mathbf{\Delta}^{(\pi,k)}$ differ only on duels involving k. Averaging over permutations $\pi_k \in \Pi_k(\mathbf{\Sigma})$ symmetrizes the contribution of all negative entries in row k, and yields a lower bound $\chi \gtrsim K_{k;\leqslant 0}\|\Delta_k^-\|_2^{-2}\log(1/\delta)$. Finally, choosing the row k that maximizes $K_{k;\leqslant 0}/\|\Delta_k^-\|_2^2$ gives (110). *(iv) Tie-breaking convention.* In the last step, we a tie-breaking convention to conclude.

Proof. Step 1: Reference and alternative instances.

For now, fix a tie-breaking convention f_{tb} and a sign matrix $\mathbf{\Sigma}$ as in (106). The specific choice of $\mathbf{\Sigma}$ will be made in the final step of the proof.

Fix a permutation $\pi \in \Pi(\mathbf{\Sigma})$ (see Equation (107)), and consider the corresponding matrix $\mathbf{\Delta}^{(\pi)}$ defined in Equation (108). We denote by $\mathbb{P}_A^{(\pi)}$ the distribution of the observations induced by the interaction between algorithm A and the environment with gap matrix $\mathbf{\Delta}^{(\pi)}$.

Fix a suboptimal arm $k \in \{2, \dots, K\}$. The precise choice of k will be made explicit in the last step of the proof. We construct the gap matrix $\mathbf{\Delta}^{(\pi,k)}$ by setting to zero all entries (k, j) with $j \in \Sigma_k^- := \{j \in [K] : \sigma_{k,j} = -1\}$. Recall that, by the definition of $\mathbf{\Sigma}$ (see (106)), if $j \in \Sigma_k^-$ then $\Delta_{k,j}^{(\pi)} \leq 0$, that is Σ_k^- contains all arms that beat strictly k, together with some arms that are tied with k. We define

$$\mathbf{\Delta}_{i,j}^{(\pi,k)} := \begin{cases} \Delta_{i,j}^{(\pi)}, & \text{if } i \neq k \text{ and } j \neq k, \\ \Delta_{i,j}^{(\pi)}, & \text{if } (i = k \text{ and } \sigma_{k,j} = 1) \text{ or } (j = k \text{ and } \sigma_{k,i} = 1), \\ 0, & \text{if } (i = k \text{ and } \sigma_{k,j} = -1) \text{ or } (j = k \text{ and } \sigma_{k,i} = -1). \end{cases} \tag{112}$$

The matrices $\mathbf{\Delta}^{(\pi)}$ and $\mathbf{\Delta}^{(\pi,k)}$ differ only in row/column k. Moreover, by construction, rows 1 and k of $\mathbf{\Delta}^{(\pi,k)}$ contain only nonnegative entries, so that $\mathbf{\Delta}^{(\pi,k)} \notin \mathbb{D}_{\mathrm{cw}}$. We denote by $\mathbb{P}_A^{(\pi,k)}$ the distribution of the observations induced by the interaction between algorithm A and the environment with gap matrix $\mathbf{\Delta}^{(\pi,k)}$.

Step 2: information-theoretic arguments.

This step is identical to Step 2 in the proof of the quantile bound (95) (Section F.2.2). Consider the event $B := \{N_\delta \leqslant \chi\}$, and define the truncated algorithm $\tilde{A}$ as in that proof: run A up to time χ, returning $\tilde{i} = \hat{i}$ if A stops before χ, and $\tilde{i} = 0$ otherwise.

By the definition of χ in (109)—a uniform upper bound on the $(1-\delta)$-quantile of N_δ under any $\mathbb{P}_A^{(\pi)}$—we have

$$\mathbb{P}_{\tilde{A}}^{(\pi)}(\tilde{i} = 0) = \mathbb{P}_A^{(\pi)}(N_\delta > \chi) \leqslant \delta \, .$$

Since $\mathbf{\Delta}^{(\pi,k)}$ can be approximated by CW instances admitting 1 or k as CW (as in (98)–(99)), we obtain

$$\mathbb{P}_{\tilde{A}}^{(\pi,k)}(\tilde{i} \neq 0) \leqslant 2\delta \, .$$

Writing $B := \{\tilde{i} = 0\}$, this yields

$$\mathrm{TV}\big(\mathbb{P}_{\tilde{A}}^{(\pi)}, \mathbb{P}_{\tilde{A}}^{(\pi,k)}\big) \;\geqslant\; (1 - 2\delta) - \delta \;=\; 1 - 3\delta. \tag{113}$$

Remarks. *The result from Step 2 can be interpreted as a reduction scheme. Consider the signal detection problem of testing $H_0 : \mu = \mathbf{0}$ versus $H_1(\pi_k) : \mu = \big(\Delta_{k,\pi_k(\ell)}\big)_{\ell \in \Sigma_k^-}$ in a bandit setting. Equation* (113) *shows that, for any permutation π_k, $\tilde{A}$ is 2δ-correct for this signal detection problem, with a budget bounded by χ (independently of the permutation). This reduction is the main novelty of our proof technique. The remaining arguments in Step 3 build upon the literature on active signal detection [Castro, 2014, Saad et al., 2023, Graf et al., 2025].*

Step 3: computing the KL divergence.

By the Bretagnolle–Huber inequality (see, e.g., Lattimore and Szepesvári, 2020), the conclusion of Step 2 (113) implies

$$\mathrm{KL}\big(\mathbb{P}_{\tilde{A}}^{(\pi,k)}, \mathbb{P}_{\tilde{A}}^{(\pi)}\big) \;\geqslant\; \log\!\left(\frac{1}{6\delta}\right). \tag{114}$$

We now compute $\mathrm{KL}\big(\mathbb{P}_{\tilde{A}}^{(\pi,k)}, \mathbb{P}_{\tilde{A}}^{(\pi)}\big)$. Observe that we take the law under the alternative instance $\mathbf{\Delta}^{(\pi,k)}$ in the left side of the divergence: this ensures that the expectation $\mathbb{E}_{\tilde{A}}^{(\pi,k)}[N_{\{k,j\}}]$ appearing

in the KL decomposition (below) does not depend on π_k, which will allow us to average over permutations $\pi_k \in \Pi_k(\mathbf{\Sigma})$.

By the standard decomposition of the Kullback–Leibler divergence for adaptive procedures (see, e.g., Lattimore and Szepesvári, 2020, Lemma 15.1),

$$\begin{aligned}\mathrm{KL}\big(\mathbb{P}_{\tilde{A}}^{(\pi,k)}, \mathbb{P}_{\tilde{A}}^{(\pi)}\big) &= \sum_{i \neq j} \mathbb{E}_{\tilde{A}}^{(\pi,k)}\big[N_{i,j}\big]\, \mathrm{KL}\big(\mathbb{P}_{i,j}^{(\pi,k)}, \mathbb{P}_{i,j}^{(\pi)}\big) \\ &= \sum_{j \in \Sigma_k^-} \mathbb{E}_{\tilde{A}}^{(\pi,k)}\big[N_{\{k,j\}}\big]\, \mathrm{KL}\big(\mathbb{P}_{k,j}^{(\pi,k)}, \mathbb{P}_{k,j}^{(\pi)}\big)\ , \end{aligned} \tag{115}$$

since the two instances differ only on pairs (k, j) or (j, k) with $j \in \Sigma_k^-$, by construction of $\mathbf{\Delta}^{(\pi,k)}$.

Now fix $j \in \Sigma_k^-$. By the definitions of $\mathbf{\Delta}^{(\pi)}$ and $\mathbf{\Delta}^{(\pi,k)}$ (Equations (108) and (112)), we have

$$\mathbb{P}_{k,j}^{(\pi,k)} = \mathcal{B}\big(\tfrac{1}{2}\big)\,, \qquad \mathbb{P}_{k,j}^{(\pi)} = \mathcal{B}\big(\tfrac{1}{2} + \Delta_{k,\pi_k(j)}\big)\,,$$

where $\Delta_{k,\pi_k(j)} \in [-1/4, 0]$. Hence, using the bound on kl from (103),

$$\mathrm{KL}\big(\mathbb{P}_{k,j}^{(\pi,k)}, \mathbb{P}_{k,j}^{(\pi)}\big) \leqslant\ 8 \log\!\left(\frac{4}{3}\right) \Delta_{k,\pi_k(j)}^2\ ,$$

Plugging these bounds into (115), we obtain

$$\mathrm{KL}\big(\mathbb{P}_{\tilde{A}}^{(\pi,k)}, \mathbb{P}_{\tilde{A}}^{(\pi)}\big) \leqslant 8 \log\!\left(\frac{4}{3}\right) \sum_{j \in \Sigma_k^-} \mathbb{E}_{\tilde{A}}^{(\pi,k)}[N_{\{k,j\}}]\, \Delta_{k,\pi_k(j)}^2\ .$$

A key property of our construction is that $\mathbf{\Delta}^{(\pi,k)}$ *does not actually depend on* π_k: all entries permuted by π_k in row k of $\mathbf{\Delta}^{(\pi)}$ are set to 0 under $\mathbf{\Delta}^{(\pi,k)}$. Consequently, $\mathbb{E}_{\tilde{A}}^{(\pi,k)}$ does not depend on π_k. Given $\pi'_k \in \Pi_k(\mathbf{\Sigma})$, we denote $\pi' = (\pi_2, \ldots, \pi_{k-1}, \pi'_k, \pi_{k+1}, \ldots)$ where we only change π_k. Averaging the previous inequality over $\pi'_k \in \Pi_k(\mathbf{\Sigma})$ while keeping $(\pi_l)_{l \neq k}$ fixed, we get

$$\begin{aligned}\frac{1}{|\Pi_k(\mathbf{\Sigma})|} \sum_{\pi'_k \in \Pi_k(\mathbf{\Sigma})} \mathrm{KL}\big(\mathbb{P}_{\tilde{A}}^{(\pi,k)}, \mathbb{P}_{\tilde{A}}^{(\pi')}\big) &\leqslant 8 \log\!\left(\frac{4}{3}\right) \frac{1}{|\Pi_k(\mathbf{\Sigma})|} \sum_{\pi'_k \in \Pi_k(\mathbf{\Sigma})} \sum_{j \in \Sigma_k^-} \mathbb{E}_{\tilde{A}}^{(\pi,k)}[N_{\{k,j\}}]\, \Delta_{k,\pi'_k(j)}^2 \\ &= 8 \log\!\left(\frac{4}{3}\right) \frac{1}{|\Sigma_k^-|} \left(\sum_{j \in \Sigma_k^-} \mathbb{E}_{\tilde{A}}^{(\pi,k)}[N_{\{k,j\}}]\right) \left(\sum_{j \in \Sigma_k^-} \Delta_{k,j}^2\right)\ , \end{aligned} \tag{116}$$

where we used Lemma H.3 to symmetrize over all permutations π'_k of Σ_k^-.

By definition of Σ_k^-, it holds that $\Sigma_k^- \subset \{j \in [K] \setminus \{k\} : \Delta_{k,j} \leqslant 0\}$, so that

$$\sum_{j \in \Sigma_k^-} \Delta_{k,j}^2 = \|\Delta_k^-\|^2\ .$$

Moreover, by construction, the modified algorithm $\tilde{A}$ has a budget upper bounded by χ, and

$$\sum_{j \in \Sigma_k^-} \mathbb{E}_{\tilde{A}}^{(\pi,k)}[N_{\{k,j\}}]\ \leqslant\ \chi\,.$$

From there, we get

$$\frac{1}{|\Pi_k(\mathbf{\Sigma})|} \sum_{\pi'_k \in \Pi_k(\mathbf{\Sigma})} \mathrm{KL}\big(\mathbb{P}_{\tilde{A}}^{(\pi,k)}, \mathbb{P}_{\tilde{A}}^{(\pi')}\big) \leqslant 8 \log\!\left(\frac{4}{3}\right) \frac{\|\Delta_k^-\|^2}{|\Sigma_k^-|}\, \chi\ . \tag{117}$$

Finally, combining (117) with the Bretagnolle–Huber bound (114), we obtain

$$\chi \geqslant \frac{1}{8 \log(4/3)} \frac{|\Sigma_k^-|}{\|\Delta_k^-\|^2} \log\!\left(\frac{1}{6\delta}\right)\ . \tag{118}$$

Step 4: choice of convention $\mathbf{\Sigma}$ and conclusion.

It remains to choose an appropriate arm k, and a convention f_{tb} in the definition of the sign matrix $\mathbf{\Sigma}$ (see Equation (106)). Consider

$$k^* \in \arg\max_{k=2}^{K} \frac{K_{k;\leqslant 0}}{\|\Delta_k^-\|^2} \,,$$

as a suboptimal arm for which detecting a negative entry in its row is the most costly.

Fix a tie-breaking convention $f_{\mathrm{tb}} : [K]^2 \to \{-1, 0, 1\}$ such that $f_{\mathrm{tb}}(k^*, i) = -1$ for any $i \neq k^*$. For this choice, we have

$$\frac{|\Sigma_{k^*}^-|}{\|\Delta_{k^*}^-\|^2} = \frac{K_{k^*;\leqslant 0}}{\|\Delta_{k^*}^-\|^2} = \max_{i \neq i^*} \frac{K_{i;\leqslant 0}}{\|\Delta_i^-\|^2} \,.$$

Plugging this into (118) yields

$$\chi \geqslant \frac{1}{8\log(4/3)} \max_{i \neq i^*} \frac{K_{i;\leqslant 0}}{\|\Delta_i^-\|^2} \log\left(\frac{1}{6\delta}\right) \,,$$

which establishes the first inequality (110) on χ. □

F.4.2 Proof of Equation (111) in Theorem F.2

Let $\delta \leqslant 1/12$. Let A be any δ-correct algorithm over the entire class $\mathbb{D}_{\mathrm{cw}}$, and fix $\mathbf{\Delta} \in \mathbb{D}_{\mathrm{cw}}$ with CW $i^*(\mathbf{\Delta}) = 1$. In this subsection, we prove Equation (111) from Theorem F.2, using the same instance construction as in the proof of bound (110). Recall this bound:

$$\chi \geqslant \frac{1}{64\log(4/3)} \frac{1}{\log(2K)} \sum_{i=2}^{K} \frac{K_{i;\leqslant 0}}{\|\Delta_i^-\|^2}.$$

Sketch of proof. We follow the three-step roadmap of Section F.1, reusing the instance construction from the proof of (110) but with a more refined separating event. The key differences are (ii) a refined event B_k that also controls the number of duels involving arm k, and (iii) the use of Pinsker's inequality and Jensen to average over multiple arms simultaneously.

(i) Reference and alternative instances. As before, we consider the local class $(\mathbf{\Delta}^{(\pi)})_{\pi \in \Pi(\mathbf{\Sigma})}$ and, for each $k \in \{2, \dots, K\}$, the alternative $\mathbf{\Delta}^{(\pi,k)}$. We define the row-wise hardness $\beta_k^2 = \|\Delta_k^-\|^2/|\Sigma_k^-|$, assumed ordered increasingly. Let I be the index that maximizes $(k-1)/\beta_k^2$, corresponding to the worst-case average hardness over the first k rows. *(ii) Separating event and total variation.* Instead of $\{N_\delta \leq \chi\}$, we use the event $B_k = \{N_\delta \leqslant \chi,\, N_{k,\cdot} \leqslant 4\chi/(I-1)\}$, where $N_{k,\cdot}$ counts duels between k and opponents in Σ_k^-. Define a truncation $\tilde{A}_k$ that outputs $\mathbb{1}_{B_k}$, using budget at most $4\chi/(I-1)$ on k. Under $\mathbf{\Delta}^{(\pi,k)}$, $\mathbb{P}(\mathbb{1}_{B_k} = 1) \leqslant 2\delta$. Under $\mathbf{\Delta}^{(\pi)}$, we use a pigeonhole argument and the average probability is larger than $3/4 - \delta$. This yields averaged TV larger than $1/2$—the key novelty of this proof technique. *(iii) KL decomposition and extraction of the sum.* By Pinsker's inequality, the averaged TV lower bound implies an averaged KL lower bound. Each KL is upper bounded as previously, and now we average over $k = 2, \dots, I$ and use the truncation constraint $N_{k,\cdot} \leqslant 4\chi/(I-1)$ to get $\chi \gtrsim (I-1)/\beta_I^2$. By the definition of I, this gives $\chi \gtrsim \sum_{i=2}^{K} K_{i;\leqslant 0}/\|\Delta_i^-\|_2^2$, up to logarithmic factors.

