Finding Optimal Arms in Non-stochastic Combinatorial Bandits with Semi-bandit Feedback and Finite Budget

Jasmin Brandt[a], **Viktor Bengs**[b], **Björn Haddenhorst**[a], **Eyke Hüllermeier**[b,c]
[a] Department of Computer Science, Paderborn University, Germany
[b]Institute of Informatics, University of Munich (LMU)
[c]Munich Center for Machine Learning
jasmin.brandt@upb.de, bjoernha@mail.upb.de, viktor.bengs@lmu.de, eyke@lmu.de

Abstract

We consider the combinatorial bandits problem with semi-bandit feedback under finite sampling budget constraints, in which the learner can carry out its action only for a limited number of times specified by an overall budget. The action is to choose a set of arms, whereupon feedback for each arm in the chosen set is received. Unlike existing works, we study this problem in a non-stochastic setting with subset-dependent feedback, i.e., the semi-bandit feedback received could be generated by an oblivious adversary and also might depend on the chosen set of arms. In addition, we consider a general feedback scenario covering both the numerical-based as well as preference-based case and introduce a sound theoretical framework for this setting guaranteeing sensible notions of optimal arms, which a learner seeks to find. We suggest a generic algorithm suitable to cover the full spectrum of conceivable arm elimination strategies from aggressive to conservative. Theoretical questions about the sufficient and necessary budget of the algorithm to find the best arm are answered and complemented by deriving lower bounds for any learning algorithm for this problem scenario.

1 Introduction

The multi-armed bandits (MAB) problem is an intensively studied problem class in the realm of machine learning, in which a learner is facing a sequential decision making problem under uncertainty [32]. A decision (action) corresponds to making a choice between a finite set of specific choice alternatives (objects, items, etc.), also called *arms* in reference to the metaphor of gambling machines in casinos. After each decision to choose a particular arm, the learner receives some form of *feedback* – typically a numerical reward – determined by a *feedback mechanism* of the chosen arm. The learner is not aware of the arms' feedback mechanisms and consequently tries to learn these in the course of time by performing actions according to its learning strategy. The concrete design of its learning strategy depends essentially on two main components of the learning setting: the assumptions on the feedback mechanisms and the learning task.

Traditionally, and even up to now the most prevalent assumption is that the feedback received by choosing one arm is generated by means of a probability distribution of the chosen arm [41, 31]. In this way, any useful learning strategy revolves around learning specific probabilistic features of the arms' distributions such as the means. These features, in turn, quite naturally provide a way to define a notion of *(sub-)optimality of an arm* as well as a *best arm*. A relaxation of this *stochastic setting* is the *non-stochastic setting*, in which no assumption is made in the form of probabilistic laws of the feedback mechanisms. Instead, either no assumptions are made on the feedback mechanisms, so that

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

these can also be generated by an adversary [7], or that the sequence of feedback observations (or a transformation thereof) of an arm converges asymptotically to a fixed point [25]. In the latter case, the notion of an arm's (sub-)optimality is again straightforward, given that the limit points can be ordered, while in the former usually the best action in hindsight plays the role of the best arm.

Regarding the learning task, the most prominent one is that of *regret minimization*, where in each learning round the learner suffers regret unless the optimal decision is made (determined by the feedback mechanisms). The main challenge for the learner is to manage the trade-off between exploration and exploitation, i.e., constantly balancing (i) the degree of new information acquisition about some arms' feedback mechanism in order to appropriately expand the current knowledge base (exploration), and (ii) the degree of choosing arms considered to be optimal given the current knowledge base in order to keep the overall regret low (exploitation). In many practical applications, however, the learning task is of a quite different kind, as the focus is rather on finding the (approximately) correct answer to a problem specific question, e.g., which arm is the (approximately) optimal one, within a reasonable time (number of learning rounds). This *pure exploration* learning task can be considered in two variants, namely the fixed confidence and the fixed budget setting. In the former the learner tries to find the answer within as few as possible learning rounds, while guaranteeing a given pre-determined confidence for the correctness of its returned answer. In the latter, it is the other way around, as the learner is provided with a limit on the possible number of learning rounds (budget) and the confidence for the returned answer should be as high as possible. In both variants the main challenge for designing a suitable learner is to specify a clever exploration strategy for finding the correct answer.

In order to model more complex learning settings in practice, the basic setup of MAB problems has been generalized in various ways, such as incorporating additional side information [3, 6, 17, 2] or infinite number of arms [11, 38], just to name a few. Of special practical interest is the generalization of the basic setup, where the learner is allowed to choose specific sets of arms as its action. Consider as an example the online algorithm selection problem with parallel runs, where for sequentially arriving problem instances one selects a subset of available algorithms (solvers) to be run in parallel in order to solve the current problem instance.

If the feedback received is of a numerical nature, this variant has manifested itself under the term *combinatorial bandits* [14] while for feedback of a qualitative nature this variant can be referred to as *preference-based bandits* as put forward by [8]. Combinatorial bandits are further distinguished with respect to the type of feedback between semi-bandit feedback, where feedback of each single arm in the selected subset is observed, and full bandit feedback, where only some aggregated value of the individual numerical feedback of the arms in the selected subset is observed. Although both combinatorial and preference-based bandits consider a similar action set and for both the learner needs to deal with the possibly exponential size of the action set, the process of learning is quite different due to the nature of the observed feedback. The main reason for this is that in preference-based feedback the mutual correlations that may exist between the arms in the chosen subset play a major role in both the assumption about the feedback mechanisms and the learning task from the outset. In contrast to this, the standard setting in combinatorial bandits with semi-bandit feedback is that the individual reward generation mechanisms are independent of the chosen subsets. However, this modeling assumption is questionable in a couple of practical applications, especially when humans provide the feedback. For example, in opinion polls or rating systems where humans rate a subset of objects (political parties/candidates, products, etc.), it is well known that the ratings of the objects may be affected by context effects, i.e., preferences in favor of an object may depend on what other objects are available. In the fields of economics and psychology, context effects are among others divided into compromise [47], attention [22] and similarity [48] effects.

In this paper, we take a step towards unifying these two variants for the best arm identification (BAI) problem in a pure exploration learning setting with fixed budget and non-stochastic feedback mechanisms. The main motivation for this unification is to derive a general purpose learner, which can tackle the BAI problem in both feedback variants. In this way, for example, one can transform a learning problem with numerical signals into a preference-based learning problem and thus conveniently apply such a general purpose learner. Recent works have demonstrated in two different learning scenarios with numerical feedback that such a transformation has great potential [37, 27].

Needless to say, the main challenge is to unify both feedback variants through suitable abstractions allowing them to be treated as instantiations of the same problem class. This bridge is built by dropping

the common independence assumption (of the chosen arm set) for the numerical combinatorial bandits and abstracting the nature of the observations. Additionally, we simply assume that the learner is provided with an appropriate statistic customized to the explicit nature of the feedback. By appropriate choice of the statistic one obtains the respective setting, e.g., the empirical mean for the case of numerical feedback and relative frequencies for the case of preference feedback.

Our contribution. Under mild assumptions on the asymptotic behavior of these statistics, we derive a proper definition of a best arm a learner seeks to find (Section 2) as well as lower bounds on the necessary budget for this task (Section 3). To the best of our knowledge, such lower bounds are novel for non-stochastic settings and the derivation is rather non-standard due to the combinatorial setup of the problem. We suggest a general algorithmic framework suitable to cover the full spectrum of conceivable arm elimination strategies from aggressive to conservative, which we analyze theoretically regarding the algorithms' sufficient and necessary budget to find the best arm (Section 4). As a consequence, we obtain to the best of our knowledge the first algorithm(s) for non-stochastic preference-based bandits as well as for combinatorial bandits under semi-bandit feedback, in which the individual (numerical) feedback received for an arm depends on the chosen subset due to possibly existing mutual correlations between the arms in the chosen subset. The mild assumptions on the asymptotics of the statistics allow to transfer our theoretical results to the stochastic counterparts of the semi-bandit combinatorial and preference-based bandits (Section 5). We demonstrate the usefulness of the generality of our setting in an experimental study for an algorithm selection problem with parallel runs (Section 6), where once again the transformation of numerical feedback to preference feedback plays a key role. Additional experiments are given in the supplementary material, where also all proofs of the theoretical results are collected.

Related Work. A large body of literature considers the combinatorial bandit problem under preference-based feedback, see [8] for an overview. Although *dueling bandits* [52] has established as an overall agreed term for the scenario with actions of size two, the terminology for action sets of larger sizes is still discordant, e.g., multi-dueling [10], battling [42], choice [4] or preselection bandits [9], mainly due to subtle nuances of the motivating practical applications. While pure exploration settings with a stochastic preference-based feedback haven been considered by a series of works [36, 39, 15, 40, 43, 21], a pure exploration setting under a non-stochastic feedback mechanism as in our case has yet to be studied.

Pure exploration has been intensively studied in the basic multi-armed bandits (MAB) setting with stochastic feedback mechanisms as well, see Section 33.5 in [32] for a detailed overview. The non-stochastic variant of the fixed budget MAB setting is considered in [25], which is the backbone for the well-known Hyperband algorithm [34] and additionally inspired in some part the assumptions we make for our work. Initiated by the work of [12] to design learners for regret minimization frameworks which can perform well in both stochastic and non-stochastic settings, the fixed budget framework has been the subject of research by [1] and [45].

Combinatorial bandits with numerical feedback have been introduced by [14] and [16] in a regret minimization framework. The fixed confidence setting for stochastic combinatorial bandits with semi-bandit feedback is studied in [26], and full bandit feedback in [18, 30]. Finally, the best-of-k bandits game introduced in [46], which in some way unifies the combinatorial bandits with binary set-dependent feedback and preference-based bandits in one joint framework similarly as we do in this work. However, they consider a fixed confidence setting with stochastic feedback mechanisms and do not provide a learner for the dependent arm case, although they derive lower bounds on the worst case sample complexity for this case. The only work assuming a set-dependent feedback mechanism in combinatorial bandits with semi-bandit feedback is in [50], where, however, the regret minimization task is studied under stochastic feedback mechanisms. In summary, there seems to be no existing work which considers a pure exploration setting for combinatorial bandits with non-stochastic or even stochastic semi-bandit feedback, where the (mean) rewards of the arms in the chosen subset of arms depends on the subset. Accordingly, our results provide new theoretical contributions to this field.

2 Problem Formulation

In our setup, we assume a set $\mathcal{A}$ of n arms, which we simply identify by their indices, i.e., $\mathcal{A} = [n] = \{1, \dots, n\}$. For some fixed $k < n$ we denote the set of all possible subsets of arms with size of at least 2 and at most k by $\mathcal{Q}_{\leq k} = \{Q \subseteq \mathcal{A} \mid 2 \leq |Q| \leq k\}$. Further, we assume that for any

$Q \in \mathcal{Q}_{\leq k}$ we can query some feedback, in the form of a feedback vector $\mathbf{o}_Q = (o_{i|Q})_{i \in Q} \in D^{|Q|}$ which in turn can be of numerical or qualitative nature specified by the domain D. If we query a subset of arms Q for t many times, then $\mathbf{o}_Q(t)$ is the corresponding feedback vector. We suppose that we are given a suitable statistic s for the type of observation vectors, which maps a multiset of observations to some specific value relevant for the decision making. With this, $\mathbf{s}_Q(t) = (s_{i|Q}(t))_{i \in Q}$ is the statistic vector derived by the sequence of feedback $(\mathbf{o}_Q(t))_t$ of the query set $Q \in \mathcal{Q}_{\leq k}$, and $s_{i|Q}(t) = s(\{o_{i|Q}(1), \ldots, o_{i|Q}(t)\})$ is the relevant statistic for decision making about arm i in the "context" Q after querying Q for t many times.

Examples. For combinatorial bandits with semi-bandit feedback, the observation $\mathbf{o}_Q(t) = (o_{i|Q}(t))_{i \in Q}$ corresponds to the reward one obtains for each arm $i \in Q$ by using Q for the t-th time, so that in particular $D = \mathbb{R}$. The most natural statistic in this case is the empirical mean given for a multiset O of observations by $s(O) = \frac{1}{|O|} \sum_{x \in O} x$, such that $s_{i|Q}(t) = \frac{1}{t} \sum_{t'=1}^{t} o_{i|Q}(t')$, which is also the arguably most prevalent statistic used in the realm of bandit problems for guiding the decision making process. However, other statistics s such as quantiles or the expected shortfall are of interest as well [13].