Proof. Step 1: construction of instances. Fix a tie-breaking convention f_{tb} and a sign matrix $\mathbf{\Sigma}$ (see (106)). Again, they will be chosen in the last step of the proof. Fix a permutation $\pi \in \Pi(\mathbf{\Sigma})$ and use $\mathbf{\Delta}^{(\pi)}$ as the reference matrix (see (108)). Define, for each $k \in \{2, \dots, K\}$, the row-wise hardness

$$\beta_k^2(\Sigma) \coloneqq \frac{\|\Delta_k^-\|^2}{|\Sigma_k^-|} \,. \tag{119}$$

Without loss of generality, assume that the arms are ordered so that $\beta_2^2 \leqslant \beta_3^2 \leqslant \cdots \leqslant \beta_K^2$. Define

$$I \coloneqq \arg\max_{k=2,\dots,K} \frac{k-1}{\beta_k^2} \,, \tag{120}$$

the index that captures the worst-case average hardness over the first k rows.

As alternative instances, we consider the family $\{\mathbf{\Delta}^{(\pi,k)}\}_{k=2,\dots,I}$.

Step 2: information-theoretic arguments. In this step, we construct an event under which algorithm A should behave differently depending on whether it interacts with $\mathbf{\Delta}^{(\pi)}$ or $\mathbf{\Delta}^{(\pi,k)}$. To capture the sum lower bound (111), we need a more refined event than in the proof of (110). For $k \in \{2,\dots,I\}$, define

$$B_k := \{N_\delta \leqslant \chi\} \cap \left\{ N_{k,\cdot} \leqslant \frac{4\chi}{I-1} \right\}, \tag{121}$$

where

$$N_{k,\cdot} := \sum_{i \in \Sigma_k^-} N_{\{k,i\}}$$

denotes the total number of duels involving arm k against opponents in Σ_k^-.

Remarks. *The event B_k is designed as follows. The bound* (110) *shows that the quantity β_k^{-2} characterizes the budget needed to find a negative entry in row k of $\mathbf{\Delta}^{(\pi)}$, uniformly over all permutations π. . Identifying the CW $i^* = 1$ amounts to solving simultaneously $K-1$ such signal detection problems, one per suboptimal row. By the definition of I in* (120) *and Lemma H.1, it is natural to think of the simplified regime where $(\beta_2^{-2},\dots,\beta_I^{-2})$ are of the same order, so that arms $2,\dots,I$ are equally hard to eliminate. In that case, any reasonable algorithm should allocate its samples roughly uniformly across rows $2,\dots,I$, and the event B_k describes this expected behavior for a δ-correct algorithm A.*

We can compute the event B_k with a truncated procedure $\tilde{A}_k$ that uses a total budget at most χ, and at most $4\chi/(I-1)$ comparisons involving arm k against an opponent in Σ_k^-.

Define the following procedure $\tilde{A}_k$. For $t = 1,\dots,\chi$, run algorithm A. During the process, stop as soon as the algorithm tries to sample more than $4\chi/(I-1)$ samples on the row k, or more than χ samples in total and return 0. Otherwise, the procedure returns 1 if the sampling stops by χ without exceeding the cap. By construction, the binary decision ψ_k computed by $\tilde{A}_k$ satisfies $\psi_k = \mathbb{1}_{B_k}$.

We write $\mathbb{P}_{\tilde{A}_k}^{(\pi)}$ (resp. $\mathbb{P}_{\tilde{A}_k}^{(\pi,k)}$) for the distribution induced by the interaction between algorithm $\tilde{A}_k$ and the environment with gap matrix $\mathbf{\Delta}^{(\pi)}$ (resp. $\mathbf{\Delta}^{(\pi,k)}$). We now lower bound the total variation distance between $\mathbb{P}_{\tilde{A}_k}^{(\pi)}$ and $\mathbb{P}_{\tilde{A}_k}^{(\pi,k)}$ using the event B_k.

First, $B_k \subset \{N_\delta \leqslant \chi\}$. Since A is δ-correct over $\mathbb{D}_{\mathrm{cw}}$ and $\mathbf{\Delta}^{(\pi,k)}$ can be approximated by instances in $\mathbb{D}_{\mathrm{cw}}$ as in the proof of (99), we obtain, for any $k \neq 1$,

$$\mathbb{P}_{\tilde{A}_k}^{(\pi,k)}(B_k) \leqslant \mathbb{P}_{\tilde{A}_k}^{(\pi,k)}(N_\delta \leqslant \chi) \leqslant 2\delta \ . \tag{122}$$

Next, consider $\mathbb{P}_{\tilde{A}_k}^{(\pi)}$ and the complement B_k^c. We have

$$B_k^c = \{N_\delta > \chi\} \ \cup \ \{N_\delta \leqslant \chi,\ N_{k,\cdot} > 4\chi/(I-1)\}.$$

Since A is δ-correct and $\mathbf{\Delta}^{(\pi)} \in \mathbb{D}_{\mathrm{cw}}$, the definition of χ implies

$$\mathbb{P}_{\tilde{A}_k}^{(\pi)}(N_\delta > \chi) = \mathbb{P}_A^{(\pi)}(N_\delta > \chi) \leqslant \delta \ . \tag{123}$$

We now average the second term of B_k^c over $k \in \{2,\dots,I\}$:

$$\frac{1}{I-1}\sum_{k=2}^{I} \mathbb{P}_{\tilde{A}_k}^{(\pi)}\big(N_\delta \leqslant \chi,\ N_{k,\cdot} > 4\chi/(I-1)\big) = \mathbb{E}_A^{(\pi)}\left[\frac{1}{I-1}\sum_{k=2}^{I} \mathbb{1}_{\{N_\delta \leqslant \chi,\ N_{k,\cdot} > 4\chi/(I-1)\}}\right].$$

Since $\sum_{k=2}^{K} N_{k,\cdot} \leqslant N_\delta$, on the event $\{N_\delta \leqslant \chi\}$ at most $(I-1)/4$ indices $k \in \{2,\dots,I\}$ can satisfy $N_{k,\cdot} > 4\chi/(I-1)$. Hence,

$$\frac{1}{I-1}\sum_{k=2}^{I} \mathbb{1}_{\{N_\delta \leqslant \chi,\ N_{k,\cdot} > 4\chi/(I-1)\}} \leqslant \frac{1}{4},$$

which yields

$$\frac{1}{I-1}\sum_{k=2}^{I}\mathbb{P}^{(\pi)}_{\tilde{A}_k}\big(N_\delta \leqslant \chi,\ N_{k,\cdot} > 4\chi/(I-1)\big) \leqslant \frac{1}{4}\ . \tag{124}$$

Combining (122), (123), and (124), we obtain

$$\begin{aligned}\frac{1}{I-1}\sum_{k=2}^{I}\mathrm{TV}\big(\mathbb{P}^{(\pi,k)}_{\tilde{A}_k},\mathbb{P}^{(\pi)}_{\tilde{A}_k}\big) &\geqslant \frac{1}{I-1}\sum_{k=2}^{I}\Big(\mathbb{P}^{(\pi,k)}_{\tilde{A}_k}(B_k^c)-\mathbb{P}^{(\pi)}_{\tilde{A}_k}(B_k^c)\Big)\\ &\geqslant 1-2\delta-\frac{1}{I-1}\sum_{k=2}^{I}\mathbb{P}^{(\pi)}_{\tilde{A}_k}(B_k^c)\\ &\geqslant (1-2\delta)-\left(\delta+\frac{1}{4}\right)\\ &\geqslant \frac{1}{2}\,,\end{aligned}$$

where the last inequality uses the assumption $\delta \leqslant 1/12$. Finally, we apply a data-processing inequality. In this regime which does not depend on δ, we use Pinsker's inequality which implies that

$$\frac{1}{2} \leqslant \frac{1}{I-1}\sum_{k=2}^{I}\mathrm{TV}\big(\mathbb{P}^{(\pi,k)}_{\tilde{A}_k},\mathbb{P}^{(\pi)}_{\tilde{A}_k}\big) \leqslant \frac{1}{I-1}\sum_{k=2}^{I}\sqrt{\frac{1}{2}\,\mathrm{KL}\big(\mathbb{P}^{(\pi,k)}_{\tilde{A}_k},\mathbb{P}^{(\pi)}_{\tilde{A}_k}\big)}\ . \tag{125}$$

Remarks. *We can again interpret this result as a reduction argument. Averaging* (122)*,* (123)*, and* (124) *over* $\pi \sim \mathcal{U}\mathrm{nif}(\Pi(\mathbf{\Sigma}))$*, we obtain that there exists some* $k \in \{2,\dots,I\}$ *(independent of* π*) such that* $\tilde{A}_k$ *is* $1/2$*-correct, with budget at most* $4\chi/(I-1)$*, for the active signal detection problem*

$$H_0 : \mu = \mathbf{0} \quad vs \quad H_1 : \mu = \big(\Delta_{k,\pi_k(\ell)}\big)_{\ell\in\Sigma_k^-},\quad \pi_k \sim \mathcal{U}\mathrm{nif}(\Pi_k(\mathbf{\Sigma}))\,.$$

Step 3: computing the KL divergence. We now conclude by computing the KL divergence above. Fix $k \in \{2,\dots,I\}$. Given $\pi'_k \in \Pi_k(\mathbf{\Sigma})$, we write $\pi' = (\pi_2,\dots,\pi_{k-1},\pi'_k,\pi_{k+1},\dots,)$. Using the same computation as in Equation (117), we obtain

$$\begin{aligned}\frac{1}{|\Pi_k(\mathbf{\Sigma})|}\sum_{\pi_k\in\Pi_k(\mathbf{\Sigma})}\mathrm{KL}\big(\mathbb{P}^{(\pi,k)}_{\tilde{A}_k},\mathbb{P}^{(\pi')}_{\tilde{A}_k}\big) &\leqslant 8\log\left(\frac{4}{3}\right)\sum_{j\in\Sigma_k^-}\mathbb{E}^{(\pi,k)}_{\tilde{A}_k}[N_{\{k,j\}}]\,\beta_k^2\\ &\leqslant 8\log\left(\frac{4}{3}\right)\cdot\frac{4\chi}{I-1}\cdot\beta_I^2\ ,\end{aligned}$$

where the last inequality uses the facts that $\sum_{j\in\Sigma_k^-}\mathbb{E}^{(\pi,k)}_{\tilde{A}_k}[N_{\{k,j\}}] \leqslant 4\chi/(I-1)$ by construction of $\tilde{A}_k$, and that, by our ordering assumption, $\beta_k^2 \leqslant \beta_I^2$ for all $k \in \{2,\dots,I\}$. Averaging additionally over all permutations $\pi = (\pi_2,\dots,\pi_K) \in \Pi(\mathbf{\Sigma})$, we obtain

$$\frac{1}{|\Pi(\mathbf{\Sigma})|}\sum_{\pi\in\Pi(\mathbf{\Sigma})}\frac{1}{I-1}\sum_{k=2}^{I}\mathrm{KL}\big(\mathbb{P}^{(\pi,k)}_{\tilde{A}_k},\mathbb{P}^{(\pi)}_{\tilde{A}_k}\big) \leqslant 8\log\left(\frac{4}{3}\right)\cdot\frac{4\chi}{I-1}\cdot\beta_I^2\ . \tag{126}$$

Finally, combining Pinsker's inequality (125) with Jensen's inequality, we get

$$\begin{aligned}\frac{1}{2} &\leqslant \frac{1}{|\Pi(\mathbf{\Sigma})|}\sum_{\pi\in\Pi(\mathbf{\Sigma})}\frac{1}{I-1}\sum_{k=2}^{I}\sqrt{\frac{1}{2}\,\mathrm{KL}\big(\mathbb{P}^{(\pi,k)}_{\tilde{A}_k},\mathbb{P}^{(\pi)}_{\tilde{A}_k}\big)}\\ &\leqslant \sqrt{\frac{1}{2}\cdot\frac{1}{|\Pi(\mathbf{\Sigma})|}\sum_{\pi\in\Pi(\mathbf{\Sigma})}\frac{1}{I-1}\sum_{k=2}^{I}\mathrm{KL}\big(\mathbb{P}^{(\pi,k)}_{\tilde{A}_k},\mathbb{P}^{(\pi)}_{\tilde{A}_k}\big)}\\ &\leqslant \sqrt{\frac{1}{2}\cdot 8\log\left(\frac{4}{3}\right)\cdot\frac{4\chi}{I-1}\cdot\beta_I^2}\ ,\end{aligned}$$

where the last inequality follows from (126). Rearranging yields

$$\chi \geqslant \frac{1}{64\log(4/3)}\frac{I-1}{\beta_I^2} = \frac{1}{64\log(4/3)}\max_{i=2,\dots,K}\frac{i-1}{\beta_i^2} \ , \tag{127}$$

where the second equality follows from the definition of I (see (120)). From Lemma H.1, we have

$$\max_{i=2,\dots,K}\frac{i-1}{\beta_i^2} \geqslant \frac{1}{\log(2K)}\sum_{i=2}^{K}\frac{1}{\beta_i^2} = \frac{1}{\log(2K)}\sum_{i=2}^{K}\frac{|\Sigma_i^-|}{\|\Delta_i^-\|^2} \ .$$

Step 4: choice of convention $\mathbf{\Sigma}$ and conclusion.

We claim that there exists a tie-breaking convention f_{tb} that satisfies

$$\sum_{i=2}^{K}\frac{|\Sigma_i^-|}{\|\Delta_i^-\|^2} \geqslant \frac{1}{2}\sum_{i=2}^{K}\frac{K_{i;\leqslant 0}}{\|\Delta_i^-\|^2} \ . \tag{128}$$

Then, the conclusion (111) directly follows from the (127) together with (128).

We finally finish with a technical construction of a tie-breaking that satisfies (128). Consider any suboptimal arm $i \neq 1$. For any $j \neq i$, it holds that

$$\big(j \in \Sigma_i^-\big) \iff \big(\Delta_{i,j} < 0\big) \text{ or } \big(\Delta_{i,j} = 0,\ f_{\mathrm{tb}}(i,j) = -1\big),$$

so that

$$|\Sigma_i^-| = \sum_{\substack{j=1\\ j\neq i}}^{K} \mathbb{1}_{\{\Delta_{i,j}<0\}} + \mathbb{1}_{\{\Delta_{i,j}=0\}}\mathbb{1}_{\{f_{\mathrm{tb}}(i,j)=-1\}} \ .$$

Summing over $i = 2, \dots, K$ gives

$$\begin{aligned}
\sum_{i=2}^{K}\frac{|\Sigma_i^-|}{\|\Delta_i^-\|^2} &= \sum_{i=2}^{K}\sum_{\substack{j=1\\ j\neq i}}^{K}\frac{1}{\|\Delta_i^-\|^2}\left(\mathbb{1}_{\{\Delta_{i,j}<0\}} + \mathbb{1}_{\{\Delta_{i,j}=0\}}\mathbb{1}_{\{f_{\mathrm{tb}}(i,j)=-1\}}\right) \\
&= \sum_{i=2}^{K}\sum_{\substack{j=1\\ j\neq i}}^{K}\frac{1}{\|\Delta_i^-\|^2}\mathbb{1}_{\{\Delta_{i,j}<0\}} + \sum_{i=2}^{K}\frac{1}{\|\Delta_i^-\|^2}\mathbb{1}_{\{\Delta_{i,1}=0\}}\mathbb{1}_{\{f_{\mathrm{tb}}(i,1)=-1\}} \\
&\quad + \sum_{2\leqslant i<j\leqslant K}\mathbb{1}_{\{\Delta_{i,j}=0\}}\left(\frac{\mathbb{1}_{\{f_{\mathrm{tb}}(i,j)=-1\}}}{\|\Delta_i^-\|^2} + \frac{\mathbb{1}_{\{f_{\mathrm{tb}}(i,j)=1\}}}{\|\Delta_j^-\|^2}\right) \ ,
\end{aligned} \tag{129}$$

where the last equality follows from a simple reorganization of the sum.

We now choose f_{tb} to make the second and third terms as large as possible. Let f_{tb} be the antisymmetric function $f_{\mathrm{tb}} : [K]^2 \to \{-1, 0, 1\}$, defined for any $i < j$ by

$$f_{\mathrm{tb}}(i,j) \coloneqq \begin{cases} 1, & \text{if } i = 1,\ j > 1, \\ \mathbb{1}_{\{\|\Delta_i^-\| > \|\Delta_j^-\|\}} - \mathbb{1}_{\{\|\Delta_i^-\| < \|\Delta_j^-\|\}}, & \text{if } 2 \leq i < j \text{ and } \|\Delta_i^-\| \neq \|\Delta_j^-\|, \\ 1, & \text{if } 2 \leq i < j \text{ and } \|\Delta_i^-\| = \|\Delta_j^-\|. \end{cases}$$

The convention f_{tb} for the row of the CW implies that

$$\sum_{i=2}^{K}\frac{1}{\|\Delta_i^-\|^2}\mathbb{1}_{\{\Delta_{i,1}=0\}}\mathbb{1}_{\{f_{\mathrm{tb}}(i,1)=-1\}} = \sum_{i=2}^{K}\frac{1}{\|\Delta_i^-\|^2}\mathbb{1}_{\{\Delta_{i,1}=0\}} \ . \tag{130}$$

Moreover, the expression of $f_{\mathrm{tb}}(i,j)$ for $2 \leqslant i < j \leqslant K$ is chosen so that

$$\begin{aligned}
\sum_{2\leqslant i<j\leqslant K}\mathbb{1}_{\{\Delta_{i,j}=0\}}\left(\frac{\mathbb{1}_{\{f_{\mathrm{tb}}(i,j)=-1\}}}{\|\Delta_i^-\|^2} + \frac{\mathbb{1}_{\{f_{\mathrm{tb}}(i,j)=1\}}}{\|\Delta_j^-\|^2}\right) &= \sum_{2\leqslant i<j\leqslant K}\mathbb{1}_{\{\Delta_{i,j}=0\}}\left(\frac{1}{\|\Delta_i^-\|^2} \vee \frac{1}{\|\Delta_j^-\|^2}\right) \\
&\geqslant \frac{1}{2}\sum_{2\leqslant i<j\leqslant K}\mathbb{1}_{\{\Delta_{i,j}=0\}}\left(\frac{1}{\|\Delta_i^-\|^2} + \frac{1}{\|\Delta_j^-\|^2}\right) \ .
\end{aligned} \tag{131}$$

Finally, gathering Equations (129), (130), and (131), we get

$$\sum_{i=2}^{K} \frac{|\Sigma_i^-|}{\|\Delta_i^-\|^2} \geqslant \sum_{i=2}^{K} \sum_{\substack{j=1 \\ j\neq i}}^{K} \frac{1}{\|\Delta_i^-\|^2} \left(\mathbb{1}_{\{\Delta_{i,j}<0\}} + \mathbb{1}_{\{j=1,\ \Delta_{i,1}=0\}} + \frac{1}{2}\mathbb{1}_{\{j\neq 1,\ \Delta_{i,j}=0\}} \right)$$
$$\geqslant \frac{1}{2} \sum_{i=2}^{K} \frac{K_{i;\leqslant 0}}{\|\Delta_i^-\|^2} \ ,$$

which proves Equation (128), hence finishing the proof. □

F.5 Proof of Corollary 3.3

Proof. Without loss of generality, assume $i^* = 1$. Let $\underline{s} = (s_i)_{i\neq 1}$ with $s_i \in [K/8]$ for all $i \neq i^*$, and let $\underline{\Delta} = (\Delta_i)_{i\neq i^*}$ with $\Delta_i \in (0, 1/4)$. Let $\epsilon > 0$ be small.

For simplicity, assume K is a multiple of 8, and set $d = K/2$. We construct the $K \times K$ antisymmetric matrix $M^\epsilon = M^\epsilon(\underline{s}, \underline{\Delta})$:

$$M^\epsilon = \begin{pmatrix} A & -D \\ D^\top & \Lambda \end{pmatrix}, \tag{132}$$

where A, D, and Λ are $d \times d$ matrices defined below.

The matrix A is the $d \times d$ antisymmetric matrix with first row $A_{1,\cdot} = (0, \epsilon, \dots, \epsilon) \in \mathbb{R}^d$, and for $i = 2, \dots, d$,

$$A_{i,\cdot} = (-\epsilon, \dots, -\epsilon, \underbrace{0}_{i\text{-th}}, \epsilon, \dots, \epsilon) \in \mathbb{R}^d.$$

The matrix D is the $d \times d$ matrix with nonnegative entries where $D_{1,\cdot} = (-\epsilon, \dots, -\epsilon) \in \mathbb{R}^d$, and for $i = 2, \dots, d$,

$$D_{i,\cdot} = (\underbrace{\Delta_i, \dots, \Delta_i}_{s_i \text{ times}}, \epsilon, \dots, \epsilon) \in \mathbb{R}^d,$$

which is possible since $s_i \leqslant d$.

To construct Λ, recall d is a multiple of 4 and $s_i \in \{1, \dots, d/4\}$. Define Λ as the block matrix:

$$\Lambda = \begin{pmatrix} J_\epsilon & -\Lambda^{(0)} & \epsilon\mathbf{1} & (\Lambda^{(3)})^\top \\ (\Lambda^{(0)})^\top & J_\epsilon & -\Lambda^{(1)} & \epsilon\mathbf{1} \\ -\epsilon\mathbf{1} & (\Lambda^{(1)})^\top & J_\epsilon & -\Lambda^{(2)} \\ -\Lambda^{(3)} & -\epsilon\mathbf{1} & (\Lambda^{(2)})^\top & J_\epsilon \end{pmatrix},$$

where $\mathbf{1}$ is the $(d/4) \times (d/4)$ all-ones matrix, J_ϵ is the $(d/4) \times (d/4)$ antisymmetric matrix with ϵ above the diagonal, and for $l \in \{0, \dots, 3\}$, $\Lambda^{(l)}$ is the $(d/4) \times (d/4)$ matrix where the i-th row is

$$\Lambda^{(l)}_{i,\cdot} = (\underbrace{\Delta_j, \dots, \Delta_j}_{s_j \text{ times}}, \epsilon, \dots, \epsilon) \in \mathbb{R}^{d/4}, \quad j = d + \tfrac{d}{4}l + i.$$

The matrix $M^\epsilon = M(\underline{s}, \underline{\Delta}, \epsilon)$ is clearly antisymmetric. Its first row is $(0, \epsilon, \dots, \epsilon)$, so $i^* = 1$ and $M^\epsilon \in \mathbb{D}_{\mathrm{cw}}$.

For each arm $i = 2, \dots, K$, row i of M^ϵ contains exactly s_i entries of magnitude $-\Delta_i$, with all other negative entries equal to $-\epsilon$. For sufficiently small $\epsilon > 0$, the optimal sparsity $\boldsymbol{s}^*_{M^\epsilon}$ achieving the minimum in (4) equals $\underline{s}$, with associated gaps $(M^\epsilon_{i,(s_i)})_{i\neq i^*} = (-\Delta_i)_{i\neq i^*}$. Moreover, M^ϵ has no ties since $\Delta_i \neq 0$ and $\epsilon > 0$.

Consider Corollary 3.3. Let $\delta \leqslant 1/12$. For $\boldsymbol{\Delta} \in \mathbb{D}_{\mathrm{cw}}$, construction yields $\epsilon > 0$ small such that $M^\epsilon(\boldsymbol{s}^*_{\boldsymbol{\Delta}}, -\boldsymbol{\Delta}_{(s^*)}) \in \mathbb{D}(\boldsymbol{\Delta})$ with no ties.

Theorem 3.2 applied to M^ϵ gives $\tilde{\boldsymbol{\Delta}} \in \mathbb{D}(M^\epsilon) = \mathbb{D}(\boldsymbol{\Delta})$ satisfying

$$\mathbb{P}_{\tilde{\boldsymbol{\Delta}},A}\left(N_\delta \ \geqslant \ \frac{1}{3} \max_{i\neq i^*} \frac{K^\epsilon_{i;<0}}{\|(M^\epsilon_i)^-\|^2} \log\left(\frac{1}{6\delta}\right) \vee \frac{1}{37\log(2K)} \sum_{i\neq i^*} \frac{K^\epsilon_{i;<0}}{\|(M^\epsilon_i)^-\|^2} \right) \ \geqslant \ \delta,$$

where $K^{\epsilon}_{i;<0}$ counts negative entries of row i. By construction, $K^{\epsilon}_{i;<0} \geqslant K/8$ for $i = 2, \dots, K$. For small ϵ,

$$s_i^* \Delta^2_{i,(s_i^*)} \leqslant \|(M_i^{\epsilon})^-\|^2 \leqslant s_i^* \Delta^2_{i,(s_i^*)} + (K - s_i^*)\epsilon^2, \leqslant 2 s_i^* \Delta^2_{i,(s_i^*)}$$

yielding

$$\mathbb{P}_{\tilde{\boldsymbol{\Delta}},A}\left(N_\delta \;\geqslant\; \frac{1}{48} \max_{i \neq i^*} \frac{K}{s_i^* \Delta^2_{i,(s_i^*)}} \log\left(\frac{1}{6\delta}\right) \vee \frac{1}{592 \log(2K)} \sum_{i \neq i^*} \frac{K}{s_i^* \Delta^2_{i,(s_i^*)}} \right) \;\geqslant\; \delta.$$

This scales as $H_{\mathrm{explore}}(\boldsymbol{s}^*)$ up to $\log K$ factors, proving the first part of Corollary 3.3.

The $H_{\mathrm{certify}}(\boldsymbol{s}^*, \delta)$ term follows from the quantile bound in Proposition 3.1: by construction of M^{ϵ}, we have $\tilde{\Delta}_{i,(1)} = \tilde{\Delta}_{i,(s_i^*)}$ for all $i \neq i^*$, so

$$H_{\mathrm{certify}}(\boldsymbol{s}^*, \delta) = \sum_{i \neq i^*} \frac{\log(1/\delta)}{\Delta^2_{i,(s_i^*)}} = \sum_{i \neq i^*} \frac{\log(1/\delta)}{\tilde{\Delta}^2_{i,(1)}},$$

and the lower bound applies directly. □

F.6 Lower Bounds Preserving CW Row Structure

Consider $\boldsymbol{\Delta} \in \mathbb{D}_{\mathrm{cw}}$. For simplicity, assume that $\boldsymbol{\Delta}$ has no ties.[7] Let $\boldsymbol{\Sigma}$ be its sign matrix as in Equation (106), and let $\pi \in \Pi(\boldsymbol{\Sigma})$ (see (107)) with associated matrix $\boldsymbol{\Delta}^{\pi}$ defined in Equation (108).

By Lemma F.1, $\boldsymbol{\Delta}^{\pi}$ has the same gap structure as $\boldsymbol{\Delta}$: it preserves all signs (hence all pairwise preferences) and gap magnitudes up to reordering. However, $\boldsymbol{\Delta}^{\pi}$ may alter the Condorcet winner row, so in general $H_{\mathrm{cw}}(\boldsymbol{\Delta}^{\pi}) \neq H_{\mathrm{cw}}(\boldsymbol{\Delta})$, and the construction can even drastically increase it: $H_{\mathrm{cw}}(\boldsymbol{\Delta}^{\pi}) \gg H_{\mathrm{cw}}(\boldsymbol{\Delta})$.

In this section, we explain how the lower bound techniques from the proof of Theorem 3.2 can be adapted to also preserve the CW row.

To this end, define $\tilde{\Pi}(\boldsymbol{\Sigma}) \subset \Pi(\boldsymbol{\Sigma})$ as the subset of permutations preserving the CW row. For each $i \in \{2, \dots, K\}$, let $\tilde{\Pi}_i(\boldsymbol{\Sigma})$ be permutations of Σ_i^- that fix $i^* = 1$:

$$\tilde{\Pi}_i(\boldsymbol{\Sigma}) \coloneqq \left\{ \pi_i : \Sigma_i^- \to \Sigma_i^- \;\middle|\; \pi_i \text{ is a bijection and } \pi_i(1) = 1 \right\}, \tag{133}$$

and set $\tilde{\Pi}(\boldsymbol{\Sigma}) \coloneqq \tilde{\Pi}_2(\boldsymbol{\Sigma}) \times \cdots \times \tilde{\Pi}_K(\boldsymbol{\Sigma})$. For $\pi \in \tilde{\Pi}(\boldsymbol{\Sigma})$, construct $\boldsymbol{\Delta}^{(\pi)}$ via Equation (108). In addition to properties 1, 3, and 4 of Lemma F.1, we have:

Lemma F.3. *For any $\pi \in \tilde{\Pi}(\boldsymbol{\Sigma})$, $\boldsymbol{\Delta}^{(\pi)}$ satisfies:*

2' $\Delta^{(\pi)}_{1,\cdot} = \Delta_{1,\cdot}$ (CW row preservation).

We can then derive the following theorem, analogous to Theorem 3.2:

Theorem F.4. *Let A be a δ-correct algorithm over $\mathbb{D}_{\mathrm{cw}}$, with $\delta \leqslant 1/12$. Let $\boldsymbol{\Delta} \in \mathbb{D}_{\mathrm{cw}}$ have no ties. Define*

$$\tilde{\chi} \coloneqq \inf \left\{ x > 0 : \sup_{\pi \in \tilde{\Pi}(\boldsymbol{\Sigma})} \mathbb{P}_{\boldsymbol{\Delta}^{(\pi)},A}(N_\delta > x) \leqslant \delta \right\}. \tag{134}$$

Then,

$$\tilde{\chi} \geqslant \frac{1}{16 \log(4/3)} \max_{i \neq i^*} \left(\frac{1}{\Delta^2_{i^*,i}} \wedge \frac{K_{i;<0}}{\|\Delta_i^-\|^2} \right) \log\left(\frac{1}{6\delta}\right), \tag{135}$$

$$\tilde{\chi} \geqslant \frac{1}{128 \log(4/3)} \frac{1}{\log(2K)} \sum_{i \neq i^*} \left(\frac{1}{\Delta^2_{i^*,i}} \wedge \frac{K_{i;<0}}{\|\Delta_i^-\|^2} \right). \tag{136}$$

[7] If $\boldsymbol{\Delta}$ contains ties, we fix the convention $f_{\mathrm{tb}} \equiv 0$, i.e., we never permute zero entries, hence keeping them uninformative.

Similarly to Section 3, we define a subclass of $\mathbb{D}(\mathbf{\Delta})$ (see (8)) that additionally preserves the CW row $\Delta_{i^*,\cdot}$:

$$\mathbb{D}_0(\mathbf{\Delta}) = \{\tilde{\mathbf{\Delta}} \in \mathbb{D}(\mathbf{\Delta}) \text{ s.t. } (\tilde{\Delta}_{i^*,i})_{i\neq i^*} = (\Delta_{i^*,i})_{i\neq i^*}\} \ . \tag{137}$$

Corollary F.5. *Let A be a δ-correct algorithm over $\mathbb{D}_{\mathrm{cw}}$, with $\delta \leqslant 1/12$. Let $\mathbf{\Delta} \in \mathbb{D}_{\mathrm{cw}}$. Then there exists $\tilde{\mathbf{\Delta}} \in \mathbb{D}_0(\mathbf{\Delta})$ such that, with $\mathbb{P}_{\tilde{\mathbf{\Delta}},A}$-probability at least δ, the budget N_δ satisfies*

$$N_\delta \gtrsim \sum_{i\neq i^*} \frac{\log(1/\delta)}{\Delta^2_{i^*,i} \vee \Delta^2_{i,(s_i^*)}} + \max_{i\neq i^*} \frac{\log(1/\delta)}{\Delta^2_{i^*,i} \vee \frac{\|\Delta_i^-\|^2}{K_{i;<0}}} + \sum_{i\neq i^*} \frac{1}{\Delta^2_{i^*,i} \vee \frac{\|\Delta_i^-\|^2}{K_{i;<0}}}, \tag{138}$$

where $\gtrsim$ hides logarithmic K factors and numerical constants.

Proof. Assume without loss of generality that $i^* = 1$. Construct M by keeping the CW row and column unchanged and, for every $i, j \neq 1$ with $\Delta_{i,j} < 0$, setting

$$M_{i,j} = \max\{\Delta_{i,j}, \Delta_{i,(s_i^*)}\}, \qquad M_{j,i} = -M_{i,j},$$

while keeping zero entries equal to zero. For every $i \neq 1$ and $1 \leqslant r \leqslant K_{i;<0}$,

$$|M_{i,(r)}| \leqslant |\Delta_{i,(r)}|, \qquad M_{i,(s_i^*)} = \Delta_{i,(s_i^*)}.$$

Thus every admissible sparsity cost can only increase, with equality at $\boldsymbol{s}^*$. Hence $\boldsymbol{s}^*$ remains optimal and, with consistent tie-breaking, $M \in \mathbb{D}_0(\mathbf{\Delta})$.

Moreover, for every $i \neq 1$,

$$K_{i;<0}(M) = K_{i;<0}, \qquad \|M_i^-\|^2 \leqslant \|\Delta_i^-\|^2, \qquad M^2_{i,(1)} = \Delta^2_{1,i} \vee \Delta^2_{i,(s_i^*)}.$$

Apply the proof of Theorem F.4 to M, leaving zero entries fixed when ties occur. The first two relations yield a hard permutation $\tilde{\mathbf{\Delta}} \in \mathbb{D}_0(\mathbf{\Delta})$ satisfying the two exploration lower bounds. Since this permutation preserves the negative-entry multisets, the last relation and the quantile bound in Proposition 3.1 give the certification lower bound on the same instance. Taking the maximum of these thresholds yields their sum up to a numerical constant. □

Remarks. *This reveals the fundamental trade-off between eliminating suboptimal arms against the CW versus finding better competitors among them.*

Proof of Theorem F.4. Assume without loss of generality that $i^* = 1$. We start with the proof of Equation (135), which follows the proof of Equation (110). From careful inspection, Steps 1, 2, and 3 apply verbatim, replacing χ by $\tilde{\chi}$ and Π by $\tilde{\Pi}$.