In the preference-based bandit setting with winner feedback we observe after the t-th usage of the query set Q only a binary (winner) information, i.e., $o_{i|Q}(t) = 1$ if arm i is preferred over the other arms in Q at "pull" t and $o_{i|Q}(t) = 0$ otherwise, so that $D = \{0, 1\}$. Once again the empirical mean of these binary observations is a quite intuitive choice for the statistic s, as in a stochastic feedback setting the corresponding statistic vector $s_{i|Q}(t) = \frac{1}{t} \sum_{t'=1}^{t} o_{i|Q}(t)$ would converge to the probability vector determining how likely an arm will be preferred over all the other arms in the query set Q. For preference-based bandits with full ranking feedback we observe after the t-th usage of the query set Q an entire ranking of the arms in Q, i.e., $o_{i|Q}(t)$ is arm i's rank among the arms in Q at "pull" t, so that $D = \{1, \ldots, k\}$. In such a case the statistic s might be a positional scoring rule [28].

Goal. The goal of the learner in our setting is to find a or the best arm (specified below) within a fixed budget of at most B samples (numbers of queries). For any Q, write $n_Q(t)$ for the number of times Q has been queried until (including) time t. An algorithm, which tackles the problem, chooses at time t a set $Q_t \in \mathcal{Q}_{\leq k}$ and observes as feedback $\mathbf{o}_{Q_t}(n_{Q_t}(t))$ leading to an update of the relevant statistic vector $\mathbf{s}_{Q_t}(n_{Q_t}(t)) = (s_{i|Q_t}(n_{Q_t}(t)))_{i \in Q_t}$.

Best arm. Inspired by the theoretical groundings of Hyperband [25, 34] for best arm identification problems in numerical bandit problems with non-stochastic rewards, we make the following assumption regarding the limit behavior of the statistics

$$(A1) : \forall Q \in \mathcal{Q}_{\leq k} \; \forall i \in Q \; : \; S_{i|Q} \coloneqq \lim_{t \to \infty} s_{i|Q}(t) \text{ exists.}$$

This assumption is in general slightly looser than assuming (stationary) stochastic feedback mechanisms, as (A1) is fulfilled for many prevalent statistics by means of a limits theorem such as the strong law of large numbers. Conceptionally, our Assumption (A1) is similar to the assumption on the sequence of losses in [25], as both have in common that the statistics (losses in [25]) converge to some fixed point, respectively. However, due to the difference of the action spaces (single arms vs. set of arms) and the nature of the feedback (scalar vs. vector observation), our assumption can be seen as a combinatorial extension of the one in [25].

Given assumption (A1) a straightforward notion of a best arm is obtained by leveraging the idea of a Borda winner from dueling bandits with pairs of arms as the possible subsets to more general subsets of arms. A *generalized Borda winner* (GBW) is then an arm which has on average the largest asympotical statistic, i.e.,

$$i^*_{\mathcal{B}} \in \arg\max_{i \in \mathcal{A}} S_i^{\mathcal{B}} = \arg\max_{i \in \mathcal{A}} \frac{\sum_{Q \in \mathcal{Q}_{=k}(i)} S_{i|Q}}{|\mathcal{Q}_{=k}(i)|},$$

where $\mathcal{Q}_{=k}(i) = \{Q \in \mathcal{Q}_{=k} \mid i \in Q\}$ and $S_i^{\mathcal{B}}$ are the asymptotic Borda scores, i.e., the limits according to (A1) of $s_i^{\mathcal{B}}(t) \coloneqq \frac{\sum_{Q \in \mathcal{Q}_{=k}(i)} s_{i|Q}(t)}{|\mathcal{Q}_{=k}(i)|}$. Similarly, a *generalized Copeland winner* (GCopeW) is an arm i, which wins w.r.t. the asymptotic statistics on average on the most query sets, i.e.,

$$i^*_{\mathcal{C}} \in \arg\max_{i \in \mathcal{A}} S_i^{\mathcal{C}} = \arg\max_{i \in \mathcal{A}} \frac{\sum_{Q \in \mathcal{Q}_{=k}(i)} \mathbf{1}\{S_{i|Q} = S_{(1)|Q}\}}{|\mathcal{Q}_{=k}(i)|}.$$

However, both these notions of best arm have two major drawbacks, as there might be multiple GBWs and GCopeWs and due to averaging over all subsets in their definition, there is no way to identify a GBW or a GCopeW within a sampling budget of $o(\binom{n-1}{k-1})$ in the worst case (see Theorem 3.1).

In light of these drawbacks, we specify another reasonable notion of a best arm, for which we leverage the concept of the *generalized Condorcet winner* [21, 4] from the preference-based bandits literature. For this purpose, we introduce the following assumption

$$(A2): \exists i^* \in \mathcal{A} \text{ such that } \forall Q \in \mathcal{Q}_{\leq k}(i^*)\ \forall j \in Q \setminus \{i^*\} \text{ it holds that } S_{i^*|Q} > S_{j|Q},$$

where $\mathcal{Q}_{\leq k}(i) = \{Q \in \mathcal{Q}_{\leq k} \mid i \in Q\}$ for $i \in [n]$. We call i^* the generalized Condorcet winner (GCW), which is the arm dominating all the other arms in each possible query set containing it. It is worth noting that such an arm may not exist, but if it exists, then it is arguably the most natural way to define the optimal arm, even though it may differ from the GBW. Nevertheless, the existence of the generalized Condorcet winner (or simply the Condorcet winner for the case $k = 2$) is a common assumption in the preference-based bandits literature [4, 21, 8]. Additionally, we will show below that identifying a GCW is possible for a sampling budget of size $\Omega(n/k)$ even in worst case scenarios.

Problem characteristics. In light of (A1) and (A2), there are two key characteristics which will determine the appeal of any learner in our setting. The first one is the speed of convergence of the statistics $s_{i|Q}$ to their limit values $S_{i|Q}$. More precisely, the function $\gamma_{i|Q} : \mathbb{N} \to \mathbb{R}$, which is the point-wise smallest non-increasing function fulfilling $|s_{i|Q}(t) - S_{i|Q}| \leq \gamma_{i|Q}(t)$ for any $t \in \mathbb{N}$, plays a major role in characterizing the difficulty of the learning problem. Moreover, the worst speed of convergence function of a query set $Q \in \mathcal{Q}_{\leq k}$ given by $\overline{\gamma}_Q(t) \coloneqq \max_{i \in Q} \gamma_{i|Q}(t)$ and the overall worst speed of convergence function $\overline{\gamma}(t) \coloneqq \max_{Q \in \mathcal{Q}_{\leq k}} \overline{\gamma}_Q(t)$ will be of relevance as well. Assuming a stochastic setting, the role of $\gamma_{i|Q}$ is played by the minimax rate of convergence of the statistic to its population counterpart, e.g., $1/\sqrt{t}$ for the empirical mean and the expected value. Usually the speed of convergence functions will appear implicitly by means of their (quasi-) inverses given by $\gamma_{i|Q}^{-1}(\alpha) \coloneqq \min\{t \in \mathbb{N} \mid \gamma_{i|Q}(t) \leq \alpha\}$, $\bar{\gamma}_Q^{-1}(t) \coloneqq \min_{i \in Q} \gamma_{i|Q}^{-1}(t)$ and $\bar{\gamma}^{-1}(t) \coloneqq \min_{Q \in \mathcal{Q}_{\leq k}} \bar{\gamma}_Q^{-1}(t)$.

The other relevant problem characteristic are the gaps of the limits statistics, i.e., $\Delta_{i|Q} \coloneqq S_{i^*|Q} - S_{i|Q}$ for $i \in [n]$, $Q \in \mathcal{Q}_{\leq k}(i) \cap \mathcal{Q}_{\leq k}(i^*)$. Such gaps are prevalent in the realm of bandit problems, as they can be used to define a measure of (sub-)optimality of an arm in the stochastic feedback case. Note that in our setting this is not straightforward, as the gaps are depending on the query set and more importantly the speed of convergence has a decisive impact on the severeness of these gaps.

3 Lower Bounds

Let us abbreviate $\mathbf{S} \coloneqq (S_{i|Q})_{Q \in \mathcal{Q}_{\leq k}, i \in Q}$ and $\boldsymbol{\gamma} \coloneqq (\gamma_{i|Q}(t))_{Q \in \mathcal{Q}_{\leq k}, i \in Q, t \in \mathbb{N}}$. Given $\mathbf{S}$ and $\boldsymbol{\gamma}$, write $\mathfrak{S}(\mathbf{S}, \boldsymbol{\gamma})$ for the set of all $\mathbf{s} = (s_{i|Q}(t))_{Q \in \mathcal{Q}_{\leq k}, i \in Q, t \in \mathbb{N}}$ that fulfill

(i) $\forall Q \in \mathcal{Q}_{\leq k}, i \in Q$: $S'_{i|Q} = \lim_{t \to \infty} s_{i|Q}(t)$ exists,
(ii) $\forall Q \in \mathcal{Q}_{\leq k}, i \in Q, t \in \mathbb{N}$: $|s_{i|Q}(t) - S'_{i|Q}| \leq \gamma_{i|Q}(t)$,
(iii) $\forall Q \in \mathcal{Q}_{\leq k}$: $\exists \pi_Q : Q \to Q$ bijective such that $S'_{i|Q} = S_{\pi(i)|Q}$ for all $i \in Q$.

For Q and $l \in \{1, \ldots, |Q|\}$ write $S_{(l)|Q}$ for the l-th order statistic of $\{S_{i|Q}\}_{i \in Q}$, i.e., $\{S_{i|Q}\}_{i \in Q} = \{S_{(l)|Q}\}_{l \in Q}$ and $S_{(1)|Q} \geq \cdots \geq S_{(|Q|)|Q}$. If Alg is a (possibly probabilistic) sequential algorithm, we denote by $B(\text{Alg}, \mathbf{s})$ the number of queries made by Alg before termination when started on $\mathbf{s}$. In the following, we provide lower bounds on $\mathbb{E}[B(\text{Alg}, \mathbf{s})]$ for algorithms Alg, which identify, for any instance $\mathbf{s} \in \mathfrak{S}(\mathbf{S}, \boldsymbol{\gamma})$, almost surely one GCW resp. GBW resp. GCopeW of $\mathbf{s}$.

Theorem 3.1. *Let* $\mathbf{S}$ *be such that* $S_{(1)|Q} > S_{(2)|Q}$ *for all* $Q \in \mathcal{Q}_{\leq k}$.
(i) There exists $\mathbf{s} \in \mathfrak{S}(\mathbf{S}, \boldsymbol{\gamma})$ *such that if* Alg *correctly identifies a GCW for any instance in* $\mathfrak{S}(\mathbf{S}, \boldsymbol{\gamma})$*, then*

$$\mathbb{E}\left[B(\text{Alg}, \mathbf{s})\right] \geq \left\lceil \frac{n}{k} \right\rceil \min_{Q \in \mathcal{Q}_{\leq k}, j \in Q} \gamma_{j|Q}^{-1}\left(\frac{S_{(1)|Q} - S_{(|Q|)|Q}}{2}\right).$$

(ii) Assume $(S_{(1)|Q}, \dots, S_{(|Q|)|Q})$ *does not depend on* Q *for any* $Q \in \mathcal{Q}_{=k}$*. If* Alg *correctly identifies a GBW for any instance in* $\mathfrak{S}(\mathbf{S}, \boldsymbol{\gamma})$*, then*

$$\sup_{\mathbf{s} \in \mathfrak{S}(\mathbf{S}, \boldsymbol{\gamma})} \mathbb{E}\left[B(\mathrm{Alg}, \mathbf{s})\right] = \Omega\left(\binom{n-1}{k-1}\right).$$

(iii) If Alg *correctly identifies a GCopeW for any instance in* $\mathfrak{S}(\mathbf{S}, \boldsymbol{\gamma})$*, then it fulfills*

$$\sup_{\mathbf{s} \in \mathfrak{S}(\mathbf{S}, \boldsymbol{\gamma})} \mathbb{E}\left[B(\mathrm{Alg}, \mathbf{s})\right] = \Omega\left(\binom{n-1}{k-1}\right).$$

The theorem is proven in Section B in the supplement, where we provide in fact slightly stronger versions of these bounds.

4 Algorithms

In this section, we present a class of algorithms along with three possible instantiations solving the corresponding learning task for the case of the generalized Condorcet winner being the best arm. In particular, we analyze all algorithms theoretically in terms of their sufficient and necessary budget to find the respective best arm. In light of the results in Theorem 3.1 for GBW and GCopeW identification, it is straightforward that a simple algorithm, which enumerates all possible subsets, pulls each of them equally often in a round-robin fashion, and returns the empirical GBW (or GCopeW) is already optimal. For sake of completeness, we show this result for the case of GBW identification in Section C and also include this simple algorithm in the experimental study below (called ROUNDROBIN).