Fix $k \neq i^*$. With the same notation and construction, one constructs $\tilde{A}$ with budget upper bounded by $\tilde{\chi}$ such that

$$\log\left(\frac{1}{6\delta}\right) \leqslant 8\log\left(\frac{4}{3}\right) \frac{1}{|\tilde{\Pi}_k(\mathbf{\Sigma})|} \sum_{\pi_k \in \tilde{\Pi}_k(\mathbf{\Sigma})} \sum_{j\in\Sigma_k^-} \mathbb{E}^{(\pi,k)}_{\tilde{A}}[N_{\{k,j\}}]\Delta^2_{k,\pi_k(j)}, \tag{139}$$

with the only difference in the subsequent computation.

If $K_{k;<0} = 1$, then $\Sigma_k^- = \{1\}$, so the preceding inequality gives

$$\log\left(\frac{1}{6\delta}\right) \leqslant 8\log(4/3)\Delta^2_{k,1}\widetilde{\chi}.$$

Since $\|\Delta_k^-\|^2 = \Delta^2_{k,1}$, this yields a stronger bound than required for this arm. For the remainder of this calculation, assume $K_{k;<0} \geqslant 2$.

For $\pi \in \tilde{\Pi}_k$, we have $\pi_k(1) = 1$ and $\pi_k|_{\Sigma_k^-\setminus\{1\}}$ is a bijection, so $|\tilde{\Pi}_k(\mathbf{\Sigma})| \simeq \mathfrak{S}_{k_i}$ where $k_i := |\Sigma_k^- \setminus \{1\}| = K_{k;<0} - 1$. Separating the role of 1 and the rest of Σ_k^- in (139),

$$\begin{aligned}
&\frac{1}{|\tilde{\Pi}_k(\mathbf{\Sigma})|} \sum_{\pi_k \in \tilde{\Pi}_k(\mathbf{\Sigma})} \sum_{j\in\Sigma_k^-} \mathbb{E}^{(\pi,k)}_{\tilde{A}}[N_{\{k,j\}}]\Delta^2_{k,\pi_k(j)} \\
&= \mathbb{E}^{(\pi,k)}_{\tilde{A}}[N_{\{k,1\}}]\Delta^2_{k,1} + \frac{1}{|\mathfrak{S}_{k_i}|} \sum_{\pi_k \in \tilde{\Pi}_k(\mathbf{\Sigma})} \sum_{j\in\Sigma_k^-\setminus\{1\}} \mathbb{E}^{(\pi,k)}_{\tilde{A}}[N_{\{k,j\}}]\Delta^2_{k,\pi_k(j)},
\end{aligned}$$

where $\mathbb{E}_{\tilde{A}}^{(\pi,k)}$ is independent of π_k. By Lemma H.3,

$$\frac{1}{8\log(4/3)}\log\left(\frac{1}{6\delta}\right) \leqslant \mathbb{E}_{\tilde{A}}^{(\pi,k)}[N_{\{k,1\}}]\Delta_{k,1}^2 + \frac{1}{K_{k;<0}-1}\sum_{j\in\Sigma_k^-\setminus\{1\}}\mathbb{E}_{\tilde{A}}^{(\pi,k)}[N_{\{k,j\}}]\sum_{j\in\Sigma_k^-\setminus\{1\}}\Delta_{k,j}^2$$
$$\leqslant \Delta_{k,1}^2\tilde{\chi} + \frac{\|\Delta_k^-\|^2-\Delta_{k,1}^2}{K_{k;<0}-1}\tilde{\chi},$$

using $\sum_{j\in\Sigma_k^-\setminus\{1\}}\mathbb{E}_{\tilde{A}}^{(\pi,k)}[N_{\{k,j\}}] \leqslant \tilde{\chi}$ and $\sum_{j\in\Sigma_k^-\setminus\{1\}}\Delta_{k,j}^2 = \|\Delta_k^-\|^2 - \Delta_{k,1}^2$. Finally,

$$\Delta_{k,1}^2 + \frac{\|\Delta_k^-\|^2-\Delta_{k,1}^2}{K_{k;<0}-1} \leqslant 2\left(\Delta_{k,1}^2 \vee \frac{\|\Delta_k^-\|^2}{K_{k;<0}}\right).$$

Taking the maximum over $k \neq i^*$ yields

$$\tilde{\chi} \geqslant \frac{1}{16\log(4/3)}\max_{k\neq i^*}\frac{1}{\Delta_{k,1}^2 \vee \frac{\|\Delta_k^-\|^2}{K_{k;<0}}}\log\left(\frac{1}{6\delta}\right),$$

which is Equation (135).

The proof of (136) follows the proof of (111) step-by-step, highlighting differences below.

Step 1: Define $\tilde{\beta}_k = \Delta_{k,1}^2 \vee \frac{\|\Delta_k^-\|^2}{K_{k;<0}}$ and $\tilde{I} := \arg\max_{k=2}^K \frac{k-1}{\tilde{\beta}_k}$.

Step 2: The same event (using $\tilde{I}$) bounds the probabilities, so (125) holds.

Step 3: For $K_{k;<0} \geqslant 2$, the KL upper bound computation adapts as above. If $K_{k;<0} = 1$, then $\widetilde{\beta}_k = \Delta_{k,1}^2$, and the direct KL calculation gives

$$\mathrm{KL}\big(\mathbb{P}_{\tilde{A}_k}^{(\pi,k)}, \mathbb{P}_{\tilde{A}_k}^{(\pi)}\big) \leqslant 8\log(4/3)\Delta_{k,1}^2\mathbb{E}_{\tilde{A}_k}^{(\pi,k)}[N_{\{k,1\}}] \leqslant 8\log(4/3)\frac{4\widetilde{\chi}}{\widetilde{I}-1}\widetilde{\beta}_k.$$

Combining both cases and averaging yields

$$\frac{1}{|\tilde{\Pi}|}\sum_{\pi\in\tilde{\Pi}}\frac{1}{I-1}\sum_{k=2}^{I}\mathrm{KL}\big(\mathbb{P}_{\tilde{A}_k}^{(\pi,k)}, \mathbb{P}_{\tilde{A}_k}^{(\pi)}\big) \leqslant 16\log\Big(\frac{4}{3}\Big)\cdot\frac{4\tilde{\chi}}{I-1}\cdot\tilde{\beta}_I. \tag{140}$$

which conclude from rearranging.

Step 4 is unnecessary since $\mathbf{\Sigma}$ avoids ties by convention.

□

G Proofs of Section E

In this section, we provide all lower bound proofs for the fixed-budget setting. As discussed at the end of Appendix E, these proofs closely parallel those in Appendix F, but the fixed-budget nature requires new arguments.

For completeness, we revisit all results with minimax-style formulations and provide a nearly self-contained presentation. In Subsection G.1, we prove Theorem E.1. Theorem E.2 follows in Subsection G.2.

Roadmap for Fixed-Budget Lower Bounds The proofs in this section follow the same three-step change-of-measure pattern introduced for the fixed-confidence case in Subsection F.1. However, the fixed-budget setting requires three important adaptations, which we now describe in detail. In the fixed-budget setting, we aim to lower bound the worst-case error

$$\inf_A \sup_{M\in\mathbb{D}} \mathbb{P}_{A,M}(\hat{i}_T \neq i^*),$$

where the infimum is over all algorithms A with fixed budget T, and $\mathbb{D}$ is some class of instances.

Step 1: Reference and alternative instances. In the fixed-confidence setting, the reference instance can be any arbitrary gap matrix $\mathbf{\Delta} \in \mathbb{D}_{\mathrm{cw}}$. Here, we instead construct a highly symmetric reference matrix $M \in \mathbb{D}$ that serves as a "hard instance" for the minimax bound. For each suboptimal arm $k \neq i^*$, we construct an alternative instance $M^{(k)}$ by setting all negative entries in row k (and corresponding column entries to preserve antisymmetry) to zero. This ensures $M^{(k)} \notin \mathbb{D}_{\mathrm{cw}}$.

Step 2: Separating event and total variation bound. Unlike fixed-confidence algorithms, which use a stopping time N_δ to construct events on which the two distributions disagree, fixed-budget algorithms have no stopping rule. Instead, we exploit the symmetry of the reference matrix M, together with the recommendation rule.

Step 3: KL decomposition. This step is conceptually identical to the fixed-confidence proofs. The KL divergence decomposes as

$$\mathrm{KL}(\mathbb{P}_M, \mathbb{P}_{M^{(k)}}) = \sum_{k<j} \mathbb{E}_M[N_{\{k,j\}}]\, \mathrm{kl}(\Delta_{k,j}, \Delta^{(k)}_{k,j}),$$

where $N_{\{k,j\}}$ counts unordered duels between arms k and j. The key difference is that the fixed budget T directly bounds $\sum_{k<j} \mathbb{E}_M[N_{\{k,j\}}] \leqslant T$, whereas fixed-confidence proofs require truncation at the $(1-\delta)$-quantile χ and reinterpretation. Combining the total variation lower bound from Step 2 with this KL upper bound from Step 3 yields the desired lower bound on ϵ_T.

G.1 Proof of Theorem E.1

Let $\underline{\Delta} = (\Delta_i)_{i\in[K]}$ be a K-dimensional vector such that $\Delta_1 = 0$ and $\Delta_i \in (0, 1/4)$ for $i \neq 1$, and assume without loss of generality that $\Delta_2 \leqslant \Delta_3 \leqslant \cdots \leqslant \Delta_K$. Recall that $\mathbb{D}^{(1)}(\underline{\Delta})$ (see Definition 83) denotes the class of gap matrices that admit a Condorcet winner whose row is equal to $\underline{\Delta}$ up to permutation.

Fix an algorithm A with a fixed budget $T \in \mathbb{N}^*$. Define the worst-case error probability of A over the class $\mathbb{D}^{(1)}(\underline{\Delta})$ as

$$\epsilon_T \coloneqq \max_{M\in\mathbb{D}^{(1)}(\underline{\Delta})} \mathbb{P}_{M,A}\left(\hat{i}_T \neq i^*(M)\right), \tag{141}$$

where $\hat{i}_T$ denotes the recommendation of algorithm A after T queries.

We aim to prove Theorem E.1, namely that

$$\epsilon_T \;\geq\; \frac{1}{4}\exp\left(-\frac{T}{\frac{1}{22}\sum_{i=2}^{K}\frac{1}{\Delta_i^2}}\right).$$

Sketch of proof. The proof follows the three-step roadmap of Subsection F.1. *(i) Reference and perturbed instances* $M^{(1)}, M^{(k)}$. Since we seek a minimax bound, we can construct a specific reference matrix $M^{(1)} \in \mathbb{D}^{(1)}(\underline{\Delta})$ with Condorcet winner $i^* = 1$, where arm 1 is the strongest opponent of every suboptimal arm. For each suboptimal $k \neq 1$, $M^{(k)}$ modifies row k's negative entries of $M^{(1)}$ to make k the CW while preserving $M^{(k)} \in \mathbb{D}^{(1)}(\underline{\Delta})$. *(ii) Separating event and TV bound.* Algorithm A is ϵ_T-correct over $\mathbb{D}^{(1)}(\underline{\Delta})$, so $\mathbb{P}_{M^{(k)}}(\hat{i}_T = k) \geq 1-\epsilon_T$. The event $B_k = \{\hat{i}_T \neq k\}$ thus satisfies $\mathbb{P}_{M^{(1)}}(B_k) \geq 1-\epsilon_T$ and $\mathbb{P}_{M^{(k)}}(B_k) \leq \epsilon_T$, yielding $\mathrm{TV}(\mathbb{P}_{M^{(1)}}, \mathbb{P}_{M^{(k)}}) \geq 1-2\epsilon_T$. *(iii) KL decomposition and computation.* Similar computation as in the proof of Theorem 3.1.

Proof of Theorem E.1. Step 1: construction of the reference instance $M^{(1)}$.

We construct within the class $\mathbb{D}^{(1)}(\underline{\Delta})$ a highly structured instance. The key structural property of this matrix is that the Condorcet winner $k^* = 1$ is the strongest opponent of every other arm.

Define a gap matrix $M^{(1)} \in \mathbb{D}_{\mathrm{cw}}$ by setting $M^{(1)}_{1,\cdot} = \underline{\Delta}$, and, for $i, j \neq 1$,

$$M^{(1)}_{i,j} \coloneqq \begin{cases} \Delta_j, & \text{if } i < j, \\ -\Delta_i, & \text{if } i > j, \\ 0, & \text{if } i = j. \end{cases}$$

Thus $M^{(1)}$ has the form

$$M^{(1)} = \begin{pmatrix} 0 & \Delta_2 & \Delta_3 & \Delta_4 & \cdots & \Delta_{K-1} & \Delta_K \\ -\Delta_2 & 0 & \Delta_3 & \Delta_4 & \cdots & \Delta_{K-1} & \Delta_K \\ -\Delta_3 & -\Delta_3 & 0 & \Delta_4 & \cdots & \Delta_{K-1} & \Delta_K \\ -\Delta_4 & -\Delta_4 & -\Delta_4 & 0 & \cdots & \Delta_{K-1} & \Delta_K \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -\Delta_{K-1} & -\Delta_{K-1} & -\Delta_{K-1} & -\Delta_{K-1} & \cdots & 0 & \Delta_K \\ -\Delta_K & -\Delta_K & -\Delta_K & -\Delta_K & \cdots & -\Delta_K & 0 \end{pmatrix}.$$

By construction, we have $M^{(1)} \in \mathbb{D}^{(1)}(\underline{\Delta})$, and its Condorcet winner is $i^*(M^{(1)}) = 1$.

For each $k \geqslant 2$, define the matrix $M^{(k)}$ as follows:

- For $i, j \neq k$, set $M^{(k)}_{i,j} = M^{(1)}_{i,j}$.
- For $j < k$, set $M^{(k)}_{k,j} = \Delta_{j+1}$ and $M^{(k)}_{j,k} = -\Delta_{j+1}$.
- For $j \geqslant k$, set $M^{(k)}_{k,j} = M^{(1)}_{k,j}$ and $M^{(k)}_{j,k} = M^{(1)}_{j,k}$.

The matrix $M^{(k)}$ can be written as

$$M^{(k)} = \begin{pmatrix} 0 & \Delta_2 & \cdots & \Delta_{k-1} & -\Delta_2 & \Delta_{k+1} & \cdots & \Delta_K \\ -\Delta_2 & 0 & \cdots & \Delta_{k-1} & -\Delta_3 & \Delta_{k+1} & \cdots & \Delta_K \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ -\Delta_{k-1} & -\Delta_{k-1} & \cdots & 0 & -\Delta_k & \Delta_{k+1} & \cdots & \Delta_K \\ \Delta_2 & \Delta_3 & \cdots & \Delta_k & 0 & \Delta_{k+1} & \cdots & \Delta_K \\ -\Delta_{k+1} & -\Delta_{k+1} & \cdots & -\Delta_{k+1} & -\Delta_{k+1} & 0 & \cdots & \Delta_K \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\ -\Delta_K & -\Delta_K & \cdots & -\Delta_K & -\Delta_K & -\Delta_K & \cdots & 0 \end{pmatrix},$$

where the blue entries indicate the differences with respect to $M^{(1)}$.

It is straightforward to check that, for each k, $M^{(k)} \in \mathbb{D}_{\mathrm{cw}}$. These matrices have three key properties: (i) $M^{(k)}$ does not have the same Condorcet winner as $M^{(1)}$, indeed $i^*(M^{(k)}) = k$; (ii) we have $M^{(k)} \in \mathbb{D}^{(1)}(\underline{\Delta})$, indeed the k-th row $M^{(k)}_{k,\cdot}$ is equal to $\underline{\Delta}$ up to a permutation; and (iii) the environment with gap matrix $M^{(k)}$ is difficult to distinguish from the one defined by $M^{(1)}$ in terms of KL divergence.

For $k \geqslant 2$, denote by $\mathbb{P}^{(k)}$ the distribution of the data when the underlying gap matrix is $M^{(k)}$.

Step 2: TV bound.

Let A be a δ-correct algorithm over $\mathbb{D}^{(1)}(\underline{\Delta})$, and let $\hat{i}$ denote its output. For any $k \geqslant 1$, when the true gap matrix is $M^{(k)}$ the Condorcet winner is k, and $M^{(k)} \in \mathbb{D}^{(1)}(\underline{\Delta})$. Then, the definition of ϵ_T (see (141)) implies

$$\forall k \in [K], \quad \mathbb{P}^{(k)}(\hat{i} \neq k) \leqslant \epsilon_T \,.$$

In particular, we have

$$1 - 2\epsilon_T \leqslant \mathbb{P}^{(1)}(\hat{i} \neq k) - \mathbb{P}^{(k)}(\hat{i} \neq k) \leqslant \mathrm{TV}\big(\mathbb{P}^{(1)}, \mathbb{P}^{(k)}\big) \,.$$

Then, with Bretagnolle–Huber inequality, we have

$$1 - 2\epsilon_T \leqslant \mathrm{TV}\big(\mathbb{P}^{(1)}, \mathbb{P}^{(k)}\big) \leqslant 1 - \tfrac{1}{2}\exp\big\{-\mathrm{KL}(\mathbb{P}^{(1)}, \mathbb{P}^{(k)})\big\}.$$

Step 3: computing the KL divergence and concluding.

For $i < j$ in $[K]$, let $N_{\{i,j\}}$ denote the total number of observed duels between i and j under algorithm A. Using the divergence decomposition lemma (Lemma 15.1 in Lattimore and Szepesvári, 2020), we have

$$\begin{aligned}\mathrm{KL}\big(\mathbb{P}^{(1)}, \mathbb{P}^{(k)}\big) &= \sum_{1 \leqslant i < j \leqslant K} \mathbb{E}^{(1)}\big[N_{\{i,j\}}\big]\ \mathrm{KL}\big(\mathbb{P}^{(1)}_{i,j}, \mathbb{P}^{(k)}_{i,j}\big) \\ &= \sum_{i=1}^{k-1} \mathbb{E}^{(1)}[N_{\{k,i\}}]\, \mathrm{KL}\big(\mathbb{P}^{(1)}_{k,i}, \mathbb{P}^{(k)}_{k,i}\big)\ , \end{aligned} \tag{142}$$

since the two instances differ only on pairs involving arm k.

For all $i < k$, the corresponding Bernoulli feedback distributions satisfy

$$\mathbb{P}^{(1)}_{k,i} = \mathcal{B}\big(\tfrac{1}{2} - \Delta_k\big)\,, \qquad \mathbb{P}^{(k)}_{k,i} = \mathcal{B}\big(\tfrac{1}{2} + \Delta_{i+1}\big)\,,$$

so that, with the same sequence of inequalities as in (93), (94), we obtain

$$\begin{aligned}\mathrm{KL}\big(\mathbb{P}^{(1)}_{k,i}, \mathbb{P}^{(k)}_{k,i}\big) &= \mathrm{kl}\big(-\Delta_k + \tfrac{1}{2},\ \Delta_{i+1} + \tfrac{1}{2}\big) \\ &\leqslant \frac{\big(-\Delta_k + \frac{1}{2} - (\Delta_{i+1} + \frac{1}{2})\big)^2}{(-\Delta_{i+1} + \frac{1}{2})(\Delta_{i+1} + \frac{1}{2})} \\ &\leqslant \frac{(2\Delta_k)^2}{3/16}\ \leq\ 22\,\Delta_k^2\,,\end{aligned}$$

where we used that $(\Delta_j)_{j \in \{2,\dots,K\}}$ is nondecreasing and $k \geqslant i+1$, which yields $\Delta_k \geqslant \Delta_{i+1}$. Also, we use that $\Delta_{i+1} \in (0, 1/4)$ and thus the denominator is bounded below by $3/16$.

Now, from (142) and the last inequality, we obtain

$$\mathrm{KL}\big(\mathbb{P}^{(1)}, \mathbb{P}^{(k)}\big) \leqslant 22 \sum_{i=1}^{k-1} \mathbb{E}^{(1)}\big[N_{\{k,i\}}\big]\, \Delta_k^2\,.$$

From Step 3, this bound implies

$$\sum_{i=1}^{k-1} \mathbb{E}^{(1)}\big[N_{\{k,i\}}\big] \geqslant \frac{1}{22\Delta_k^2}\ \log\frac{1}{4\epsilon_T}\ .$$

Summing over $k \geqslant 2$, and using that A has a budget T, we conclude that

$$T \geqslant \sum_{k=2}^{K} \sum_{i=1}^{k-1} \mathbb{E}^{(1)}\ \big[N_{\{k,i\}}\big] \geqslant \frac{1}{22} \sum_{k=2}^{K} \frac{1}{\Delta_k^2} \log\frac{1}{4\epsilon_T}\ ,$$

which, after rearranging, is exactly the claimed lower bound. □

G.2 Proof of Theorem E.2

Before proving Theorem E.2, we extend the notion of Condorcet winner to a broader class of matrices. Let $\mathbf{\Delta}$ be an antisymmetric matrix. We denote as a weak Condorcet winner, an arm $i^* \in [K]$ such that

$$\forall i \neq i^*, \Delta_{i^*,i} \geqslant 0 \text{ and, } \forall i \neq i^*, \min_{j \neq i} \Delta_{i,j} < 0\ . \tag{143}$$

Note that such arm is unique if it exists. Consider $\mathbb{D}_{\mathrm{wcw}}$ as the class of dueling bandit environments for which there exists such weak Condorcet winner

$$\mathbb{D}_{\mathrm{wcw}} := \Big\{\mathbf{\Delta} \in [-\tfrac{1}{4}, \tfrac{1}{4}]^{K \times K} : \exists!\, i^* \in [K] \text{ such that } \forall j,\ \Delta_{i^*,j} \geqslant 0\Big\}. \tag{144}$$

Observe that $\mathbb{D}_{\mathrm{cw}} \subset \mathbb{D}_{\mathrm{wcw}}$, and that naturally, any Condorcet winner is a weak Condorcet winner. Moreover, for a matrix $\mathbf{\Delta} \in \mathbb{D}_{\mathrm{wcw}}$, the quantities $(\boldsymbol{s}^*_{\mathbf{\Delta}}, \Delta_{(\boldsymbol{s}^*)})$ defined in Section 3 are still well

defined. Indeed, the minimum in Equation (4) for which $\boldsymbol{s}^*_{\boldsymbol{\Delta}}$ is defined as the argmin is still well defined under weak Condorcet winner assumption.

As a variant of the class $\mathbb{D}^{(3)}(\underline{\Delta}, \underline{s})$ introduced in (84), we introduce the class:

$$\mathbb{D}^{(3)}(\underline{\Delta}, \underline{s}) := \left\{ \boldsymbol{\Delta} \in \mathbb{D}_{\mathrm{wcw}} : \exists \sigma \in \mathfrak{S}_K \text{ s.t. } \boldsymbol{s}^*_{\boldsymbol{\Delta}} = \sigma(\underline{s}) \text{ and } \boldsymbol{\Delta}_{(s^*)} = -\sigma(\underline{\Delta}) \right\}. \tag{145}$$

The introduction of this definition is motivated by the fact that we want to consider in the lower bounds, matrices where the Condorcet winner may have some ties. The following lemma implies that this is feasible.

Lemma G.1. *Let A be an algorithm with a fixed budget T. Then, we have*

$$\sup_{\boldsymbol{\Delta} \in \mathbb{D}^{(3)}(\underline{\Delta}, \underline{s})} \mathbb{P}_{A, \boldsymbol{\Delta}}(\hat{i}_T \neq i^*(\boldsymbol{\Delta})) = \sup_{\boldsymbol{\Delta} \in \mathbb{D}^{(2)}(\underline{\Delta}, \underline{s})} \mathbb{P}_{A, \boldsymbol{\Delta}}(\hat{i}_T \neq i^*(\boldsymbol{\Delta}))$$

Proof of Lemma G.1. Let $\boldsymbol{\Delta} \in \mathbb{D}^{(3)}(\underline{\Delta}, \underline{s})$, with a weak Condorcet winner i^*. By definition, for any $i \neq i^*$, $\Delta_{i,\cdot}$ contains at least one negative entry. Let $\epsilon > 0$. Consider the modified matrix $\boldsymbol{\Delta}^\epsilon$ obtained from $\boldsymbol{\Delta}$ by lifting to $\epsilon > 0$ all off-diagonal null entries of on CW row i^* [and $-\epsilon$ to (j, i^*)], so that, for ϵ small enough, $\boldsymbol{\Delta}^\epsilon$ admits i^* as a (strong) Condorcet winner, and $\boldsymbol{\Delta}^\epsilon \in \mathbb{D}_{\mathrm{cw}}$. Now, for ϵ small enough, it holds that $\boldsymbol{s}^*_{\boldsymbol{\Delta}^\epsilon} = \boldsymbol{s}^*_{\boldsymbol{\Delta}}$ and $\Delta^\epsilon_{(s^*)} = \Delta_{(s^*)}$. Then,

$$\mathbb{P}_{A, \boldsymbol{\Delta}^\epsilon}(\hat{i}_T \neq i^*(\boldsymbol{\Delta})) \leqslant \sup_{\boldsymbol{\Delta} \in \mathbb{D}^{(2)}(\underline{\Delta}, \underline{s})} \mathbb{P}_{A, \boldsymbol{\Delta}}(\hat{i}_T \neq i^*(\boldsymbol{\Delta})).$$

Since A has a fixed budget T, one can take the limit $\epsilon \to 0$ in the inequality above. Taking a maximum over $\boldsymbol{\Delta} \in \mathbb{D}^{(3)}(\underline{\Delta}, \underline{s})$, one therefore obtains

$$\sup_{\boldsymbol{\Delta} \in \mathbb{D}^{(3)}(\underline{\Delta}, \underline{s})} \mathbb{P}_{A, \boldsymbol{\Delta}}(\hat{i}_T \neq i^*(\boldsymbol{\Delta})) \leqslant \sup_{\boldsymbol{\Delta} \in \mathbb{D}^{(2)}(\underline{\Delta}, \underline{s})} \mathbb{P}_{A, \boldsymbol{\Delta}}(\hat{i}_T \neq i^*(\boldsymbol{\Delta})) \,.$$

We have $\mathbb{D}^{(2)}(\underline{\Delta}, \underline{s}) \subset \mathbb{D}^{(3)}(\underline{\Delta}, \underline{s})$, so the other side of the inequality is clear. □

Now, we are ready to prove Theorem E.2. Assume that K is a multiple of 8, and denote $d = K/2$.

Let $(\underline{\Delta}, \underline{s})$ be such that $\underline{\Delta} = (\Delta_i)_{i \in [K]}$ with $\Delta_1 = 0$ and $(\Delta_2, \dots, \Delta_K) \in (0, 1/4)^{K-1}$. Let $\underline{s} = (s_1, s_2, \dots, s_K)$ with $s_1 = 0$, $1 \leqslant s_i \leqslant K/8$ for $i = 2, \dots, K$.

Consider the class $\mathbb{D}^{(3)}(\underline{\Delta}, \underline{s})$ as defined in Equation (145). Fix an algorithm A with a fixed budget T, and define the maximum error of A across $\mathbb{D}^{(3)}(\underline{\Delta}, \underline{s})$ as

$$\epsilon_T \coloneqq \sup_{\boldsymbol{\Delta} \in \mathbb{D}^{(3)}(\underline{\Delta}, \underline{s})} \mathbb{P}_{A, \boldsymbol{\Delta}}(\hat{i}_T \neq i^*(\boldsymbol{\Delta})) \,. \tag{146}$$

The proof of Theorem E.2 is divided in three lemmas, corresponding to the three terms in the lower bound.

Lemma G.2. *We have*

$$\epsilon_T \geqslant \frac{1}{4} \exp\left(-16 \log(4/3) \, \frac{T}{\max_{i=2}^d \frac{K}{s_i \Delta_i^2}} \right) \,. \tag{147}$$

Lemma G.3. *If $(\underline{\Delta}, \underline{s})$ are constants on the indices $i \in \{2, \dots, d\}$, that is, $\exists (\mu, s)$ such that, for any $i \in \{2, \dots, d\}$, $\Delta_i = \mu$, and $s_i = s$, then*

$$\epsilon_T \geqslant \frac{1}{2} - \sqrt{128 \log(4/3) \frac{T}{\frac{K^2}{s\mu^2}}} \,. \tag{148}$$

Lemma G.4. *With the same assumption as Lemma G.3, then*

$$\epsilon_T \geqslant \frac{1}{4} \exp\left(-32 \log(4/3) \, \frac{T}{\frac{K}{\mu^2}} \right) \,. \tag{149}$$

Proof of Theorem E.2. Recall that ϵ_T (see (146)) is the worst-case error over $\mathbb{D}^{(3)}(\underline{\Delta}, \underline{s})$. From Lemma G.1, it is also equal to the worst-case over $\mathbb{D}^{(2)}(\underline{\Delta}, \underline{s})$. Together, Lemmas G.2, G.3, and G.4 directly imply Theorem E.2. □

Proof of Lemma G.2. Recalling the definition of ϵ_T from (146), $\epsilon_T = \sup_{\boldsymbol{\Delta} \in \mathbb{D}^{(3)}(\underline{\Delta}, \underline{s})} \mathbb{P}_{A, \boldsymbol{\Delta}}(\hat{i}_T \neq i^*(\boldsymbol{\Delta}))$, we want to prove the following bound, equivalent to (147)

$$T \geqslant \frac{1}{16 \log(4/3)} \max_{i=2}^{d} \frac{K}{s_i \Delta_i^2} \log\left(\frac{1}{4\epsilon_T}\right) \ .$$

We may assume that $\epsilon_T < 1/4$, since otherwise the bound is immediate.

Step 1: reference matrix M^π. For reference, we consider the same matrix as in Subsection F.5, taking $\epsilon = 0$. For completeness, we recall this construction here.

We assumed for simplicity that K is a multiple of 8, and denote $d = K/2$. Consider the $K \times K$ antisymmetric matrix M defined by

$$M = \begin{pmatrix} \mathbf{0} & -D \\ D^\top & \Lambda \end{pmatrix} \ , \tag{150}$$

where D and Λ are two $d \times d$ matrices specified below.

The matrix D is the $d \times d$ matrix with nonnegative entries such that the first row is $D_{1,\cdot} = (0, \dots, 0) \in \mathbb{R}^d$, and for any $i = 2, \dots, d$,

$$D_{i,\cdot} = (\underbrace{\Delta_i, \dots, \Delta_i}_{s_i \text{ times}}, 0, \dots, 0) \in \mathbb{R}^d \ ,$$

which is possible since $s_i \leqslant d$ for $i = 2, \dots, d$.

To construct Λ, recall that d is assumed to be a multiple of 4 and that we assumed $s_i \in \{1, \dots, d/4\}$ for all $i \in \{d+1, \dots, K\}$. Define Λ as the following block matrix:

$$\Lambda = \begin{pmatrix} 0 & -\Lambda^{(0)} & 0 & (\Lambda^{(3)})^\top \\ (\Lambda^{(0)})^\top & 0 & -\Lambda^{(1)} & 0 \\ 0 & (\Lambda^{(1)})^\top & 0 & -\Lambda^{(2)} \\ -\Lambda^{(3)} & 0 & (\Lambda^{(2)})^\top & 0 \end{pmatrix} ,$$

where, for $l \in \{0, \dots, 3\}$, the sub-matrix $\Lambda^{(l)}$ is the $d/4 \times d/4$ matrix such that, for $i \in \{1, \dots, d/4\}$, the i-th row of $\Lambda^{(l)}$ is

$$\Lambda^{(l)}_{i,\cdot} = (\underbrace{\Delta_j, \dots, \Delta_j}_{s_j \text{ times}}, 0, \dots, 0) \in \mathbb{R}^{d/4} \quad \text{with } j = d + \frac{d}{4} l + i \ . \tag{151}$$

Overall, M is clearly antisymmetric by construction. Moreover, for each arm $i = 2, \dots, K$, the i-th row of M contains exactly s_i negative entries of magnitude Δ_i. The first row is equal to 0, so that $M \in \mathbb{D}_{\text{wcw}}$ (see (143)) and the (weak) Condorcet winner is $i^* = 1$. Finally, since in each row the negative entries are constant, we have $s^*_M = (s_1, \dots, s_K)$ and $M \in \mathbb{D}^{(3)}(\underline{\Delta}, \underline{s})$.

We use the same permutation construction as in the proof of Theorem 3.2. We recall this construction here. Let Π be the set of permutations, where $\pi = (\pi_1, \dots, \pi_d) \in \Pi$ if π_i is a permutation of $\{1, \dots, d\}$ for any $i \in [d]$.

From any $\pi \in \Pi$, define M^π as the matrix obtained by permuting the d first columns of D according to π in the following way:

$$M^\pi = \begin{pmatrix} \mathbf{0} & -D^\pi \\ (D^\pi)^\top & \Lambda \end{pmatrix} \ , \tag{152}$$

where, for any $(i, j) \in [d]^2$,

$$D^\pi_{i,j} = D_{i, \pi_i(j)} \ .$$

By construction, for any $\pi \in \Pi$, we still have $M^\pi \in \mathbb{D}^{(3)}(\underline{\Delta}, \underline{s})$. We denote by $\mathbb{P}^\pi_A$ the distribution of the observations induced by the interaction between algorithm A and the environment with gap matrix M^π.