4.1 Generalized Condorcet Winner Identification

In the following we introduce a general class of algorithms in Algorithm 1 which is instantiable with various different elimination strategies of the arms. Below, we present some instantiations which build on commonly used elimination strategies in the standard multi-armed bandit setting. The idea of Algorithm 1 is simply to maintain a set of active arms, which is successively reduced by following a specific arm elimination strategy (Algorithm 2) referred to as elimination rounds. In each elimination round $r \in \{1, 2, \dots, R\}$ the set of active arms $\mathbb{A}_r$ is partitioned into P_r many sets of size k (up to a possible remainder set) denoted by $(\mathbb{A}_{r,j})_{j \in P_r}$ for which the elimination strategy is applied with a roundwise-dependent budget b_r. The budget allocated to a partition $\mathbb{A}_{r,j}$ in round r is of the form $b_r = \lceil B/(R \cdot P_r) \rceil$ following the idea to split up the available budget equally first for each round and second for all partitions in each round. The explicit arm elimination strategy used in Algorithm 1 is specified by Algorithm 2 and corresponds to pulling the chosen query set Q for a fixed number of times and afterwards keeping only the best $f(|Q|)$ arms of Q. Here, $f : [k] \to [k]$ is an arbitrary function with $f(x) \leq x - 1$ for all x, which essentially determines the aggressiveness or conservativeness of an arm elimination strategy as we will see below.

Algorithm 1 Combinatorial Successive Elimination

Input: set of arms $[n]$, subset size $k \leq n$, sampling budget B, a function $f : [k] \to [k]$, sequence $\{P_r\}_r$ (number of partitions at round r), R (number of rounds in total)
Initialization: $\mathbb{A}_1 \leftarrow [n]$, $r \leftarrow 1$

while $|\mathbb{A}_r| \geq k$ **do**
  $b_r \leftarrow \lfloor B/(P_r R) \rfloor$, $J \leftarrow P_r$
  $\mathbb{A}_{r,1}, \mathbb{A}_{r,2}, \dots, \mathbb{A}_{r,J} \leftarrow \mathrm{Partition}(\mathbb{A}_r, k)$
  if $|\mathbb{A}_{r,J}| < k$ **then**
    $\mathcal{R} \leftarrow \mathbb{A}_{r,J}$, $J \leftarrow J - 1$
  else
    $\mathcal{R} \leftarrow \emptyset$
  end if
  $\mathbb{A}_{r+1} \leftarrow \mathcal{R}$
  for $j \in [J]$ **do**
    $\mathcal{R} \leftarrow \mathrm{ArmElimination}(\mathbb{A}_{r,j}, b_r, f(|\mathbb{A}_{r,j}|))$
    $\mathbb{A}_{r+1} \leftarrow \mathbb{A}_{r+1} \cup \mathcal{R}$
  end for
  $r \leftarrow r + 1$
end while
$\mathbb{A}_{r+1} \leftarrow \emptyset$
while $|\mathbb{A}_r| > 1$ **do**
  $\mathbb{A}_{r+1} \leftarrow \mathrm{ArmElimination}(\mathbb{A}_{r+1}, b_r, f(|\mathbb{A}_{r+1}|))$, $r \leftarrow r + 1$
end while

Output: The remaining item in $\mathbb{A}_r$

In the following we provide three possible instantiations of Algorithm 1 inspired by commonly used elimination strategies in the standard multi-armed bandit setting for pure exploration tasks.

Combinatorial Successive Winner Stays. The most aggressive elimination strategy is to keep only the arm with the best statistic (the winner) of each partition in each round and discard all others from the set of active arms for the next round. Concretely, we use $f^{\mathrm{CSWS}}(s) = 1$ for f in Algorithm 1 in this case.

Algorithm 2 ArmElimination$(\mathbb{A}', b, l)$

1: Use $\mathbb{A}'$ for b times
2: For all $i \in \mathbb{A}'$, update $s_{i|\mathbb{A}'}(b)$
3: Choose an ordering $i_1, \dots, i_{|\mathbb{A}'|}$ of $(s_{i|\mathbb{A}'}(b))_{i \in \mathbb{A}'}$
4: **return** $\{i_1, \dots, i_l\}$

The resulting instantiation of Algorithm 1 is called *Combinatorial Successive Winner Stays* (CSWS), which has at most $R^{CSWS} = \lceil \log_k(n) \rceil + 1$ many rounds in total (at most $\lceil \log_k(n) \rceil$ rounds in the first while-loop and at most 1 in the second). The total number of partitions in round r is at most $P_r^{CSWS} = \lceil n/k^r \rceil$.

Combinatorial Successive Reject. On the other extreme regarding the aggressiveness of the arm elimination strategy is to dismiss only the worst arm of each partition in each round and keep all others in the set of active arms for the next round. More specifically, we use $f^{\mathrm{CSR}}(s) = s - 1$ for this variant, which can be seen as a variant of the Successive Reject algorithm [5] for best arm identification adopted to the combinatorial bandit problem. Consequently, we call the resulting instantiation of Algorithm 1 the *Combinatorial Successive Reject* (CSR) algorithm, whose number of rounds in the first while-loop is at most $\lceil \log_{k-1/k}(1/n) \rceil$ and in the second at most $k-1$. Overall, we have a maximal number of rounds $R^{CSR} = \lceil \log_{k-1/k}(1/n) \rceil + k - 1$ and a maximal number of partitions per round $P_r^{CSR} = \lceil n(1-\frac{1}{k})^{(r-1)}/k \rceil$.

Combinatorial Successive Halving. As a compromise between the aggressive elimination strategy of CSWS and the conservative elimination strategy of CSR one could discard in every elimination round the worse half of all arms in the partition, i.e., using $f^{\mathrm{CSH}}(s) = \lceil s/2 \rceil$ for f in Algorithm 1. This can be seen as a generalization of the successive halving algorithm [25] adopted to the combinatorial bandit problem we are considering. Thus, the instantiation of Algorithm 1 in this spirit will be called the *Combinatorial Successive Halving* (CSH) algorithm. Note that we have at most $\lceil \log_2(n) \rceil$ rounds in the first while-loop and additional $\lceil \log_2(k) \rceil$ in the second while-loop resulting in at most $R^{CSH} = \lceil \log_2(n) \rceil + \lceil \log_2(k) \rceil$ many rounds throughout a run of CSH. Furthermore, we have at most $P_r^{CSH} = \lceil \frac{n}{2^{r-1}k} \rceil$ partitions in round r.

4.2 Theoretical Guarantees

In the following, we derive the sufficient budget for Algorithm 1 to return under assumptions (A1) and (A2) the best arm i^*, i.e., the generalized Condorcet winner. For this purpose, we write $\mathbb{A}_r(i^*)$ for the unique set $\mathbb{A}_{r,j} \in \{\mathbb{A}_{r,1}, \dots, \mathbb{A}_{r,P_r}\}$ with $i^* \in \mathbb{A}_{r,j}$ and define $\Delta_{(l)|Q} = S_{i^*|Q} - S_{(l)|Q}$ for any $Q \subseteq [n]$ with $i^* \in Q$.

Theorem 4.1. *Assume P_r, R are such that Algorithm 1 called with B does not exceed B as total budget. Under Assumptions (A1) and (A2) Algorithm 1 returns i^* if B is larger than*

$$z\left(f, R, \{P_r\}_{1 \le r \le R}\right) \coloneqq R \max_{r \in [R]} P_r \cdot \left\lceil \bar{\gamma}^{-1}_{\mathbb{A}_r(i^*)}\left(\Delta_{(f(|\mathbb{A}_r(i^*)|)+1)|\mathbb{A}_r(i^*)}/2\right) \right\rceil$$

The following theorem indicates optimality of z in the theorem above (cf. Sec. D.2 for the proofs).

Theorem 4.2. *For any distinct asymptotic values* $\mathbf{S}$*, there exists a family of statistics $\{s_{i|Q}(t)\}_{t \in \mathbb{N}, Q \in \mathcal{Q}_{\le k}, i \in Q}$ with $s_{i|Q}(t) \to S_{i|Q}$ for all $i \in [n], Q \in \mathcal{Q}_{\le k}$ such that if Algorithm 1 is used with a budget $B < z\left(f, R, \{P_r\}_{1 \le r \le R}\right)$ then it does* ***not*** *return i^*.*

By means of Theorem 4.1, we can infer the following result regarding the sufficient sampling budget B for the three instantiations to output i^* (cf. Sec. D.3 of the appendix for the proof).

Corollary 4.3. *Under Assumptions (A1) and (A2),* $\mathrm{CSX} \in \{\mathrm{CSWS}, \mathrm{CSR}, \mathrm{CSH}\}$ *returns i^* if it is executed with a budget $B \ge z_{\mathrm{CSX}}$, where $z_{\mathrm{CSX}} \coloneqq z\left(f^{\mathrm{CSX}}, R^{\mathrm{CSX}}, \{P_r^{\mathrm{CSX}}\}_{1 \le r \le R^{\mathrm{CSX}}}\right)$.*

By substituting the concrete values for P_r, R and f of the corresponding instantiation into Corollary 4.3 and using a rough estimate for the inverse function of the speed of convergence, we see that all of

the resulting sufficient budgets are essentially $\tilde{\mathcal{O}}(n/k)$ (see Table 1) almost[1] matching the dependency on n and k in Theorem 3.1. If we would allow the special case of singleton sets of arms as query sets, i.e., $k = 1$, the sufficient budget for CSH matches the one derived in [25] for its non-combinatorial counterpart in the special case of numerical feedback.

Table 1: Sufficient budget for CSWS, CSR and CSH. Here, π is as in (iii) in Section 3.

z_{CSWS}	$\lceil \frac{n}{k} \rceil (\lceil \log_k(n) \rceil + 1) \cdot \max_{Q \in \mathcal{Q}_{\leq k}: i^* \in Q} \max_{i \in Q \setminus \{i^*\}} \left\lceil \bar{\gamma}^{-1} \left(\frac{S_{i^*\|Q} - S_{i\|Q}}{2} \right) \right\rceil$
z_{CSR}	$\lceil \frac{n}{k} \rceil \left(\left\lceil \log_{1-\frac{1}{k}} \left(\frac{1}{n} \right) \right\rceil + k - 1 \right) \cdot \max_{Q \in \mathcal{Q}_{\leq k}: i^* \in Q} \min_{i \in Q \setminus \{i^*\}} \left\lceil \bar{\gamma}^{-1} \left(\frac{S_{i^*\|Q} - S_{i\|Q}}{2} \right) \right\rceil$
z_{CSH}	$\lceil \frac{n}{k} \rceil (\lceil \log_2(n) \rceil + \lceil \log_2(k) \rceil) \cdot \max_{Q \in \mathcal{Q}_{\leq k}: i^* \in Q} \left\lceil \bar{\gamma}^{-1} \left(\frac{S_{i^*\|Q} - S_{\pi(Q)\|Q}}{2} \right) \right\rceil$

Regarding n and k both lower and upper bounds coincide, but the gap-term in the lower bounds include a min-term over $\mathcal{Q}_{\leq k}$, while the gap-term in the upper bound are coming with a max-term over $\mathcal{Q}_{\leq k}$. The difference between these terms depends on the underlying hardness of the bandit problem in terms of $\bar{\gamma}^{-1}$, i.e., how fast the considered statistics converge to their limit values. Due to the generality of our setting it is difficult to specify this difference more explicitly and it would be worth considering this for special cases, i.e., the numerical bandits or preference-based bandits separately.

Finally, it is worth mentioning that all of the three instantiations of Algorithm 1 have only been studied for the case of single arm pulls, but not for pulls of subsets of arms, where additionally a dependency on the set might be present. Thus, the theoretical guarantees are novel in this regard.

5 Applications to Stochastic Settings

Numerical feedback. In stochastic combinatorial bandits [16], each arm-query set pair (i, Q) is associated with a probability distribution $\nu_{i|Q}$ and querying Q for the t-th time results in the feedback $o_{i|Q}(t) \sim \nu_{i|Q}$, usually referred to as a reward (i.e., $D = \mathbb{R}$). The sequence of rewards $\{o_{i|Q}(t)\}_t$ is supposed to be independent and the statistic s is the empirical mean such that (A1) holds by the law of large numbers with $S_{i|Q} = \mathbb{E}_{X \sim \nu_{i|Q}}[X]$. If the $\nu_{i|Q}$ are sub-Gaussian, an anytime confidence bound by [24] based on the law of iterated logarithm ensures $|s_{i|Q}(t) - S_{i|Q}| \leq c_\delta(t)$ for all $t \in \mathbb{N}$ with probability at least $1 - \delta$ for some appropriate function $c_\delta(t) \in \mathcal{O}(\sqrt{t \ln(\ln(t)/\delta)})$. This implies the following result, the proof of which is deferred to Section E.