Alternative instance $M^{(\pi,k)}$. Fix a suboptimal arm $k \in \{2, \dots, d\}$.

Construct the gap matrix $M^{(\pi,k)}$ by setting to zero all entries in the k-th row and the k-th column of M^{π}. By construction, rows 1 and k of $M^{(\pi,k)}$ only contain zero entries, so that $M^{(\pi,k)}$ does not admit a unique weak Condorcet winner. We denote by $\mathbb{P}_A^{(\pi,k)}$ the distribution of the observations induced by the interaction between algorithm A and the environment with gap matrix $M^{(\pi,k)}$.

Also define $\widetilde{M}^{\pi}$ as the matrix obtained by simultaneously exchanging rows 1 and k and columns 1 and k of M^{π}. Then $\widetilde{M}^{\pi} \in \mathbb{D}^{(3)}(\underline{\Delta}, \underline{s})$ and $i^*(\widetilde{M}^{\pi}) = k$. We denote its induced distribution by $\widetilde{\mathbb{P}}_A^{\pi}$. Since rows and columns 1 and k of $M^{(\pi,k)}$ are zero, $M^{(\pi,k)}$ is unchanged by this exchange.

Step 2: information-theoretic arguments.

Consider the recommendation rule $\hat{i}$ and the budget T of algorithm A. By definition of ϵ_T, A can be considered as ϵ_T-correct over $\mathbb{D}^{(3)}(\underline{\Delta}, \underline{s})$.

Consider the event

$$B \coloneqq \{\hat{i} \neq 1\} \, .$$

We have

$$\frac{1}{|\Pi|} \sum_{\pi \in \Pi} \left[\mathbb{P}_A^{(\pi,k)}(B) + \mathbb{P}_A^{(\pi,k)}(B^c) \right] = 1 \, . \tag{153}$$

For any π, we have $i^*(M^{\pi}) = 1$ and $i^*(\widetilde{M}^{\pi}) = k$. We can then use the fact that A is ϵ_T-correct over this class, by definition of ϵ_T ((146)), to get

$$\mathbb{P}_A^{\pi}(B) \leqslant \epsilon_T \quad \text{and} \quad \widetilde{\mathbb{P}}_A^{\pi}(B^c) = \widetilde{\mathbb{P}}_A^{\pi}(\hat{i} = 1) \leqslant \widetilde{\mathbb{P}}_A^{\pi}(\hat{i} \neq k) \leqslant \epsilon_T \ . \tag{154}$$

Now we use Lemma H.4, a Fano-type inequality presented as Proposition 4 in Gerchinovitz et al. [2020], to obtain

$$\mathbb{P}_A^{(\pi,k)}(B) \leqslant \frac{\mathrm{KL}\big(\mathbb{P}_A^{(\pi,k)}, \mathbb{P}_A^{\pi}\big) + \log(2)}{\log(1/\epsilon_T)} \quad \text{and} \quad \mathbb{P}_A^{(\pi,k)}(B^c) \leqslant \frac{\mathrm{KL}\big(\mathbb{P}_A^{(\pi,k)}, \widetilde{\mathbb{P}}_A^{\pi}\big) + \log(2)}{\log(1/\epsilon_T)} \, .$$

Adding these inequalities, averaging over $\pi \in \Pi$, and using (153), we get

$$1 \leqslant \frac{\frac{1}{|\Pi|} \sum_{\pi \in \Pi} \left[\mathrm{KL}\big(\mathbb{P}_A^{(\pi,k)}, \mathbb{P}_A^{\pi}\big) + \mathrm{KL}\big(\mathbb{P}_A^{(\pi,k)}, \widetilde{\mathbb{P}}_A^{\pi}\big) \right] + \log(4)}{\log(1/\epsilon_T)} \, .$$

Observe that $M^{(\pi,k)}$ does not depend on the permutation π_k. Denote as $\pi^{(-k)}$ the vector of permutations obtained from $\pi = (\pi_1, \dots, \pi_d) \in \Pi$ by removing the k-th component, that is, $\pi^{(-k)} = (\pi_1, \dots, \pi_{k-1}, \pi_{k+1}, \dots, \pi_d)$. Denote as $\Pi^{(-k)}$ the family $\{\pi^{(-k)}\}_{\pi \in \Pi}$. For a fixed $\pi^{(-k)} \in \Pi^{(-k)}$, we have $M^{(\pi,k)} = M^{(\pi^{(-k)},k)}$. Then, we write the inequality above as

$$\frac{1}{|\Pi^{(-k)}|} \sum_{\pi^{(-k)} \in \Pi^{(-k)}} \frac{1}{|\mathfrak{S}_d|} \sum_{\pi'_k \in \mathfrak{S}_d} \left[\mathrm{KL}\big(\mathbb{P}_A^{(\pi^{(-k)},k)}, \mathbb{P}_A^{\pi'}\big) + \mathrm{KL}\big(\mathbb{P}_A^{(\pi^{(-k)},k)}, \widetilde{\mathbb{P}}_A^{\pi'}\big) \right] + \log(4) \geqslant \log(1/\epsilon_T) \, , \tag{155}$$

where inside the sum, $\pi' = (\pi_1, \dots, \pi_{k-1}, \pi'_k, \pi_{k+1}, \dots, \pi_d)$.

Step 3: computing the KL divergence. We now bound, for a fixed $\pi^{(-k)} \in \Pi^{(-k)}$,

$$\frac{1}{|\mathfrak{S}_d|} \sum_{\pi'_k \in \mathfrak{S}_d} \left[\mathrm{KL}\big(\mathbb{P}_A^{(\pi^{(-k)},k)}, \mathbb{P}_A^{\pi'}\big) + \mathrm{KL}\big(\mathbb{P}_A^{(\pi^{(-k)},k)}, \widetilde{\mathbb{P}}_A^{\pi'}\big) \right] .$$

The matrices $M^{\pi'}$ and $M^{(\pi^{(-k)},k)}$ differ only on pairs $\{k, j\}$ with $j \in \{d+1, \dots, K\}$, whereas $\widetilde{M}^{\pi'}$ and $M^{(\pi^{(-k)},k)}$ differ only on pairs $\{1, j\}$ with $j \in \{d+1, \dots, K\}$. Using the same computation as

in Equation (117), and the fact that the law under $M^{(\pi^{(-k)},k)}$ does not depend on π'_k, we obtain

$$\begin{aligned}
&\frac{1}{|\mathfrak{S}_d|}\sum_{\pi'_k\in\mathfrak{S}_d}\Big[\mathrm{KL}\big(\mathbb{P}_A^{(\pi^{(-k)},k)},\mathbb{P}_A^{\pi'}\big)+\mathrm{KL}\big(\mathbb{P}_A^{(\pi^{(-k)},k)},\widetilde{\mathbb{P}}_A^{\pi'}\big)\Big]\\
&\qquad\leqslant 8\log\left(\frac{4}{3}\right)\frac{\|M_{k,\cdot}\|^2}{d}\sum_{j=d+1}^{K}\mathbb{E}_A^{(\pi^{(-k)},k)}\left[N_{\{k,j\}}+N_{\{1,j\}}\right]\\
&\qquad\leqslant 8\log\left(\frac{4}{3}\right)\frac{\|M_{k,\cdot}\|^2}{d}\,T\ ,
\end{aligned}\tag{156}$$

where $N_{\{i,j\}}$ denotes the number of duels of the unordered pair $\{i,j\}$. The last inequality holds because the two families of pairs are disjoint. By construction of M, we have $\|M_{k,\cdot}\|^2=s_k\Delta_k^2$.

Averaging (156) over $\Pi^{(-k)}$ and combining with (155), we get

$$T\geqslant\frac{1}{8\log(4/3)}\frac{d}{s_k\Delta_k^2}\log\left(\frac{1}{4\epsilon_T}\right)=\frac{1}{16\log(4/3)}\frac{K}{s_k\Delta_k^2}\log\left(\frac{1}{4\epsilon_T}\right)\ .$$

This holds for every $k\in\{2,\dots,d\}$. Taking the maximum over k gives the desired bound (147). □

Proof of Lemma G.3. Assume additionally that there exist $\mu>0$ and $s\in[d]$ such that, for every $i\in\{2,\dots,d\}$, $\Delta_i=\mu$ and $s_i=s$. We want to prove Bound (148), that is

$$\epsilon_T\geqslant\frac{1}{2}-\sqrt{128\log(4/3)\frac{s\mu^2}{K^2}T}\ .$$

Step 1: reference and alternative instances.

Consider $\bar{M}$ as the matrix defined by

$$\bar{M}=\begin{pmatrix}\mathbf{0} & -\bar{D}\\ \bar{D}^\top & \Lambda\end{pmatrix}\ ,\tag{157}$$

where Λ is as in (151), and $\bar{D}$ is the $d\times d$ matrix such that, for any $i=1,\dots,d$,

$$\bar{D}_{i,\cdot}=(\underbrace{\mu,\dots,\mu}_{s\text{ times}},0,\dots,0)\in\mathbb{R}^d\ .$$

Observe that the first d rows of $\bar{M}$ are equal, and that $\bar{M}$ does not admit a Condorcet winner; in particular, $\bar{M}\notin\mathbb{D}^{(3)}(\underline{\Delta},\underline{s})$.

Again, for any $\pi\in\Pi$, define $\bar{M}^\pi$ as the matrix obtained by permuting the d first rows of $\bar{M}$ according to $\pi_1,\dots,\pi_d$ as in (152). For this part of the proof, we denote by $\mathbb{P}_A^\pi$ the distribution of the observations induced by the interaction between algorithm A and the environment with gap matrix $\bar{M}^\pi$.

Construction of perturbed instances $\bar{M}^{(\pi,k)}$. Consider any arm $k\in[d]$.

We construct $\bar{M}^{(\pi,k)}$ as the matrix obtained from $\bar{M}^\pi$ by setting to zero all entries in the k-th row and the k-th column. By construction, $\bar{M}^{(\pi,k)}\in\mathbb{D}^{(3)}(\underline{\Delta},\underline{s})$ and $i^*(\bar{M}^{(\pi,k)})=k$. We denote by $\mathbb{P}_A^{(\pi,k)}$ the distribution of the observations induced by the interaction between algorithm A and the environment with gap matrix $\bar{M}^{(\pi,k)}$.

Step 2: bound on the total variation distance.

Consider the recommendation rule $\hat{i}$ and the budget T of algorithm A, which is ϵ_T-correct over $\mathbb{D}^{(3)}(\underline{\Delta},\underline{s})$, by definition of the maximum error ϵ_T.

Intuitively, under $\bar{M}$ there is no Condorcet winner among the first d arms, so algorithm A cannot systematically decide in favour of a specific subset of them, and it must make a large error on at least half of these arms. Indeed, it always holds that $\{\hat{i}\notin[|1;d/2|]\}$ or $\{\hat{i}\notin[|d/2+1;d|]\}$. Therefore,

$$\frac{1}{|\Pi|}\sum_{\pi\in\Pi}\mathbb{P}_A^\pi(\hat{i}\notin[1;d/2])\geqslant\frac{1}{2}\quad\text{or}\quad\frac{1}{|\Pi|}\sum_{\pi\in\Pi}\mathbb{P}_A^\pi(\hat{i}\notin[d/2+1;d])\geqslant\frac{1}{2}.$$

Without loss of generality[8] we assume that

$$\frac{1}{|\Pi|}\sum_{\pi\in\Pi}\mathbb{P}_A^\pi(\hat{i}\notin[1;d/2])\geqslant\frac{1}{2}\,. \tag{158}$$

For any $k\in[d/2]$, consider the event

$$B_k\coloneqq\big\{\hat{i}=k\big\}\cup\big\{N_{\{k,\cdot\}}>\tfrac{16T}{K}\big\}\,,$$

$N_{\{k,\cdot\}}$ denotes the number of duels involving arm k and an adversary in $[d+1;K]$ between time $t=1$ and time T, that is,

$$N_{\{k,\cdot\}}=\Big|\Big\{t\in[T]\ :\ \exists j\in[d+1;K]\text{ with }\{I_t,J_t\}=\{k,j\}\Big\}\Big|\ .$$

Observe first that, for any fixed π, $\mathbb{P}_A^\pi$ does not depend on k, so that

$$\frac{1}{d/2}\sum_{k=1}^{d/2}\mathbb{P}_A^\pi(\hat{i}=k)=\frac{1}{d/2}\,\mathbb{P}_A^\pi(\hat{i}\in[1;d/2])\,.$$

Averaging over $\pi\in\Pi$ and using (158), we obtain

$$\frac{1}{d/2}\sum_{k=1}^{d/2}\frac{1}{|\Pi|}\sum_{\pi\in\Pi}\mathbb{P}_A^\pi(\hat{i}=k)\leqslant\frac{1}{d}\,. \tag{159}$$

Now, by definition, the family $\big(N_{\{k,\cdot\}}\big)_{k\in[d/2]}$ counts duels with pairwise disjoint sets of arms, for time-steps between 1 and T, so that

$$\sum_{k=1}^{d/2}N_{\{k,\cdot\}}\leqslant T\ .$$

From this upper bound, a simple counting argument implies that at most a fraction $1/4$ of the arms in $[d/2]$ can satisfy $N_{\{k,\cdot\}}>\frac{16T}{K}=\frac{4T}{d/2}$. Hence,

$$\frac{1}{d/2}\sum_{k=1}^{d/2}\mathbb{1}_{\{N_{\{k,\cdot\}}>\frac{16T}{K}\}}\leqslant\frac{1}{4}\ .$$

Taking expectation with respect to the probability $\frac{1}{|\Pi|}\sum_{\pi\in\Pi}\mathbb{P}_A^\pi$, we obtain

$$\frac{1}{d/2}\sum_{k=1}^{d/2}\frac{1}{|\Pi|}\sum_{\pi\in\Pi}\mathbb{P}_A^\pi\Big(N_{\{k,\cdot\}}>\tfrac{16T}{K}\Big)\leqslant\frac{1}{4}\ . \tag{160}$$

Combining (159) and (160), and using $d\geqslant 4$, we get

$$\frac{1}{d/2}\sum_{k=1}^{d/2}\frac{1}{|\Pi|}\sum_{\pi\in\Pi}\mathbb{P}_A^\pi(B_k)\leqslant\frac{1}{4}+\frac{1}{d}\leqslant\frac{1}{2}\ . \tag{161}$$

Now, consider B_k under $\mathbb{P}_A^{(\pi,k)}$. Observe that $B_k^c\subset\{\hat{i}\neq k\}$. The environment $\bar{M}^{(\pi,k)}$ admits k as Condorcet winner and belongs to $\mathbb{D}^{(3)}(\underline{\Delta},\underline{s})$. Using that A is ϵ_T-correct over this class, by definition of ϵ_T, we obtain, for any $\pi\in\Pi$,

$$\mathbb{P}_A^{(\pi,k)}(B_k^c)\leqslant\mathbb{P}_A^{(\pi,k)}(\hat{i}\neq k)\leqslant\epsilon_T\,. \tag{162}$$

The event B_k can be computed by a procedure $\widetilde{A}_k$ using at most $16T/K$ duels between arm k and opponents in $\{d+1,\dots,K\}$. For $t=1,\dots,T$, let A choose its next duel. If this duel is between k and an opponent in $\{d+1,\dots,K\}$ and would make the count $N_{\{k,\cdot\}}$ exceed $16T/K$, stop before

[8] in the other case, we consider $k\in[d/2+1,d]$ and run the same arguments

querying that duel and return $\psi_k = 1$. Otherwise, perform the duel and continue. If the procedure has not stopped after T duels, compute $\hat{i}_T$ and output $\psi_k := \mathbf{1}_{\{\hat{i}_T = k\}}$.