Corollary 5.1. *Let f, R and $\{P_r\}_{r \in [R]}$ be as in Theorem 4.1 and suppose the reward distributions $\nu_{i|Q}$ to be σ-sub-Gaussian and such that their means $S_{i|Q}$ satisfy (A2). There is a function $C(\delta, \varepsilon, k, R, \sigma)$ in $\mathcal{O}\left(\sigma^2 \varepsilon^{-2} \ln\left(kR/\delta \ln\left(kR\sigma/\varepsilon\delta\right)\right)\right)$ such that if i^* is the optimal arm for $(S_{i|Q})_{Q \in \mathcal{Q}_{\leq k}, i \in Q}$ and $\sup_{Q \in \mathcal{Q}_{\leq k}(i^*)} \Delta_{(f(|Q|)+1)|Q} \leq \varepsilon$, then Algorithm 1 used with a budget B larger than $C(\delta, \varepsilon, k, R, \sigma) \cdot R \max_{r \in [R]} P_r$ returns i^* with probability at least $1 - \delta$.*

Other statistics for numerical feedback. A rich class of statistics can be obtained by applying a linear functional $U(F) = \int r(x) \mathrm{d}F(x)$, where F is a cumulative distribution function (CDF) and $r : \mathbb{R} \to \mathbb{R}$ some measurable function [49], on the empirical CDF, i.e., $\tilde{s}(O, x) = |O|^{-1} \sum_{o \in O} \mathbb{1}\{x \leq o\}$, for any $x \in \mathbb{R}$ and any multiset of (reward) observations O. This leads to the statistics

$$s_{i|Q}(t) = U(\tilde{s}(o_{i|Q}(1), \ldots, o_{i|Q}(t), \cdot)) = \sum_{s=1}^{t} \frac{r(o_{i|Q}(s))}{t},$$

which converge to $S_{i|Q} = \mathbb{E}_{X \sim \nu_{i|Q}}[r(X)]$ by the law of large numbers, provided these expected values exist. For this class of statistics we can show a quite similar result to Corollary 5.1 by generalizing the findings in [24] (see Section E). However, the result is in fact more general than Corollary 5.1 as for r being the identity function we can recover the empirical mean.

[1] Here, $\tilde{\mathcal{O}}$ hides logarithmic factors.

Preference feedback. In the preference-based bandits, we observe when querying Q for the t-th time a categorical random variable with values in Q, i.e., $o_{i|Q}(t) \sim \mathrm{Cat}_Q(\mathbf{p}_Q)$ for some underlying unknown parameter $\mathbf{p}_Q = (p_{i|Q})_{i \in Q}$. Let $w_{i|Q}(t) := \sum_{s \leq t} \mathbf{1}\{o_{i|Q}(t) = i\}$ be the number of times arm i has won in the query set Q until the t-th pull of $\bar{Q}$. We consider as the relevant statistics $s_{i|Q}(t) = \frac{w_{i|Q}(t)}{t}$, which converge to $p_{i|Q} =: S_{i|Q}$ by the law of large numbers. The Dvoretzky-Kiefer-Wolfowitz inequality [19] ensures a concentration inequality on $\sup_{i \in Q} |s_{i|Q}(t) - S_{i|Q}|$, which can be used to deduce the following result (cf. Sec. E for the proof).

Corollary 5.2. *Let f,R and $\{P_r\}_{r \in [R]}$ be as in Theorem 4.1 and suppose preference-based winner feedback with parameter $(p_{i|Q})_{Q \in \mathcal{Q}_{\leq k}, i \in Q}$, which satisfies (A2). There is a function $C(\delta, \varepsilon, k, R) \in \mathcal{O}\left(\varepsilon^{-2} \ln\left(R/\delta\varepsilon^4\right)\right)$ with the following property: If i^* is the optimal arm and $\sup_{Q \in \mathcal{Q}_{\leq k}(i^*)} \Delta_{(f(|Q|)+1)|Q} \leq \varepsilon$, then Algorithm 1 used with a budget B larger than $C(\delta, \varepsilon, k, R) \cdot R \max_{r \in [R]} P_r$ returns i^* with probability at least $1 - \delta$.*

By substituting the concrete values for P_r, R and f of the corresponding instantiation of Algorithm 1 into the bound on the budget in Corollary 5.2 (compare to Table 1), we see that each of the three resulting bounds almost matches the optimal sample complexity bounds for identifying the (generalized) Condorcet Winner under fixed confidence in preference-based bandits [8, 21] indicating near optimality of the algorithms in stochastic settings. However, since no stochastic counterpart of our combinatorial setting for the numerical case exists, it would be interesting to investigate whether the analogous implication by means of Corollary 5.1 for the three algorithms is nearly optimal as well. We leave this to future work, as it is beyond the scope of our work.

6 Experimental Section

In this section we present an experimental study for our proposed algorithms on an algorithm selection problem. Further experiments, also on synthetic data and with other statistics are provided in the supplementary material in Section G.

Setting. In the following, we consider an algorithm selection problem, where the goal is to select the most efficient algorithm for solving an instance of a satisfiability (SAT) problem. For this, we randomly chose $n = 20$ parameterizations of the SAPS solver [23] which represent our candidate algorithms and correspond to the arms in our terminology. Our possible problem instances are sampled from the first 5000 problem instances from the sat_SWGCP folder of the AClib[2]. We compare CSWS, CSR, CSH and ROUNDROBIN on this problem with the Successive Halving (SH) algorithm [25]. To the best of our knowledge, there are no algorithms available as baselines, which are designed for the pure exploration problem with finite budget and subsets of arms as the actions, e.g., [4] investigates a regret minimization problem, while [21] is dealing with a stochastic pure exploration setting with fixed-confidence. However, Successive Halving serves as a baseline, which we included as a representative for the algorithms dealing with a pure exploration problem with finite budget and single arms as the actions. In each learning round, we randomly draw a problem instance from the 5000 problem instances without replacement and then start a parallel solution process with the SAPS parameterizations chosen by the corresponding learning algorithm (only one parameterization for the case of SH), where the process is stopped as soon as the first algorithm has solved the current instance. In particular, one obtains only for the "finisher" SAPS parameterization an explicit numerical value (its runtime) among the chosen set of SAPS parameterizations, as the others are right-censored. Since our proposed algorithms are designed for the case, in which feedback for all arms in the pulled query set is observed, while SH is designed for the case in which only a single arm is queried resulting in a single feedback, we enlarge the available budget for SH to $k \cdot B$ for a fairer comparison.

Instantiation of CSE. Although we could consider the negative runtimes of the parameterizations as rewards (i.e., runtimes correspond to losses) and use a statistic suitable for numerical feedback for the combinatorial successive elimination (CSE) approaches, there might be a major disadvantage due to the censored runtimes. Indeed, in order to apply a statistic suitable for numerical feedback, some sensible imputation technique is required to deal with the censored observations, which in turn could introduce a severe bias. However, thanks to the generality of our framework, we can simply

[2] http://www.aclib.net

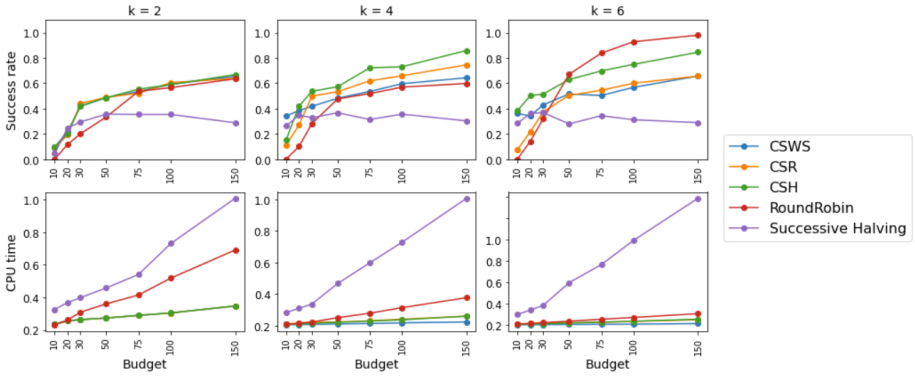

Figure 1: Success rates and runtimes for different subset sizes k and budgets B.

interpret the observed feedback as a preference in the sense that the "finisher" SAPS parameterization is preferred over the others in the chosen set of parameterizations. In this way, using a statistic based on preference-based feedback defuses the bias issue. Quite naturally, we use the relative frequency statistic for preference feedback as specified in Section 5.

Analysis. As the best arm (SAPS parameterization) we use the one having the smallest empirical runtime over all problem instances such that ROUNDROBIN and SH will tendentially return this arm if the budget is sufficiently large. The resulting success rates for our proposed algorithms and SH of identifying the best arm are illustrated in the top panel of Figure 1. One can see, that the algorithms which follow our CSE strategy significantly outperform SH if the budget is sufficiently large. In addition, CSWS, CSR and CSH identify the best arm more often than ROUNDROBIN if the subset sizes k are small, which is a realistic situation in practice. Moreover, the bottom panel in Figure 1 shows the overall runtimes of the algorithms revealing that SH takes much longer than CSWS, CSR and CSH and as expected the difference in the runtimes gets larger with the subset sizes k. Quite interestingly, even ROUNDROBIN needs a longer runtime than the CSE approaches, although it queries the same number of subsets and also stops the respective run as soon as the "finisher" SAPS parameterization is clear. Thus, the differences in the runtimes of ROUNDROBIN and the CSE approaches are only due to the fact that the latter discard the slowest SAPS parameterizations quickly and do not run them again, while ROUNDROBIN uses throughout all subsets the same amount of time, even if they contain only bad performing parameterizations. In other words, the differences are due to the sophisticated strategies of the CSE approaches.

7 Future Work

For future work, it would be interesting to investigate whether switching the elimination strategy during the learning process leads to any performance improvements both theoretically and empirically. A similar question could be asked regarding the considered statistic for numerical feedback variants. Further, the goal of identifying the best set of arms in our scenario would also be interesting. However, in the case where the observations depend on the chosen set of arms, it is far from obvious how to define a suitable optimality term (cf. Sec. 6.3.2 in [8]). Finally, a more extensive experimental study would definitely be a relevant future direction of research, especially for hyperparameter optimization problems with possible parallelization options such as in [33] or for more general algorithm configuration problems [44].

Acknowledgments and Disclosure of Funding

This work was partially supported by the German Research Foundation (DFG) within the project "Online Preference Learning with Bandit Algorithms" (project no. 317046553) and by the research training group "Dataninja" (Trustworthy AI for Seamless Problem Solving: Next Generation Intelligence Joins Robust Data Analysis) funded by the German federal state of North Rhine-Westphalia.

References

[1] Yasin Abbasi-Yadkori, Peter Bartlett, Victor Gabillon, Alan Malek, and Michal Valko. Best of both worlds: Stochastic & adversarial best-arm identification. In *Proceedings of Annual Conference on Learning Theory (COLT)*, pages 918–949, 2018.

[2] Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved Algorithms for Linear Stochastic Bandits. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, pages 2312–2320, 2011.

[3] Naoki Abe and Philip M Long. Associative reinforcement learning using linear probabilistic concepts. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 3–11, 1999.

[4] Arpit Agarwal, Nicholas Johnson, and Shivani Agarwal. Choice bandits. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[5] Jean Yves Audibert, Sébastien Bubeck, and Rémi Munos. Best arm identification in multi-armed bandits. In *Proceedings of Annual Conference on Learning Theory (COLT)*, pages 41–53, 2010.

[6] Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422, 2002.

[7] Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In *Proceedings of IEEE 36th Annual Foundations of Computer Science*, pages 322–331. IEEE, 1995.

[8] Viktor Bengs, Róbert Busa-Fekete, Adil El Mesaoudi-Paul, and Eyke Hüllermeier. Preference-based online learning with dueling bandits: A survey. *Journal of Machine Learning Research*, 22:1–108, 2021.

[9] Viktor Bengs and Eyke Hüllermeier. Preselection bandits. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 778–787, 2020.

[10] Brian Brost, Yevgeny Seldin, Ingemar J. Cox, and Christina Lioma. Multi-dueling bandits and their application to online ranker evaluation. In *Proceedings of ACM International Conference on Information and Knowledge Management (CIKM)*, pages 2161–2166, 2016.

[11] Sébastien Bubeck, Rémi Munos, Gilles Stoltz, and Csaba Szepesvári. X-armed bandits. *Journal of Machine Learning Research*, 12(5), 2011.

[12] Sébastien Bubeck and Aleksandrs Slivkins. The best of both worlds: Stochastic and adversarial bandits. In *Proceedings of Annual Conference on Learning Theory (COLT)*, pages 42.1–42.23, 2012.

[13] Asaf Cassel, Shie Mannor, and Assaf Zeevi. A general approach to multi-armed bandits under risk criteria. In *Proceedings of Annual Conference on Learning Theory (COLT)*, pages 1295–1306, 2018.

[14] Nicolò Cesa-Bianchi and Gábor Lugosi. Combinatorial bandits. *Journal of Computer and System Sciences*, 78(5):1404–1422, 2012.

[15] Wei Chen, Yihan Du, Longbo Huang, and Haoyu Zhao. Combinatorial pure exploration for dueling bandit. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 1531–1541. PMLR, 2020.

[16] Wei Chen, Yajun Wang, and Yang Yuan. Combinatorial multi-armed bandit: General framework and applications. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 151–159, 2013.

[17] Wei Chu, Lihong Li, Lev Reyzin, and Robert E. Schapire. Contextual bandits with linear payoff functions. In *Proceedings of International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 208–214, 2011.

[18] Yihan Du, Yuko Kuroki, and Wei Chen. Combinatorial pure exploration with full-bandit or partial linear feedback. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 7262–7270, 2021.

[19] Aryeh Dvoretzky, Jack Kiefer, and Jacob Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, 27(3):642 – 669, 1956.

[20] Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Proceedings of Annual Conference on Learning Theory (COLT)*, pages 998–1027, 2016.

[21] Björn Haddenhorst, Viktor Bengs, and Eyke Hüllermeier. Identification of the generalized Condorcet winner in multi-dueling bandits. *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 34, 2021.

[22] Joel Huber and Christopher Puto. Market boundaries and product choice: Illustrating attraction and substitution effects. *Journal of Consumer Research*, 10(1):31–44, 1983.

[23] Frank Hutter, Dave AD Tompkins, and Holger H Hoos. Scaling and probabilistic smoothing: Efficient dynamic local search for SAT. In *International Conference on Principles and Practice of Constraint Programming*, pages 233–248. Springer, 2002.

[24] Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lil'UCB : An optimal exploration algorithm for multi-armed bandits. In *Proceedings of Annual Conference on Learning Theory (COLT)*, volume 35, pages 423–439, 2014.

[25] Kevin Jamieson and Ameet Talwalkar. Non-stochastic best arm identification and hyperparameter optimization. In *Proceedings of International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 240–248, 2016.

[26] Marc Jourdan, Mojmír Mutnỳ, Johannes Kirschner, and Andreas Krause. Efficient pure exploration for combinatorial bandits with semi-bandit feedback. In *Proceedings of the International Conference on Algorithmic Learning Theory (ALT)*, pages 805–849, 2021.

[27] Johannes Kirschner and Andreas Krause. Bias-robust Bayesian optimization via dueling bandits. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 5595–5605, 2021.

[28] Anna Korba. *Learning from ranking data: theory and methods*. PhD thesis, Université Paris-Saclay (ComUE), 2018.

[29] Michael R. Kosorok. *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York, USA, 2008.

[30] Yuko Kuroki, Liyuan Xu, Atsushi Miyauchi, Junya Honda, and Masashi Sugiyama. Polynomial-time algorithms for multiple-arm identification with full-bandit feedback. *Neural Computation*, 32(9):1733–1773, 2020.

[31] Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4–22, 1985.

[32] Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020.

[33] Liam Li, Kevin Jamieson, Afshin Rostamizadeh, Ekaterina Gonina, Jonathan Ben-tzur, Moritz Hardt, Benjamin Recht, and Ameet Talwalkar. A system for massively parallel hyperparameter tuning. *Proceedings of Machine Learning and Systems (MLSys)*, 2:230–246, 2020.

[34] Lisha Li, Kevin Jamieson, Giulia DeSalvo, Afshin Rostamizadeh, and Ameet Talwalkar. Hyperband: A novel bandit-based approach to hyperparameter optimization. *Journal of Machine Learning Research*, 18(1):6765–6816, 2017.

[35] Pascal Massart. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The Annals of Probability*, 18(3):1269 – 1283, 1990.

[36] Soheil Mohajer, Changho Suh, and Adel Elmahdy. Active learning for top-k rank aggregation from noisy comparisons. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 2488–2497, 2017.

[37] Felix Mohr, Viktor Bengs, and Eyke Hüllermeier. Single player Monte-Carlo tree search based on the Plackett-Luce Model. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 35, pages 12373–12381, 2021.

[38] Rémi Munos. From bandits to Monte-Carlo tree search: The optimistic principle applied to optimization and planning. *Foundations and Trends® in Machine Learning*, 7(1):1–129, 2014.

[39] Wenbo Ren, Jia Liu, and Ness Shroff. On sample complexity upper and lower bounds for exact ranking from noisy comparisons. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, pages 10014–10024, 2019.

[40] Wenbo Ren, Jia Liu, and Ness Shroff. The sample complexity of best-k items selection from pairwise comparisons. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 8051–8072, 2020.

[41] Herbert Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535, 1952.

[42] Aadirupa Saha and Aditya Gopalan. Battle of bandits. In *Proceedings of Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 805–814, 2018.

[43] Aadirupa Saha and Aditya Gopalan. From PAC to instance-optimal sample complexity in the Plackett-Luce model. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 8367–8376, 2020.

[44] Elias Schede, Jasmin Brandt, Alexander Tornede, Marcel Wever, Viktor Bengs, Eyke Hüllermeier, and Kevin Tierney. A survey of methods for automated algorithm configuration. *Journal of Artificial Intelligence Research*, 75, 2022.

[45] Cong Shen. Universal best arm identification. *IEEE Transactions on Signal Processing*, 67(17):4464–4478, 2019.

[46] Max Simchowitz, Kevin Jamieson, and Benjamin Recht. Best-of-k-bandits. In *Proceedings of Annual Conference on Learning Theory (COLT)*, pages 1440–1489, 2016.

[47] Itamar Simonson. Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, 16(2):158–174, 1989.

[48] Amos Tversky. Elimination by aspects: A theory of choice. *Psychological Review*, 79(4):281, 1972.

[49] Larry Wasserman. *All of Statistics: A concise Course in Statistical Inference*. Springer Science & Business Media, 2013.

[50] Shuo Yang, Tongzheng Ren, Inderjit S Dhillon, and Sujay Sanghavi. Combinatorial bandits without total order for arms. *arXiv preprint arXiv:2103.02741*, 2021.

[51] Andrew Chi-Chin Yao. Probabilistic computations: Toward a unified measure of complexity. In *18th Annual Symposium on Foundations of Computer Science (SFCS)*, pages 222–227, 1977.

[52] 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.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] See Section 1.
   (b) Did you describe the limitations of your work? [Yes] See Section 5 and also our discussion about future work (Section 7).
   (c) Did you discuss any potential negative societal impacts of your work? [No] The results are rather theoretical and do not have societal impacts.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes] In Section 2 we describe and motivate our assumptions (A1) and (A2).
   (b) Did you include complete proofs of all theoretical results? [Yes] See Section B for the proofs of the lower bounds results in Section 3, Section D for the results in Section 4.2 and finally Section E for the results in Section 5.
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] See supplemental material.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See Section 6.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [No]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [No] The runs of the experiments are so fast, that the resources do not matter.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] See Section 6.
   (b) Did you mention the license of the assets? [No] But assets are cited.
   (c) Did you include any new assets either in the supplemental material or as a URL? [No]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [No] But the data is freely available.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [No] Not necessary, because the used data are runtimes of algorithms and does not contain personally identifiable information or offensive content.
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [No] Not applicable.
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [No] Not applicable.
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [No]

A List of Symbols

The following table contains a list of symbols that are frequently used in the main paper as well as in the following supplementary material.

Basics	
$\mathbb{1}\{\cdot\}$	indicator function
$\mathbb{N}$	set of natural numbers (without 0), i.e., $\mathbb{N} = \{1, 2, 3, \dots\}$
$\mathbb{R}$	set of real numbers
D	observation domain (categorical or numerical)
$\mathcal{A} = [n]$	set of arms
n	number of arms
k	maximal possible subset size
B	budget for the learner
$\mathcal{Q}_{\leq k}$	all subsets of $\mathcal{A}$ of size $\leq k$: $\{Q \subseteq \mathcal{A} \mid 2 \leq \lvert Q\rvert \leq k\}$
$\mathcal{Q}_{\leq k}(i)$	all subsets in $\mathcal{Q}_{\leq k}$ which contain arm i: $\{Q \in \mathcal{Q}_{\leq k} \mid i \in Q\}$
$\mathcal{Q}_{=k}$	all subsets of $\mathcal{A}$ of size k : $\{Q \subseteq \mathcal{A} \mid \lvert Q\rvert = k\}$
$\mathcal{Q}_{=k}(i)$	all subsets of $\mathcal{A}$ of size k which contain arm i : $\{Q \in \mathcal{Q}_{=k} \mid i \in Q\}$
$\mathbf{o}_Q(t)$	observed feedback vector by querying Q for the t-th time
Modelling related	
s	relevant statistic for the decision making process
$s_{i\lvert Q}(t)$	statistics for arm $i \in Q$ derived by the observed feedback at the t-th usage of query set Q
$\mathbf{s}_Q(t)$	vector of statistics for all arms in the query set Q after its t-th usage: $(s_{i\lvert Q})_{i \in Q}(t)$
$S_{i\lvert Q}$	limit of the statistics for arm i in query set Q: $\lim_{t\to\infty} s_{i\lvert Q}(t)$
i^*	best arm or generalized Condorcet winner: $\forall Q \in \mathcal{Q}_{\leq k}$ with $i^* \in Q$ it holds that $S_{i^*\lvert Q} > S_{j\lvert Q}$ for any $j \in Q \backslash \{i^*\}$
$s_i^{\mathcal{B}}(t)$	Borda score of arm i at time t: $\sum_{Q \in \mathcal{Q}_{=k}(i)} s_{i\lvert Q}(t) / \lvert \mathcal{Q}_{=k}(i)\rvert$
$S_i^{\mathcal{B}}$	limit Borda score of arm i: $\lim_{t\to\infty} s_i^{\mathcal{B}}(t)$
$i_{\mathcal{B}}^*$	generalized Borda winner: $i_{\mathcal{B}}^* \in \arg\max_{i \in \mathcal{A}} S_i^{\mathcal{B}}$
$n_Q(t)$	number of times query set Q was used until time t
$\gamma_{i\lvert Q}(t)$	point-wise smallest non-increasing function bounding the difference $\lvert s_{i\lvert Q}(t) - S_{i\lvert Q}\rvert$ (rate of convergence)
$\bar{\gamma}_Q(t)$	maximal $\gamma_{i\lvert Q}(t)$ over all $i \in Q$
$\bar{\gamma}(t)$	maximal $\gamma_Q(t)$ over all $Q \in \mathcal{Q}_{\leq k}$
$\gamma_{i\lvert Q}^{-1}(\alpha)$	quasi-inverse of $\gamma_{i\lvert Q}$: $\min\{t \in \mathbb{N} \mid \gamma_{i\lvert Q}(t) \leq \alpha\}$
$\bar{\gamma}_Q^{-1}(t)$	minimal $\gamma_{i\lvert Q}(t)$ over all $i \in Q$
$\bar{\gamma}^{-1}(t)$	minimal $\gamma_Q(t)$ over all $Q \in \mathcal{Q}_{\leq k}$
$\hat{\gamma}_i(t)$	rate of convergence of the Borda score for arm i: $\frac{1}{\lvert \mathcal{Q}_{=k}(i)\rvert} \sum_{Q \in \mathcal{Q}_{=k}(i)} \gamma_{i\lvert Q}(t)$
$\hat{\gamma}_{i,j}^{\max}(t)$	$\max\{\hat{\gamma}_i(t), \hat{\gamma}_j(t)\}$.
$\Delta_{i\lvert Q}$	gap of the limit statistic of arm $i \in Q$ to the limit statistic of the generalized Condorcet winner: $\lvert S_{i^*\lvert Q} - S_{i\lvert Q}\rvert$ for any $Q \in \mathcal{Q}_{\leq k}(i) \cap \mathcal{Q}_{\leq k}(i^*)$
$S_{(l)\lvert Q}, \Delta_{(l)\lvert Q}$	l-th order statistic of $\{S_{i\lvert Q}\}_{i \in Q}$ for $l \in \{1, 2, \dots, \lvert Q\rvert\}$ and its gap $\Delta_{(l)\lvert Q} = S_{i^*\lvert Q} - S_{(l)\lvert Q}$
Algorithm related	
f	function from $[k]$ to $[k]$ specifying the nature of the arm elimination strategy
$R, R^{\mathbb{A}}$	number of rounds of the learning algorithm ($\mathbb{A}$)
$P_r, P_r^{\mathbb{A}}$	number of partitions of the learning algorithm ($\mathbb{A}$) in round r
$\mathbb{A}_{r,j}$	j-th partition in round r
$\mathbb{A}_r(i^*)$	the partition in round r containing i^* (emptyset otherwise)
b_r	budget used in round r for a partition
$z_{\mathbb{A}}$	sufficient budget for learning algorithm $\mathbb{A}$ to return i^* (or $i_{\mathcal{B}}^*$ if $\mathbb{A}$ is ROUNDROBIN)
ROUNDROBIN	the naïve algorithm introduced in Section C
CSE	the generic *combinatorial successive elimination* algorithm (Algorithm 1)
CSWS	the *combinatorial successive winner stays* algorithm resulting by using $f(x) = 1$ in CSE
CSR	the *combinatorial successive rejects* algorithm resulting by using $f(x) = x - 1$ in CSE
CSH	the *combinatorial successive halving* algorithm resulting by using $f(x) = \lceil x/2 \rceil$ in CSE
SH	the *successive halving* algorithm for pure exploration settings in standard multi-armed bandits (cf. [25])
GBW	Generalized Borda winner
GCW	Generalized Condorcet winner