By construction, the decision ψ_k produced by $\tilde{A}_k$ satisfies $\psi_k = \mathbb{1}_{B_k}$. Moreover, for any environment ν, we have $\mathbb{P}_{\tilde{A}_k,\nu}(B_k) = \mathbb{P}_{A,\nu}(B_k)$. From these observations, we deduce

$$\begin{aligned}
&\mathrm{TV}\left(\frac{1}{d/2}\sum_{k=1}^{d/2}\frac{1}{|\Pi|}\sum_{\pi\in\Pi}\mathbb{P}_{\tilde{A}_k}^{(\pi,k)},\ \frac{1}{d/2}\sum_{k=1}^{d/2}\frac{1}{|\Pi|}\sum_{\pi\in\Pi}\mathbb{P}_{\tilde{A}_k}^{\pi}\right)\\
&\quad\geqslant \frac{1}{d/2}\sum_{k=1}^{d/2}\frac{1}{|\Pi|}\sum_{\pi\in\Pi}\mathbb{P}_{\tilde{A}_k}^{(\pi,k)}(B_k) - \frac{1}{d/2}\sum_{k=1}^{d/2}\frac{1}{|\Pi|}\sum_{\pi\in\Pi}\mathbb{P}_{\tilde{A}_k}^{\pi}(B_k)\\
&\quad= \frac{1}{d/2}\sum_{k=1}^{d/2}\frac{1}{|\Pi|}\sum_{\pi\in\Pi}\mathbb{P}_{A}^{(\pi,k)}(B_k) - \frac{1}{d/2}\sum_{k=1}^{d/2}\frac{1}{|\Pi|}\sum_{\pi\in\Pi}\mathbb{P}_{A}^{\pi}(B_k)\,.
\end{aligned}$$

Using (161) and (162), we obtain

$$\mathrm{TV}\left(\frac{1}{d/2}\sum_{k=1}^{d/2}\frac{1}{|\Pi|}\sum_{\pi\in\Pi}\mathbb{P}_{\tilde{A}}^{(\pi,k)},\ \frac{1}{d/2}\sum_{k=1}^{d/2}\frac{1}{|\Pi|}\sum_{\pi\in\Pi}\mathbb{P}_{\tilde{A}}^{\pi}\right) \geqslant 1 - \epsilon_T - \frac{1}{2} = \frac{1}{2} - \epsilon_T \tag{163}$$

Finally, using the convexity of the total variation distance together with Pinsker's inequality and (163), we get

$$\begin{aligned}
\frac{1}{2} - \epsilon_T &\leqslant \mathrm{TV}\left(\frac{1}{d/2}\sum_{k=1}^{d/2}\frac{1}{|\Pi|}\sum_{\pi\in\Pi}\mathbb{P}_{\tilde{A}_k}^{(\pi,k)},\ \frac{1}{d/2}\sum_{k=1}^{d/2}\frac{1}{|\Pi|}\sum_{\pi\in\Pi}\mathbb{P}_{\tilde{A}_k}^{\pi}\right)\\
&\leqslant \sqrt{\frac{1}{2}\frac{1}{d/2}\sum_{k=1}^{d/2}\frac{1}{|\Pi|}\sum_{\pi\in\Pi}\mathrm{KL}\big(\mathbb{P}_{\tilde{A}_k}^{(\pi,k)},\mathbb{P}_{\tilde{A}_k}^{\pi}\big)}\,.
\end{aligned} \tag{164}$$

Step 3: computing the KL divergence.

An important property of procedure $\tilde{A}_k$ is that the budget spent on duels with arm k is upper bounded by $16T/K$; precisely, for any environment ν,

$$\mathbb{E}_{\tilde{A}_k,\nu}\left[\sum_{i=d+1}^{K} N_{\{k,i\}}\right] \leqslant \frac{16T}{K}\ . \tag{165}$$

We now upper bound

$$\frac{1}{|\Pi|}\sum_{\pi\in\Pi}\mathrm{KL}\big(\mathbb{P}_{\tilde{A}_k}^{(\pi,k)},\mathbb{P}_{\tilde{A}_k}^{\pi}\big).$$

As in previous proofs, we fix $\pi_1,\dots,\pi_{k-1},\pi_{k+1},\dots$ and we average over the π_k's. Hence, for any $k\in[d/2]$, we get

$$\begin{aligned}
\frac{1}{|\mathfrak{S}_d|}\sum_{\pi_k\in\mathfrak{S}_d}\mathrm{KL}\big(\mathbb{P}_{\tilde{A}_k}^{(\pi,k)},\mathbb{P}_{\tilde{A}_k}^{\pi}\big) &\leqslant 8\log\left(\frac{4}{3}\right)\frac{\|\bar{M}_{k,\cdot}\|^2}{d}\,\mathbb{E}_{\tilde{A}_k}^{(\pi,k)}\left[\sum_{i=d+1}^{K}N_{\{k,i\}}\right]\\
&\leqslant 8\log\left(\frac{4}{3}\right)\frac{s\mu^2}{d}\,\frac{16T}{K}\ ,
\end{aligned}$$

where we use $\|\bar{M}_{k,\cdot}\|^2 = s\mu^2$ and Equation (165).

Gathering Equation (164) with the bound above, and rearranging (using $d = K/2$), we obtain

$$\epsilon_T \geqslant \frac{1}{2} - \sqrt{128\log(4/3)\frac{s\mu^2}{K^2}T}\ ,$$

which is exactly the desired bound (148). □

Proof of Lemma G.4. Consider again the constant case where there exist $\mu > 0$ and $s \in [d]$ such that, for every $i \in \{2, \dots, d\}$, $\Delta_i = \mu$ and $s_i = s$. We want to prove the following bound, equivalent to (149):

$$T \geqslant \frac{1}{32 \log(4/3)} \frac{K}{\mu^2} \log\left(\frac{1}{4\epsilon_T}\right) .$$

We may assume that $\epsilon_T < 1/4$, since otherwise the bound is immediate.

Step 1. Take the matrix $\bar{M}$ defined in (157).

Fix for now $k \in \{1, \dots, d\}$. Consider $\bar{M}^{(k)}$, the matrix obtained from $\bar{M}$ by setting to zero the k-th row and the k-th column of $\bar{M}$.

Recall that $\bar{M}$ does not admit a Condorcet winner, while $\bar{M}^{(k)} \in \mathbb{D}^{(3)}(\underline{\Delta}, \underline{s})$ with $i^*(\bar{M}^{(k)}) = k$. We denote by $\mathbb{P}_A$ (resp. $\mathbb{P}_A^{(k)}$) the distribution of the observations induced by the interaction between algorithm A and the environment with gap matrix $\bar{M}$ (resp. $\bar{M}^{(k)}$).

Step 2: bound on the total variation distance.

For each $k \in [d]$, consider the event

$$B_k \coloneqq \{\hat{i} = k\} .$$

Since these events are pairwise disjoint, we have

$$\frac{1}{d} \sum_{k=1}^{d} \mathbb{P}_A(B_k) = \frac{1}{d} \mathbb{P}_A(\hat{i} \in [d]) \leqslant \frac{1}{d} \leqslant \frac{1}{2} . \tag{166}$$

Now, consider B_k under $\mathbb{P}_A^{(k)}$. By definition of the maximum error ϵ_T, A is ϵ_T-correct over $\mathbb{D}^{(3)}$, so that

$$\mathbb{P}_A^{(k)}(B_k^c) = \mathbb{P}_A^{(k)}(\hat{i} \neq k) \leqslant \epsilon_T . \tag{167}$$

By the Fano-type inequality from Lemma H.4, it holds that

$$\mathbb{P}_A(B_k^c) \leqslant \frac{\mathrm{KL}(\mathbb{P}_A, \mathbb{P}_A^{(k)}) + \log(2)}{-\log(\mathbb{P}_A^{(k)}(B_k^c))} .$$

Averaging over $k \in [d]$ and using (166) and (167), we obtain

$$\begin{aligned} \frac{1}{2} &\leqslant \frac{1}{d} \sum_{k=1}^{d} \mathbb{P}_A(B_k^c) \\ &\leqslant \frac{\frac{1}{d} \sum_{k=1}^{d} \mathrm{KL}(\mathbb{P}_A, \mathbb{P}_A^{(k)}) + \log(2)}{-\log(\epsilon_T)} . \end{aligned} \tag{168}$$

Step 3: computing the KL divergence.

We now upper bound $\mathrm{KL}\left(\mathbb{P}_A, \mathbb{P}_A^{(k)}\right)$.

From the decomposition of the KL divergence and the definition of $\bar{M}^{(k)}$, we have

$$\begin{aligned} \mathrm{KL}\left(\mathbb{P}_A, \mathbb{P}_A^{(k)}\right) &= \sum_{i=d+1}^{K} \mathbb{E}_A[N_{\{k,i\}}] \, \mathrm{kl}\left(\tfrac{1}{2} + \bar{M}_{k,i}, \tfrac{1}{2}\right) \\ &\leqslant \sum_{i=d+1}^{K} \mathbb{E}_A[N_{\{k,i\}}] \, 8 \log(4/3) \, \mu^2 , \end{aligned} \tag{169}$$

where we use the fact that the row $\bar{M}_{k,\cdot}$ only takes values in $\{0, -\mu\}$, together with

$$\mathrm{kl}\left(\tfrac{1}{2} - \mu, \tfrac{1}{2}\right) \leqslant -\frac{1}{2} \log(1 - 4\mu^2) \leqslant 8 \log(4/3) \, \mu^2 .$$

Combining (168) with (169) and rearranging, we obtain

$$\frac{1}{16\log(4/3)}\,\frac{d}{\mu^2}\,\log\!\left(\frac{1}{4\epsilon_T}\right) \leqslant \frac{1}{8\log(4/3)}\,\frac{1}{\mu^2}\sum_{k=1}^{d}\mathrm{KL}\left(\mathbb{P}_A,\mathbb{P}_A^{(k)}\right) \leqslant \sum_{k=1}^{d}\sum_{i=d+1}^{K}\mathbb{E}_A[N_{\{k,i\}}] \leqslant T\,,$$

where the last inequality holds because each comparison is counted at most once. Since $d = K/2$, this gives

$$T \geqslant \frac{1}{32\log(4/3)}\,\frac{K}{\mu^2}\,\log\!\left(\frac{1}{4\epsilon_T}\right)\,,$$

which is the desired bound (149). □

H Technical Results

H.1 Deterministic bounds

Lemma H.1 (Section 6.1 of Audibert et al. [2010])**.** *Let $x_1,\dots,x_K$ denote a decreasing sequence of positive numbers. We have:*

$$\max_{k\in\{1,\dots,K\}} kx_k^2 \leq \sum_{i=1}^{K} x_i^2 \leq \log(4K)\max_{k\in\{1,\dots,K\}} kx_k^2.$$

Lemma H.2. *Let $x_1,\dots x_n$ denote a sequence of positive numbers such that $x_1 \leq \cdots \leq x_n$. Then we have for any $p \in (0,1)$*

$$\frac{pn}{x_{(\lceil pn\rceil)}} \leq \sum_{i=1}^{n}\frac{1}{x_i}\,.$$

Proof. We have $\frac{1}{x_n} \leq \cdots \leq \frac{1}{x_1}$. Therefore

$$\frac{pn}{x_{\lceil pn\rceil}} \leq \frac{\lceil pn\rceil}{x_{\lceil pn\rceil}} \leq \sum_{i=1}^{\lceil pn\rceil}\frac{1}{x_i} \leq \sum_{i=1}^{n}\frac{1}{x_i}\,.$$

□

Lemma H.3. *Let $\mathfrak{S}_d$ denote the set of permutations of $\{1,\dots,d\}$. Let $a_1,\dots,a_d$ and $b_1,\dots,b_d$ be two sequences of numbers. We have*

$$\frac{1}{d!}\sum_{\sigma\in\mathfrak{S}_d}\sum_{i=1}^{d}a_i b_{\sigma(i)} = \frac{1}{d}\left(\sum_{i=1}^{d}a_i\right)\left(\sum_{i=1}^{d}b_i\right)\,.$$

Proof. The result is just a consequence of summation manipulation. We have

$$\sum_{\sigma}\sum_{i=1}^{d}a_i b_{\sigma(i)} = \sum_{i=1}^{d}a_i\sum_{\sigma}b_{\sigma(i)} = \sum_{i=1}^{d}a_i\frac{d!}{d}\sum_{i=1}^{d}b_i = (d-1)!\left(\sum_{i=1}^{d}a_i\right)\left(\sum_{i=1}^{d}b_i\right)\,.$$

□

Below we present a useful Fano-type inequality presented as Proposition 4 in Gerchinovitz et al. [2020]

Lemma H.4. *Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability distributions, and let A be an event such that $\mathbb{Q}(A) \in (0,1)$. We have*

$$\mathbb{P}(A) \leq \frac{KL(\mathbb{P},\mathbb{Q}) + \log(2)}{-\log(\mathbb{Q}(A))}\,.$$

More generally, for all probability pairs $\mathbb{P}_i, \mathbb{Q}_i$ and all events A_i, where $i \in \{1, \dots, N\}$, with $0 < \frac{1}{N}\sum_{i=1}^N \mathbb{Q}_i(A_i) < 1$, we have

$$\frac{1}{N}\sum_{i=1}^N \mathbb{P}_i(A_i) \le \frac{\frac{1}{N}\sum_{i=1}^N \boldsymbol{KL}\left(\mathbb{P}_i, \mathbb{Q}_i\right) + \log(2)}{-\log\left(\frac{1}{N}\sum_{i=1}^N \mathbb{Q}_i(A_i)\right)} .$$

Lemma H.5. *For any $x > 0$, we have*

$$0 < \frac{\frac{1}{2} - e^{-x}}{\log(1-e^{-x}) - \log(e^{-x})} < \frac{1}{2x} ,$$

where the middle expression is defined by continuity to equal $1/4$ at $x = \log 2$.

Proof. Let $y = e^{-x} \in (0,1)$ and define

$$h(y) := \log\Big(\frac{1-y}{y}\Big) = \log(1-y) - \log y, \qquad R(y) := \frac{\frac{1}{2} - y}{h(y)} \quad (y \neq \tfrac{1}{2}) .$$

Let us start with a proof of the positivity of the middle expression. The function h is strictly decreasing on $(0,1)$ and satisfies $h(\frac{1}{2}) = 0$, hence $\operatorname{sign}(h(y)) = \operatorname{sign}(\frac{1}{2} - y)$. Therefore $R(y) > 0$ for all $y \neq \frac{1}{2}$. At $y = \frac{1}{2}$, both numerator and denominator vanish, by l'Hôpital's rule,

$$\lim_{y \to 1/2} R(y) = \frac{-1}{h'(1/2)} = \frac{-1}{-\left(\frac{1}{1-y} + \frac{1}{y}\right)\big|_{y=1/2}} = \frac{1}{4} .$$

Thus the middle expression is well-defined by continuity and is strictly positive for all $x > 0$.

Now let us prove the stated upper bound. Since $x = \log(1/y) > 0$, we need to show

$$R(y) < \frac{1}{2\log(1/y)} .$$

If $y < \frac{1}{2}$ (so $h(y) > 0$) we need to prove that

$$2\log(1/y)\Big(\tfrac{1}{2} - y\Big) < h(y) .$$

which is equivalent to $2y\log(1/y) + \log(1-y) > 0$, define $g(y) := 2y\log(1/y) + \log(1-y)$, therefore we need to show that

$$\forall y \in (0, 1/2), \quad g(y) > 0 .$$

If $y > \frac{1}{2}$ (so $h(y) < 0$), the same manipulation yields the equivalent condition $g(y) < 0$. Hence it suffices to prove $g(y) > 0$ on $(0, \frac{1}{2})$ and $g(y) < 0$ on $(\frac{1}{2}, 1)$.

A direct computation shows

$$g''(y) = -\frac{2}{y} - \frac{1}{(1-y)^2} < 0 \qquad (y \in (0,1)),$$

so g is strictly concave. Moreover,

$$\lim_{y \downarrow 0} g(y) = 0 \quad \text{and} \quad g(1/2) = 2 \cdot \frac{1}{2}\log 2 + \log(1/2) = 0.$$

By strict concavity, this implies $g(y) > 0$ for all $y \in (0, 1/2)$. Finally,

$$g'(y) = 2\log(1/y) - 2 - \frac{1}{1-y}, \qquad g'(1/2) = 2\log 2 - 4 < 0.$$

Since g is concave, g' is nonincreasing, so $g'(y) \le g'(1/2) < 0$ for all $y \ge 1/2$. Thus g is strictly decreasing on $[1/2, 1)$, and hence $g(y) < 0$ for all $y \in (1/2, 1)$.