B Proofs for Section 3

In this section, we prove the general lower bounds on the necessary budget for identifying the generalized Condorcet winner (GCW), the generalized Borda winner (GBW) or the generalized Copeland winner (GCopeW). For this purpose, let us first fix some further notation. If Alg is a possibly probabilistic algorithm and $\mathbf{s}$ is fixed, we write $\mathrm{Alg}(\mathbf{s})$ for the output of Alg executed on the instance $\mathbf{s}$. We restrict ourselves only to algorithms whose output is solely determined by the sequence of observations it has received as well as the corresponding statistics. Moreover, for $Q \in \mathcal{Q}_{\leq k}$, we write $B_Q(\mathrm{Alg}, \mathbf{s}) \in \mathbb{N} \cup \{\infty\}$ for the number of times Alg queries Q when started on instance $\mathbf{s}$. Note that $\mathrm{Alg}(\mathbf{s})$ as well as $B_Q(\mathrm{Alg}, \mathbf{s})$ and $B(\mathrm{Alg}, \mathbf{s}) = \sum_{Q \in \mathcal{Q}_{\leq k}} B_Q(\mathrm{Alg}, \mathbf{s})$ are random variables, because they depend on the innate randomness of Alg.
Given $\mathbf{s}$, let us write $\mathrm{GCW}(\mathbf{s})$, $\mathrm{GBW}(\mathbf{s})$ and $\mathrm{GCopeW}(\mathbf{s})$ for the set of all GCWs, GBWs and GCopeWs of $\mathbf{s}$, respectively. In case $|\mathrm{GCW}(\mathbf{s})| = 1$, $|\mathrm{GBW}(\mathbf{s})| = 1$ resp. $|\mathrm{GCopeW}(\mathbf{s})| = 1$, with a slight abuse of notation, we may denote by $\mathrm{GCW}(\mathbf{s})$, $\mathrm{GBW}(\mathbf{s})$ resp. $\mathrm{GCopeW}(\mathbf{s})$ simply the only GCW, GBW resp. GCopeW of $\mathbf{s}$. Recall that the GCW, the GBWs and the GCopeWs of $\mathbf{s}$ only depend on the limits $\mathbf{S} = (S_{i|Q})_{Q \in \mathcal{Q}_{\leq k}, i \in Q}$ with $S_{i|Q} = \lim_{t \to \infty} s_{i|Q}(t)$.

Definition B.1. *Let* Alg *be a (possibly probabilistic) sequential algorithm.*

(i) Alg *solves* $\mathcal{P}_{\mathrm{GCW}}(\mathbf{S}, \boldsymbol{\gamma})$ *if* $\mathbb{P}(\mathrm{Alg}(\mathbf{s}) \in \mathrm{GCW}(\mathbf{s})) = 1$ *for any* $\mathbf{s}$ *in* $\mathfrak{S}(\mathbf{S}, \boldsymbol{\gamma})$.

(ii) Alg *solves* $\mathcal{P}_{\mathrm{GBW}}(\mathbf{S}, \boldsymbol{\gamma})$ *if* $\mathbb{P}(\mathrm{Alg}(\mathbf{s}) \in \mathrm{GBW}(\mathbf{s})) = 1$ *for any* $\mathbf{s}$ *in* $\mathfrak{S}(\mathbf{S}, \boldsymbol{\gamma})$.

(iii) Alg *solves* $\mathcal{P}_{\mathrm{GCopeW}}(\mathbf{S}, \boldsymbol{\gamma})$ *if* $\mathbb{P}(\mathrm{Alg}(\mathbf{s}) \in \mathrm{GCopeW}(\mathbf{s})) = 1$ *for any* $\mathbf{s}$ *in* $\mathfrak{S}(\mathbf{S}, \boldsymbol{\gamma})$.

B.1 Proof of Theorem 3.1 (i): Lower Bound for GCW Identification

The proof of (i) in Theorem 3.1 is prepared with the next lemma.

Lemma B.2. *Let* Alg *be a deterministic solution to* $\mathcal{P}_{\mathrm{GCW}}(\mathbf{S}, \boldsymbol{\gamma})$ *and* $\mathbf{s}, \mathbf{s}' \in \mathfrak{S}(\mathbf{S}, \boldsymbol{\gamma})$.

(i) If $\mathrm{Alg}(\mathbf{s}) \neq \mathrm{Alg}(\mathbf{s}')$*, then*

$$\exists Q \in \mathcal{Q}_{\leq k}, i \in Q,\ t \in \{1, \dots, \min\{B_Q(\mathrm{Alg}, \mathbf{s}), B_Q(\mathrm{Alg}, \mathbf{s}')\}\} \,:\, s_{i|Q}(t) \neq s'_{i|Q}(t).$$

(ii) If $\mathbf{s}$ *and* $\mathbf{s}'$ *coincide on* $\{t < B'\}$ *and on* $\widetilde{\mathcal{Q}} \subseteq \mathcal{Q}_{\leq k}$ *in the sense that*

$$\forall Q \in \mathcal{Q}_{\leq k}, \forall i \in Q, \forall t < B' \,:\, s_{i|Q}(t) = s'_{i|Q}(t) \tag{1}$$

and

$$\forall Q \in \widetilde{\mathcal{Q}}, \forall i \in Q, \forall t \in \mathbb{N} \,:\, s_{i|Q}(t) = s'_{i|Q}(t), \tag{2}$$

then $\mathrm{Alg}(\mathbf{s}) \neq \mathrm{Alg}(\mathbf{s}')$ *implies*

$$\exists Q \in \mathcal{Q}_{\leq k} \setminus \widetilde{\mathcal{Q}} \,:\, \min\{B_Q(\mathrm{Alg}, \mathbf{s}), B_Q(\mathrm{Alg}, \mathbf{s}')\} \geq B'.$$

Proof. (i) To prove the contraposition, suppose that

$$\forall Q \in \mathcal{Q}_{\leq k}, i \in Q,\ t \in \{1, \dots, \min\{B_Q(\mathrm{Alg}, \mathbf{s}), B_Q(\mathrm{Alg}, \mathbf{s}')\}\} \,:\, s_{i|Q}(t) = s'_{i|Q}(t) \tag{3}$$

holds.
Claim 1: $B_Q(\mathrm{Alg}, \mathbf{s}) = B_Q(\mathrm{Alg}, \mathbf{s}')$ for any $\mathcal{Q} \in \mathcal{Q}_{\leq k}$.
Proof: Assume this was not the case. Let $Q \in \mathcal{Q}_{\leq k}$ be the first set, for which Alg exceeds its budget on one of $\mathbf{s}$, $\mathbf{s}'$ but does not reach it on the other instance, and suppose w.l.o.g. $B_Q(\mathrm{Alg}, \mathbf{s}) > B_Q(\mathrm{Alg}, \mathbf{s}')$. Since Alg has observed until this point exactly the same feedback on $\mathbf{s}$ as on $\mathbf{s}'$, this is a contradiction as Alg is deterministic. ■
Combining Claim 1 and (3) yields that Alg observes on $\mathbf{s}$ exactly the same feedback as on $\mathbf{s}'$ until its termination. Since Alg is deterministic, this implies $\mathrm{Alg}(\mathbf{s}) = \mathrm{Alg}(\mathbf{s}')$.

(ii) If $\mathrm{Alg}(\mathbf{s}) \neq \mathrm{Alg}(\mathbf{s}')$, then (i) together with (2) yields

$$\exists Q \in \mathcal{Q}_{\leq k} \setminus \widetilde{\mathcal{Q}}, i \in Q, t \leq \min\{B_Q(\mathrm{Alg}, \mathbf{s}), B_Q(\mathrm{Alg}, \mathbf{s}')\} \,:\, s_{i|Q}(t) \neq s'_{i|Q}(t),$$

and thus (1) implies

$$\exists Q \in \mathcal{Q}_{\leq k} \setminus \widetilde{\mathcal{Q}} \,:\, \min\{B_Q(\mathrm{Alg}, \mathbf{s}), B_Q(\mathrm{Alg}, \mathbf{s}')\} \geq B'.$$

□

Lemma B.2 is the main ingredient for the proof of Theorem 3.1, as we first analyze the lower bound for deterministic algorithms and then apply Yao's minimax principle [51] to infer the lower bound for any randomized algorithm.

Proof of Theorem 3.1 (i). We split the proof into two parts.

Part 1: The statement holds in case Alg **is a deterministic algorithm.**
Abbreviate $B' \coloneqq \min_{Q \in \mathcal{Q}_{\leq k}} \min_{j \in Q} \gamma_{j|Q}^{-1}\left(\frac{S_{(1)|Q} - S_{(|Q|)|Q}}{2}\right)$. Fix a family $\{\pi_Q\}_{Q \in \mathcal{Q}_{\leq k}}$ of permutations $\pi_Q : Q \mapsto Q$ such that $S_{\pi_Q(1)|Q} = S_{(1)|Q}$ holds for any $Q \in \mathcal{Q}_{\leq k}(1)$, and define $\mathbf{s} = (s_{i|Q}(t))_{Q \in \mathcal{Q}_{\leq k}, i \in Q, t \in \mathbb{N}}$ via

$$s_{i|Q}(t) \coloneqq \begin{cases} \frac{S_{(1)|Q} + S_{(|Q|)|Q}}{2}, & \text{if } t < B', \\ S_{\pi_Q(i)|Q}, & \text{if } t \geq B'. \end{cases}$$

Regarding our assumption on $\mathbf{S}$, $\mathrm{GCW}(\mathbf{s}) = 1$ holds by construction. For $t < B' \leq \gamma_{i|Q}^{-1}\left(\frac{S_{(1)|Q} - S_{(|Q|)|Q}}{2}\right)$, which implies $\gamma_{i|Q}(t) \geq \frac{S_{(1)|Q} - S_{(|Q|)|Q}}{2}$, we have due to $S_{(1)|Q} \geq S_{i|Q} \geq S_{(|Q|)|Q}$ the inequality

$$\begin{aligned} &\left|s_{i|Q}(t) - \lim_{t \to \infty} s_{i|Q}(t)\right| = \left|\frac{S_{(1)|Q} + S_{(|Q|)|Q}}{2} - S_{i|Q}\right| \\ &\leq \max\left\{S_{(1)|Q} - \frac{S_{(1)|Q} + S_{(|Q|)|Q}}{2}, \frac{S_{(1)|Q} + S_{(|Q|)|Q}}{2} - S_{(|Q|)|Q}\right\} \\ &= \frac{S_{(1)|Q} - S_{(|Q|)|Q}}{2} \\ &\leq \gamma_{i|Q}(t) \end{aligned}$$

for any $i \in Q$. This shows $\mathbf{s} \in \mathfrak{S}(\mathbf{S}, \boldsymbol{\gamma})$.

For any $l \in \{2, \ldots, n\}$ define an instance $\mathbf{s}^l = (s^l_{i|Q}(t))_{Q \in \mathcal{Q}_{\leq k}, i \in Q, t \in \mathbb{N}}$ such that $s^l_{\cdot|Q}(\cdot) = s_{\cdot|Q}(\cdot)$ for any $Q \in \mathcal{Q}_{\leq k}$ with $l \notin Q$ and

$$s^l_{i|Q}(t) \coloneqq \begin{cases} \frac{S_{(1)|Q} + S_{(|Q|)|Q}}{2}, & \text{if } t < B', \\ S_{(1)|Q}, & \text{if } t \geq B' \text{ and } i = l, \\ S_{l|Q}, & \text{if } t \geq B' \text{ and } i = \mathrm{argmax}_{j \in Q} S_{j|Q} \\ s_{i|Q}(t), & \text{else}, \end{cases}$$

for all $Q \in \mathcal{Q}_{\leq k}(l)$, $i \in Q$ and $t \in \mathbb{N}$. According to its definition, we have $\mathrm{GCW}(\mathbf{s}^l) = l$, and similarly as above one may check $\mathbf{s}^l \in \mathfrak{S}(\mathbf{S}, \boldsymbol{\gamma})$.

Since Alg solves $\mathcal{P}_{\mathrm{GCW}}(\mathbf{S}, \boldsymbol{\gamma})$, it satisfies $\mathrm{Alg}(\mathbf{s}) = 1 \neq 2 = \mathrm{Alg}(\mathbf{s}^2)$. Regarding that $\mathbf{s}$ and $\mathbf{s}^2$ coincide on $\{t < B'\}$ and on $\{Q \in \mathcal{Q}_{\leq k} \,|\, 1 \notin Q \text{ or } 2 \notin Q\}$ in the sense of (1) and (2), Lemma B.2 (ii) assures the existence of some $Q_1 \in \mathcal{Q}_{\leq k}$ with $1 \in Q_1$ and $i_1 \coloneqq 2 \in Q_1$ such that $B_{Q_1}(\mathrm{Alg}, \mathbf{s}) \geq \min\{B_{Q_1}(\mathrm{Alg}, \mathbf{s}), B_{Q_1}(\mathrm{Alg}, \mathbf{s}^{i_1})\} \geq B'$. Let $F_1 \coloneqq [n] \setminus Q_1$ and fix an arbitrary $i_2 \in F_1$. Then, $\mathrm{Alg}(\mathbf{s}) = 1 \neq i_2 = \mathrm{Alg}(\mathbf{s}^{i_2})$ and since $\mathbf{s}$ and $\mathbf{s}^{i_2}$ coincide on $\{t < B'\}$ and $\{Q \in \mathcal{Q}_{\leq k} \,|\, i_2 \notin Q\}$, Lemma B.2 (ii) yields the existence of some $Q_2 \in \mathcal{Q}_{\leq k}$ with $i_2 \in Q_2$ such that $B_{Q_2}(\mathrm{Alg}, \mathbf{s}) \geq \min\{B_{Q_2}(\mathrm{Alg}, \mathbf{s}), B_{Q_2}(\mathrm{Alg}, \mathbf{s}^{i_2})\} \geq B'$. From $i_2 \in F_1 = [n] \setminus Q_1$ and $i_2 \in Q_2$ we infer $Q_1 \neq Q_2$. With this, we define $F_2 \coloneqq F_1 \setminus Q_2 = [n] \setminus (Q_1 \cup Q_2)$.

Inductively, whenever $F_l \neq \emptyset$, we may select an element $i_{l+1} \in F_l$ and infer from Lemma B.2 (ii), due to $\mathrm{Alg}(\mathbf{s}) = 1 \neq i_{l+1} = \mathrm{Alg}(\mathbf{s}^{i_{l+1}})$ and the similarity of $\mathbf{s}$ and $\mathbf{s}^{i_{l+1}}$ on $\{t < B'\}$ and $\{Q \in \mathcal{Q}_{\leq k} | i_{l+1} \notin Q\}$, the existence of a set $Q_{l+1} \in \mathcal{Q}_{\leq k}$ with $i_{l+1} \in Q_{l+1}$ such that $B_{Q_{l+1}}(\mathrm{Alg}, \mathbf{s}) \geq B'$, and define $F_{l+1} \coloneqq F_l \setminus Q_{l+1}$. Then, $i_{l+1} \in F_l = [n] \setminus (Q_1 \cup \cdots \cup Q_l)$

and $i_{l+1} \in Q_{l+1}$ assure $Q_{l+1} \notin \{Q_1, \dots, Q_l\}$. This procedure terminates at the smallest l' such that $F_{l'} = \emptyset$, and $Q_1, \dots, Q_{l'}$ are distinct. Regarding that $|F_{l+1}| - |F_l| \leq |Q_l| \leq k$ for all $l \in \{1, \dots, l'-1\}$, we have $l' \geq \lceil \frac{n}{k} \rceil$. Consequently,

$$B(\mathrm{Alg}, \mathbf{s}) \geq \sum_{l=1}^{l'} B_{Q_l}(\mathrm{Alg}, \mathbf{s}) \geq \left\lceil \frac{n}{k} \right\rceil B'$$

holds, which shows the claim for deterministic algorithms with regard to the definition of B'.

Part 2: The statement holds for arbitrary Alg**.**
Let $\mathfrak{A}$ be the set of all deterministic algorithms[3] and $\mathbf{s}$ be the instance from the first part. Write $\delta_{\mathbf{s}}$ for the probability distribution on $\{\mathbf{s}\}$, which assigns $\mathbf{s}$ probability one, i.e., the Dirac measure on $\mathbf{s}$. Note that for any randomized algorithm Alg there exists a probability distribution P on $\mathfrak{A}$ such that $\mathrm{Alg} \sim P$. By applying Yao's minimax principle [51] and using part one we conclude

$$\begin{aligned}\mathbb{E}[B(\mathrm{Alg}, \mathbf{s})] = \mathbb{E}_{\mathrm{Alg}' \sim P}[B(\mathrm{Alg}', \mathbf{s})] &\geq \inf_{\mathrm{Alg} \in \mathfrak{A}} \mathbb{E}_{\mathbf{s}' \sim \delta_{\mathbf{s}}}[B(\mathrm{Alg}, \mathbf{s}')] \\ &= \inf_{\mathrm{Alg} \in \mathfrak{A}} B(\mathrm{Alg}, \mathbf{s}) \geq \left\lceil \frac{n}{k} \right\rceil B',\end{aligned}$$

where B' is as in part one. □

Remark B.3. *(i) The above proof reveals even a stronger version of Theorem 3.1 (i). Indeed, in the proof we explicitly construct n distinct instances $\mathbf{s}^1 := \mathbf{s}, \dots, \mathbf{s}^n \in \mathfrak{S}(\mathbf{S}, \boldsymbol{\gamma})$ with $\mathrm{GCW}(\mathbf{s}^l) = l$ for all $l \in [n]$, and in fact show: Any (possibly random) algorithm* Alg*, which is able to correctly identify the best arm for any $\mathbf{s}' \in \{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ (i.e.,* Alg *does not necessarily have to solve $\mathcal{P}_{\mathrm{GCW}}(\mathbf{S}, \boldsymbol{\gamma})$) fulfills*

$$\mathbb{E}\left[B(\mathrm{Alg}, \mathbf{s})\right] \geq \left\lceil \frac{n}{k} \right\rceil \min_{Q \in \mathcal{Q}_{\leq k}} \min_{j \in Q} \gamma_{j|Q}^{-1} \left(\frac{S_{(1)|Q} - S_{(|Q|)|Q}}{2} \right).$$

(ii) Condition (iii) in the definition of $\mathfrak{S}(\mathbf{S}, \boldsymbol{\gamma})$ assures that the term $S_{(1)|Q}$ resp. $S_{(|Q|)|Q}$ in our lower bound from Theorem 3.1 coincides with $S'_{(1)|Q}$ resp. $S'_{(|Q|)|Q}$, when $S'_{i|Q} := \lim_{t \to \infty} s_{i|Q}(t)$ for $\mathbf{s} \in \mathfrak{S}(\mathbf{S}, \boldsymbol{\gamma})$.

B.2 Proof of Theorem 3.1 (ii): Lower Bound for GBW Identification

Recall that $\mathrm{GBW}(\mathbf{s})$ is the set of elements $i \in [n]$, for which the limits $S_{i|Q} = \lim_{t \to \infty} s_{i|Q}(t)$ have the highest Borda score

$$S_i^{\mathcal{B}} = \frac{\sum_{Q \in \mathcal{Q}_{=k}(i)} S_{i|Q}}{|\mathcal{Q}_{=k}(i)|} = \frac{\sum_{Q \in \mathcal{Q}_{=k}(i)} S_{i|Q}}{\binom{n-1}{k-1}}.$$

We call $\mathbf{S} = (S_{i|Q})_{Q \in \mathcal{Q}_{\leq k}, i \in Q}$ **homogeneous** if $(S_{(1)|Q}, \dots, S_{(|Q|)|Q})$ does not depend on Q. Thus, if $\mathbf{S}$ is homogeneous, we may simply write $S_{(l)}$ for $S_{(l)|Q}$ for any $Q \in \mathcal{Q}_{=k}$.

The next two lemmata serves as a preparation for the proof of (ii) and (iii) in Theorem 3.1.

Lemma B.4. *For any $\mathcal{W} \subseteq \mathcal{Q}_{=k}$ we have $\sum_{j=1}^n |\mathcal{Q}_{=k}(j) \cap \mathcal{W}| = k|\mathcal{W}|$.*

Proof of Lemma B.4. Let $\mathcal{W} \subseteq \mathcal{Q}_{=k}$ be fixed. For any $Q = \{i_1, \dots, i_k\} \in \mathcal{Q}_{=k} \cap \mathcal{W}$ we have that $Q \in \mathcal{Q}_{=k}(i_l) \cap \mathcal{W}$ for any $l \in [k]$, whereas $Q \notin \mathcal{Q}_{=k}(j) \cap \mathcal{W}$ for any $j \in [n] \setminus \{i_1, \dots, i_k\}$. Hence,

$$\sum_{j=1}^n |\mathcal{Q}_{=k}(j) \cap \mathcal{W}| = k \left| \bigcup_{j=1}^n (\mathcal{Q}_{=k}(j) \cap \mathcal{W}) \right| = k \left| \left(\bigcup_{j=1}^n \mathcal{Q}_{=k}(j) \right) \cap \mathcal{W} \right| = k|\mathcal{W}|.$$

□

Lemma B.5. *For any $\mathcal{W}' \subseteq \mathcal{Q}_{=k}$ and $\mathcal{W} := \mathcal{Q}_{=k} \setminus \mathcal{W}'$ with $|\mathcal{W}'| < \frac{(1-1/n)k}{k+n-2}\binom{n}{k}$ there exists $j \in [n] \setminus \{1\}$ with $|\mathcal{Q}_{=k}(j) \cap \mathcal{W}| > |\mathcal{Q}_{=k}(1) \cap \mathcal{W}'|$.*

[3] At any time $t \in \mathbb{N}$, a deterministic algorithm $\mathrm{Alg} \in \mathfrak{A}$ may either make a query $Q \in \mathcal{Q}_{\leq k}$ or terminate with a decision $\mathrm{X} \in \{1, \dots, n\}$. Thus, $\mathfrak{A}$ is a countable set.