Therefore $R(y) < 1/(2\log(1/y)) = 1/(2x)$ for all $x > 0$, concluding the proof. □

Lemma H.6. *Let d be an integer greater than 1. Let $M \in \mathbb{R}^{d\times d}$ such that M is skew symmetric (i.e. $\forall i,j \in [d] : M_{i,j} = -M_{j,i}$). Then the number of lines of M with at least $\lceil d/4 \rceil$ non-positive off-diagonal entries is at least $\lceil d/4 \rceil$.*

Proof. For $i \in [d]$ define

$$s_i := |\{j \in [d] : \ M_{i,j} \le 0\}| \ .$$

Since the matrix M is skew symmetric (i.e. $M_{i,j} = -M_{j,i}$ for any i, j), for every unordered pair $\{i, j\}$ where $i \neq j$, at least one of the two quantities $M_{i,j}$ or $M_{j,i}$ is nonpositive. Therefore, the number of off-diagonal non-positive entries is at least $\binom{d}{2}$. We conclude by taking into account the diagonal entries that

$$\sum_{i=1}^{d} s_i \ge \binom{d}{2} + d = \frac{d(d+1)}{2} \ .$$

The conclusion follows by a simple contradiction argument. □

Lemma H.7. *Let $n \ge 3$ and M be an $n \times n$ skew-symmetric matrix, let E denote the set of rows such that the number of non-positive entries is at least $\lceil n/4 \rceil + 1$, then the intersection of E with any subset of $\{1, \dots, n\}$ of cardinality $n - \lceil n/8 \rceil$ is non-empty.*

Proof. Assume $n \ge 3$ and set $a := \lceil n/8 \rceil$, $b := \lceil n/4 \rceil$. Suppose by contradiction that there exists $S \subseteq [n]$ with $|S| = n - a$ and $S \cap E = \varnothing$. Then each row $i \in S$ has at most b non-positive entries, hence at least $n - b$ positive entries. Let $P := |\{(i, j) : M_{ij} > 0\}|$ be the total number of positive entries. Summing over rows in S gives

$$P \ \ge \ |S|\,(n-b) = (n-a)(n-b).$$

On the other hand, skew-symmetry implies that for each unordered pair $\{i, j\}$ with $i \neq j$, at most one of M_{ij}, M_{ji} is positive, hence

$$P \ \le \ \binom{n}{2} = \frac{n(n-1)}{2}.$$

Thus $(n-a)(n-b) \le \frac{n(n-1)}{2}$. But for $n \ge 5$, using $\lceil x \rceil \le x + 1$,

$$n - a \ge \frac{7(n-1)}{8}, \qquad n - b \ge \frac{3(n-1)}{4} \quad \Rightarrow \quad (n-a)(n-b) \ge \frac{21}{32}(n-1)^2 > \frac{n(n-1)}{2},$$

a contradiction; and the remaining cases $n = 3, 4$ are checked directly: $(n-a)(n-b) = 4 > 3$ and $9 > 6$, respectively. Hence no such S exists, i.e. every S with $|S| = n - \lceil n/8 \rceil$ intersects E. □

Lemma H.8. *Let $x > 10^3$ and $y > 8$. Then*

$$2\,\frac{\log^2(x)\,\log^2(y)}{\log\big(xy(\log y)^5\big)} \ \ge \ \log(xy).$$

Proof. Since $x > 10^3$ and $y > 8$, we have $\log x > \log(10^3) > 6$ and $\log y > \log 8 > 2$. Hence

$$\log x\,\log y - (\log x + \log y) = (\log x - 1)(\log y - 1) - 1 > 0 \ ,$$

so

$$\log x\,\log y \ge \log(xy) \ . \tag{170}$$

Moreover,

$$\log\big(xy(\log y)^5\big) = \log(xy) + 5\log\log y.$$

Since $y > 8$ implies $\log y > 2 > 1$, we have $\log\log y \le \log y$, and thus

$$\log\big(xy(\log y)^5\big) \le \log(xy) + 5\log y = \log x + 6\log y.$$

Therefore,

$$\begin{aligned} 2\log x\,\log y - (\log x + 6\log y) &= (2\log y - 1)\log x - 6\log y \\ &\ge 6(2\log y - 1) - 6\log y \\ &= 6(\log y - 1) > 0 \ , \end{aligned}$$

so

$$2\log x\,\log y \;\geq\; \log\bigl(xy(\log y)^5\bigr). \tag{171}$$

Multiplying (170) and (171) yields

$$2\log^2(x)\,\log^2(y) \;\geq\; \log(xy)\,\log\bigl(xy(\log y)^5\bigr).$$

Since $\log\bigl(xy(\log y)^5\bigr) > 0$, dividing by it gives the claim. □

Lemma H.9. *Let $K \geq 2$, $H \geq 4$, and $T \geq 8K\log_{8/7}(K)$. Define*

$$p_k \coloneqq \frac{1}{18} \wedge (\log T + K)\exp\Bigl(-c\,\frac{T}{\log^4(K)\log(T)\,H}\Bigr).$$

Assume $T \geq c_0\,H\log^6(H)$ for some numerical constant c_0 large enough (depending only on c). Then there exists a numerical constant $c' > 0$ (e.g. $c' = c/2$) such that

$$p_k \leq \exp\Bigl(-c'\,\frac{T}{\log^4(K)\log(T)\,H}\Bigr).$$

Proof. Let us introduce the following notation

$$x \coloneqq \frac{T}{\log^4(K)\log(T)\,H}, \qquad A \coloneqq \log T + K\,.$$

Then $p_k \leq Ae^{-cx}$. Since $T \geq 8K\log_{8/7}(K)$ and $K \geq 2$, we have $A = \log T + K \leq T$, so $\log A \leq \log T$ and

$$p_k \leq \exp(-cx + \log T).$$

Thus it suffices to prove $\log T \leq \frac{c}{2}x$, i.e.

$$\frac{T}{(\log T)^2} \;\geq\; \frac{2H\log^4(K)}{c}. \tag{172}$$

Since $T \geq 8K\log_{8/7}(K) > K$, we have $\log^4(K) \leq (\log T)^4$. Therefore (172) follows from

$$\frac{T}{(\log T)^2} \;\geq\; \frac{2H}{c}(\log T)^4\,,$$

which is equivalent to

$$\frac{T}{(\log T)^6} \;\geq\; \frac{2H}{c}.$$

Let $g(t) \coloneqq t/(\log t)^6$, which is increasing for $t \geq e^6$. Choose c_0 large enough so that $T \geq c_0H\log^6(H) \geq e^6$ for all $H \geq 4$. Then, with $T_0 \coloneqq c_0H\log^6(H)$, we have $g(T) \geq g(T_0)$ and

$$g(T_0) = \frac{c_0H\log^6(H)}{\bigl(\log(c_0H\log^6(H))\bigr)^6}.$$

For $H \geq 4$, we have $\log\log H \leq \log H$, so

$$\begin{aligned}\log(c_0H\log^6(H)) &= \log c_0 + \log H + 6\log\log H \leq \log c_0 + 7\log H\\ &\leq \Bigl(7 + \frac{\log c_0}{\log 4}\Bigr)\log H =: C_0\log H\,.\end{aligned}$$

Hence $g(T_0) \geq \frac{c_0}{C_0^6}H$. Taking c_0 large enough so that $\frac{c_0}{C_0^6} \geq \frac{2}{c}$ yields $g(T) \geq g(T_0) \geq \frac{2H}{c}$, proving (172). Therefore $p_k \leq \exp(-(c/2)x)$, i.e. the claim with $c' = c/2$. □

H.2 Concentration inequalities

Below is Hoeffding concentration inequality.

Lemma H.10. *Let $X_1, \dots, X_n$ be independent random variables such that $a_i \leq X_i \leq b_i$ almost surely. Let $S_n = \sum_{i=1}^n X_i$. Then we have for all $t > 0$:*

$$\mathbb{P}\left(S_n - \mathbb{E}[S_n] \leq -t\right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

The right-hand side is defined to be 0 when $\sum_{i=1}^n (b_i - a_i)^2 = 0$.

Below we restate two results on the concentration of the sum of independent binary random variable from Buldygin and Moskvichova [2013]. First, let us introduce some notation. Let ξ denote a sub-Gaussian random variable, its sub-Gaussian standard is defined by:

$$\tau(\xi) := \inf\left\{a \geq 0 : \mathbb{E}\left[\exp\left(\lambda\xi\right)\right] \leq \exp\left(\frac{a^2\lambda^2}{2}\right), \lambda \in \mathbb{R}\right\}.$$

Lemma H.11 (Theorem 2.1 in Buldygin and Moskvichova [2013])**.** *Let X denote a Bernoulli random variable with parameter $p \in [0, 1]$. Then we have*

$$\tau^2\left(X - p\right) = \phi(p),$$

where $\phi(.)$ is the function defined by:

$$\phi(p) = \begin{cases} 0, & p \in \{0, 1\}; \\ \frac{1}{4}, & p = \frac{1}{2}; \\ \frac{\frac{1}{2} - p}{\log(1-p) - \log(p)}, & p \in (0, 1) \setminus \left\{\frac{1}{2}\right\}. \end{cases}$$

The lemma below gives a concentration bound on the binomial random variables.

Lemma H.12. *Let X_j for $j \in \{1, \dots, n\}$ denote a sequence of independent Bernoulli random variables with parameter $p \in [0, 1]$. Define $S_n = \sum_{j=1}^n (X_j - p)$. Then, we have for all $x > 0$*

$$\mathbb{P}\left(S_n \geq x\right) \leq \exp\left(\frac{-x^2}{2n\phi(p)}\right),$$

where ϕ is defined in Lemma H.11. We also have for all $x > 0$

$$\mathbb{P}\left(S_n \leq -x\right) \leq \exp\left(\frac{-x^2}{2n\phi(p)}\right).$$

For $p \in \{0, 1\}$, both right-hand sides are defined to be 0.

Proof. This is a direct consequence of Chernoff's bound with Lemma H.11. □

I Additional experiments

We report two complementary experiments. The first is a Bradley–Terry difficulty sweep, which tests a classical structured regime with a total order. The second uses unstructured random sparse matrices, to evaluate robustness beyond the controlled circulant model used in the main text.

For the Bradley–Terry experiment, arm 0 is the Condorcet winner. We assign latent scores $\theta_i = K - i$ for $i = 0, \dots, K - 1$ and set

$$q_{i,j} = \frac{1}{1 + \exp(-\lambda_{\mathrm{BT}}(\theta_i - \theta_j))}, \qquad \Delta_{i,j} = q_{i,j} - \frac{1}{2}.$$

The parameter λ_{BT} controls the difficulty: smaller values produce smaller gaps and therefore harder instances. We fix $K = 200$, $\delta = 0.1$, and run 5 trials for

$$\lambda_{\mathrm{BT}} \in \{0.04, 0.05, 0.06, 0.07, 0.08\},$$

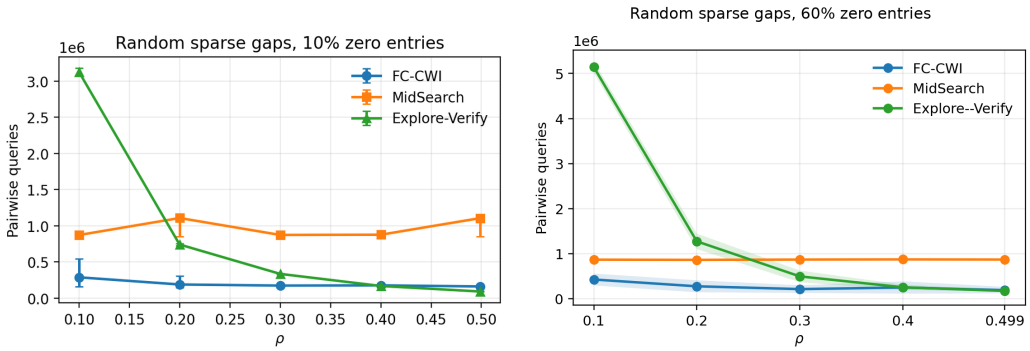


Figure 3: Query complexity on random sparse gap matrices. Left: 10% zero entries. Right: 60% zero entries. FC-CWI is most effective for small and intermediate ρ, while Explore–Verify improves as stronger random witnesses become easier to find and certify.

with a cap of 7×10^6 pairwise queries.

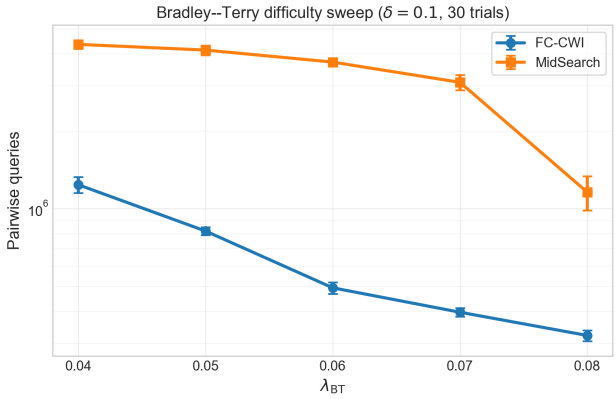


Figure 2: Query complexity on Bradley–Terry instances as the difficulty parameter λ_{BT} varies. FC-CWI remains consistently more sample-efficient than MidSearch across the sweep.

Figure 2 shows that FC-CWI identifies the Condorcet winner in all runs and uses substantially fewer queries than MidSearch. The gap is most pronounced at intermediate difficulty; for instance, when $\lambda_{\mathrm{BT}} = 0.05$, FC-CWI uses about 8.4×10^5 queries on average, whereas MidSearch uses about 5.3×10^6 queries and reaches the query cap once.

We also evaluate random sparse gap matrices. Arm 0 is again the Condorcet winner, with

$$\Delta_{0,i} \sim \mathrm{Unif}(0.05, 0.1), \qquad \Delta_{i,0} = -\Delta_{0,i}.$$

For each pair of suboptimal arms, the entry is set to zero with probability p_0 and otherwise sampled from $\mathrm{Unif}[-\rho, \rho]$, with skew-symmetry imposed. We use $K = 50$, $\delta = 0.1$, and average over 10 trials for $\rho \in \{0.1, 0.2, 0.3, 0.4, 0.499\}$.

Overall, these experiments show that FC-CWI remains competitive in classical structured instances and is particularly effective when useful suboptimal-suboptimal comparisons exist but are sparse or heterogeneous.

Calibration of numerical constants. All three algorithms contain numerical constants that affect their empirical sampling budgets. Since the theoretical constants are conservative, we calibrate one scalar constant per method on a single representative instance and then keep these constants fixed in all subsequent experiments. The calibration instance is the sparse controlled model with

$$K = 128, \qquad \delta = 0.1, \qquad \alpha = 0.2, \qquad \gamma = 0.05, \qquad \rho = 0.3.$$

Table 1: Selected calibration constants on the sparse controlled instance. All selected constants identify the Condorcet winner in all 10 paired trials.

Method	Selected constant	Errors / 10	Median queries	90% quantile
FC-CWI	0.5	$0/10$	1.95×10^6	2.14×10^6
MidSearch	0.3	$0/10$	8.42×10^6	8.75×10^6
Explore–Verify	4	$0/10$	4.00×10^6	4.03×10^6

Table 2: Robustness of the calibration. Query-complexity ranges are computed over local sweeps around the selected constants. All entries in the table have zero errors over 10 paired trials.

Method	Tested constants	Median query range	90% quantile range
FC-CWI	$\{0.5, 0.7, 0.9, 1.1, 1.3, 1.5\}$	$[1.95, 2.11]\times10^6$	$[2.14, 2.14]\times10^6$
MidSearch	$\{0.2, 0.3, 0.5\}$	$[6.09, 13.08]\times10^6$	$[6.44, 13.12]\times10^6$
Explore–Verify	$\{1, 2, 4\}$	$[4.00, 4.02]\times10^6$	$[4.03, 4.05]\times10^6$

For FC-CWI, the scalar c_{FC} is the constant used in the elimination/certification threshold. For MidSearch, the scalar c_{MS} rescales the sampling budgets used inside the MidSearch subroutines. For Explore–Verify, the scalar c_{EV} rescales the initial budget in the doubling schedule. In all cases, the confidence radii and verification thresholds are left unchanged.

To make the comparison insensitive to random fluctuations, we use paired trials: for a fixed method and trial index, all constants are evaluated on the same randomly permuted instance and with the same initial random seed. Each entry below is computed over 10 trials.

The stability table shows that the empirical conclusions are not driven by a finely tuned constant. FC-CWI remains essentially unchanged over the tested range of c_{FC}, while Explore–Verify is also stable near the selected constant. MidSearch is more sensitive because its scalar directly multiplies several internal sampling allocations, but the selected value $c_{\mathrm{MS}} = 0.3$ lies in a clean zero-error regime and the resulting query complexity remains well separated from FC-CWI on this calibration instance.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: All claims are proven rigorously.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We clearly discuss the optimality of our results, and present possible improvement as future reaseach topics in the conclusion.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: All theoretical statement are supported with proofs.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The code is presented in a seperate folder, the description of all the experiments are provided in the paper.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: Data is synthetic, its generation mechanism is presented in the paper, the code is provided.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Details about the experiments are provided in the Numerical simulations section and in the appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: All plots are supported with deviation shaded regions. Note that for some instances, the deviations are very small to appear in the plot.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)

- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The experiments are lightweight synthetic simulations; the paper specifies implementation details and query budgets, and reports that they were run on standard CPU hardware without specialized accelerators.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research conducted in the paper conforms in all aspects with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: This work is part of theoretical/foundational research, the authors are not aware of any particular concern that requires to be raised.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: Not applicable

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The authors are the only contributors.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.

- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: Not applicable.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: Not applicable.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: Not applicable

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: LLMs were not used as any important, original, or non-standard components.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.