Proof of Lemma B.5. For $j \in [n] \setminus \{1\}$ abbreviate $a_j := |\mathcal{Q}_{=k}(j) \cap \mathcal{W}| - |\mathcal{Q}_{=k}(1) \cap \mathcal{W}'|$. Due to

$$\begin{aligned}
|\mathcal{W}| &= \binom{n}{k} - |\mathcal{W}'| \\
&> \binom{n}{k} - \left(1 - \frac{1}{n}\right) \frac{k}{k+n-2} \binom{n}{k} \\
&= \binom{n}{k} - \frac{k\binom{n}{k} - \frac{k}{n}\binom{n}{k}}{k+n-2} \\
&= \frac{1}{k+n-2} \left(\binom{n-1}{k-1} + (n-2)\binom{n}{k} \right)
\end{aligned}$$

we have

$$k|\mathcal{W}| - \binom{n-1}{k-1} - (n-2)\left(\binom{n}{k} - |\mathcal{W}|\right) > 0.$$

By using Lemma B.4 and the fact that $(\mathcal{W} \cap \mathcal{Q}_{=k}(1)) \cup (\mathcal{W}' \cap \mathcal{Q}_{=k}(1)) = \mathcal{Q}_{=k}(1)$ is a disjoint union, we obtain

$$\begin{aligned}
&\sum\nolimits_{j \neq 1} a_j = \sum\nolimits_{j \neq 1} |\mathcal{Q}_{=k}(j) \cap \mathcal{W}| - (n-1)|\mathcal{Q}_{=k}(1) \cap \mathcal{W}'| \\
&= \sum\nolimits_{j \in [n]} |\mathcal{Q}_{=k}(j) \cap \mathcal{W}| - |\mathcal{Q}_{=k}(1) \cap \mathcal{W}| - |\mathcal{Q}_{=k}(1) \cap \mathcal{W}'| - (n-2)|\mathcal{Q}_{=k}(1) \cap \mathcal{W}'| \\
&= k|\mathcal{W}| - |\mathcal{Q}_{=k}(1)| - (n-2)|\mathcal{Q}_{=k}(1) \cap \mathcal{W}'| \\
&\geq k|\mathcal{W}| - \binom{n-1}{k-1} - (n-2)|\mathcal{W}'| \\
&= k|\mathcal{W}| - \binom{n-1}{k-1} - (n-2)\left(\binom{n}{k} - |\mathcal{W}|\right) > 0.
\end{aligned}$$

Consequently, there exists $j \in [n] \setminus \{1\}$ with $a_j > 0$. □

Proof of Theorem 3.1 (ii). Similarly as in the proof of Theorem 3.1 (i), we proceed in two steps.

Part 1: The statement holds in case Alg **is deterministic.**

Abbreviate $B' := \overline{\gamma}^{-1}\left(\frac{S_{(1)} - S_{(|Q|)}}{2}\right)$ and fix a family of permutations $(\pi_Q)_{Q \in \mathcal{Q}_{\leq k}}$ with $S_{(1)|Q} = S_{\pi_Q(1)|Q}$ for all $Q \in \mathcal{Q}_{\leq k}(1)$. Exactly as in the proof of Theorem 3.1 (i), we define $\mathbf{s} = (s_{i|Q}(t))_{Q \in \mathcal{Q}_{\leq k}, i \in Q, t \in \mathbb{N}}$ via

$$s_{i|Q}(t) := \begin{cases} \frac{S_{(1)|Q} + S_{(|Q|)|Q}}{2}, & \text{if } t < B' \\ S_{\pi_Q(i)|Q}, & \text{if } t \geq B'. \end{cases}$$

In the proof of Theorem 3.1 (i) we have already verified $\mathbf{s} \in \mathfrak{S}(\mathbf{S}, \boldsymbol{\gamma})$. For any $j \in \{2, \ldots, m\}$ and $Q \in \mathcal{Q}_{=k}(1) \cap \mathcal{Q}_{=k}(j)$ we have $S_{1|Q} > S_{j|Q}$, and using that $|\mathcal{Q}_{=k}(i') \setminus \mathcal{Q}_{=k}(j')|$ is the same for every distinct $i', j' \in [n]$ we thus have

$$\begin{aligned}
\sum\nolimits_{Q \in \mathcal{Q}_{=k}(1)} S_{1|Q} &= \sum\nolimits_{Q \in \mathcal{Q}_{=k}(1) \cap \mathcal{Q}_{=k}(j)} S_{1|Q} + S_{(1)} \cdot |\mathcal{Q}_{=k}(1) \setminus \mathcal{Q}_{=k}(j)| \\
&> \sum\nolimits_{Q \in \mathcal{Q}_{=k}(j) \cap \mathcal{Q}_{=k}(1)} S_{j|Q} + S_{(1)} \cdot |\mathcal{Q}_{=k}(j) \setminus \mathcal{Q}_{=k}(1)| \\
&> \sum\nolimits_{Q \in \mathcal{Q}_{=k}(j)} S_{j|Q}.
\end{aligned}$$

As $|\mathcal{Q}_{=k}(1)| = |\mathcal{Q}_{=k}(j)|$, this shows $\mathrm{GBW}(\mathbf{s}) = 1$.

In the following, we will show that

$$\mathcal{W}' := \{Q \in \mathcal{Q}_{=k} \,:\, \text{Alg started on } \mathbf{s} \text{ queries } Q \text{ at least } B' \text{ times}\}$$

contains at least $\frac{(1-1/n)k}{k+n-2}\binom{n}{k}$ elements. For this, let us assume on the contrary $|\mathcal{W}'| < \frac{(1-1/n)k}{k+n-2}\binom{n}{k}$ and write $\mathcal{W} := \mathcal{Q}_{=k} \setminus \mathcal{W}'$. Lemma B.5 allows us to fix a $j \in [n] \setminus \{1\}$ with $|\mathcal{Q}_{=k}(j) \cap \mathcal{W}| >$

$|\mathcal{Q}_{=k}(1) \cap \mathcal{W}'|$. Now, define $\mathbf{s}' = (s'_{i|Q}(t))_{Q \in \mathcal{Q}_{\leq k}, i \in Q, t \in \mathbb{N}}$ via $s'_{\cdot|Q}(\cdot) = s_{\cdot|Q}(\cdot)$ for any $Q \in (\mathcal{Q}_{\leq k} \setminus (\mathcal{Q}_{=k}(1) \cup \mathcal{Q}_{=k}(j))) \cup \mathcal{W}'$ and[4]

$$s'_{i|Q}(t) := \begin{cases} s_{i|Q}(t), & \text{if } t < B' \text{ or } \{1, j\} \not\subseteq Q, \\ S_{(1)}, & \text{if } i = j \in Q \text{ and } t \geq B', \\ S_{(|Q|)}, & \text{if } i = 1 \in Q \text{ and } t \geq B', \\ S_{1|Q}, & \text{if } t \geq B', i = \arg\min_{l' \in Q} S_{l'|Q} \text{ and } 1 \in Q \not\ni j, \\ S_{j|Q}, & \text{if } t \geq B', i = \operatorname{argmax}_{l' \in Q} S_{l'|Q} \text{ and } j \in Q \not\ni 1, \\ S_{i|Q}, & \text{otherwise,} \end{cases}$$

for $Q \in (\mathcal{Q}_{=k}(1) \cup \mathcal{Q}_{=k}(j)) \cap \mathcal{W}$. Similarly as for $\mathbf{s}$, we see $\mathbf{s}' \in \mathfrak{S}(\mathbf{S}, \boldsymbol{\gamma})$. The corresponding limit values $S'_{i|Q} = \lim_{t \to \infty} s'_{i|Q}(t)$ fulfill

$$\forall Q \in \mathcal{Q}_{=k}(1) \cap \mathcal{W} : S'_{1|Q} = S_{(|Q|)} \quad \text{and} \quad \forall Q \in \mathcal{Q}_{=k}(j) \cap \mathcal{W} : S'_{j|Q} = S_{(1)},$$

and trivially also $S_{(|Q|)} \leq S'_{i|Q} \leq S_{(1)}$ for any $Q \in \mathcal{Q}_{=k}, i \in Q$. Therefore, by choice of j, the corresponding Borda scores $(S')^{\mathcal{B}}_i$ for $\mathbf{s}'$ fulfill

$$\begin{aligned} \binom{n-1}{k-1} (S')^{\mathcal{B}}_1 &= \sum\nolimits_{Q \in \mathcal{Q}_{=k}(1)} S'_{1|Q} = \sum\nolimits_{Q \in \mathcal{Q}_{=k}(1) \cap \mathcal{W}'} S_{(1)} + \sum\nolimits_{Q \in \mathcal{Q}_{=k}(1) \cap \mathcal{W}} S_{(|Q|)} \\ &= |\mathcal{Q}_{=k}(1) \cap \mathcal{W}'| \cdot S_{(1)} + |\mathcal{Q}_{=k}(1) \cap \mathcal{W}| \cdot S_{(|Q|)} \\ &< |\mathcal{Q}_{=k}(j) \cap \mathcal{W}| \cdot S_{(1)} + |\mathcal{Q}_{=k}(j) \cap \mathcal{W}'| \cdot S_{(|Q|)} \\ &\leq \sum\nolimits_{Q \in \mathcal{Q}_{=k}(j)} S'_{j|Q} = \binom{n-1}{k-1} (S')^{\mathcal{B}}_j, \end{aligned}$$

where we have used that $|\mathcal{Q}_{=k}(1) \cap \mathcal{W}'| + |\mathcal{Q}_{=k}(1) \cap \mathcal{W}| = |\mathcal{Q}_{=k}(1)| = |\mathcal{Q}_{=k}(j) \cap \mathcal{W}'| + |\mathcal{Q}_{=k}(j) \cap \mathcal{W}|$. This show $1 \notin \mathrm{GBW}(\mathbf{s}')$. But since $s_{\cdot|\cdot}(\cdot) = s'_{\cdot|\cdot}(\cdot)$ holds on $\{t < B'\}$ as well as on $\mathcal{W}'$, Alg observes for $\mathbf{s}$ until termination exactly the same feedback as for $\mathbf{s}'$. Consequently, it outputs for both instances the same decision. Since $\mathrm{GBW}(\mathbf{s}) = 1 \notin \mathrm{GBW}(\mathbf{s}')$, it makes on at least one of the instances a mistake, which contradicts the correctness of Alg.
Thus, $|\mathcal{W}'| \geq \frac{(1-1/n)k}{k+n-2} \binom{n}{k}$ has to hold and we conclude

$$B(\mathrm{Alg}, \mathbf{s}) \geq \sum\nolimits_{Q \in \mathcal{W}'} B_Q(\mathrm{Alg}, \mathbf{s}) \geq |\mathcal{W}'| \cdot B' \geq \left(1 - \frac{1}{n}\right) \frac{k}{k+n-2} \binom{n}{k} B'.$$

Since $1 - \frac{1}{n} \geq 1/2$ and $k \leq n+2$ hold by assumption, we have in particular

$$B(\mathrm{Alg}, \mathbf{s}) \geq \frac{k}{4n} \binom{n}{k} \overline{\gamma}^{-1} \left(\frac{S_{(1)} - S_{(|Q|)}}{2} \right) = \frac{1}{4} \binom{n-1}{k-1} \overline{\gamma}^{-1} \left(\frac{S_{(1)} - S_{(|Q|)}}{2} \right) \in \Omega\left(\binom{n-1}{k-1} \right).$$

Part 2: The statement holds for arbitrary Alg**.**
Similarly as for the proof of (i) in Theorem 3.1, the proof follows by means of Yao's minimax principle. □

Remark B.6. *(i) To compare the bounds for* ROUNDROBIN *in Theorem C.1 with the lower bound from Theorem 3.1 (ii) suppose in the following* $\mathbf{S}$ *to be homogeneous with* $S_{(1)} > S_{(2)}$ *and let* $\boldsymbol{\gamma}$ *be homogeneous in the sense that* $\gamma_{i|Q}(t) = \gamma(t)$ *for all* $Q \in \mathcal{Q}_{=k}, i \in Q, t \in \mathbb{N}$ *for some* $\gamma : \mathbb{N} \to [0, \infty)$*. Moreover, let* $\mathbf{s}$ *be the instance from the proof of Theorem 3.1 (ii), and denote by* $\mathbf{S}$ *the family of limits* $S_{i|Q} = \lim_{t \to \infty} s_{i|Q}(t)$*,* $Q \in \mathcal{Q}_{\leq k}, i \in Q$*. Let us write* $S^{\mathcal{B}}_{(1)}, \ldots, S^{\mathcal{B}}_{(n)}$ *for the order statistics of* $\{S^{\mathcal{B}}_i\}_{i \in [n]}$*, i.e.,* $S^{\mathcal{B}}_{(1)} \geq \cdots \geq S^{\mathcal{B}}_{(n)}$*. Then,* ROUNDROBIN *returns a GBW of* $\mathbf{s} \in \mathfrak{S}(\mathbf{S}, \boldsymbol{\gamma})$ *if it is executed with a budget* B *at least*

$$z_{\mathrm{RR}} = \binom{n}{k} B_1 \quad \textit{with} \quad B_1 := \overline{\gamma}^{-1} \left(\frac{S^{\mathcal{B}}_{(1)} - S^{\mathcal{B}}_{(2)}}{2} \right).$$

[4] That is, for constructing $\mathbf{s}'$, we proceed for $Q \in \mathcal{W}$ as follows: If $\{1, j\} \subseteq Q$, we exchange $S_{1|Q}$ with $S_{j|Q}$. If $1 \in Q \not\ni j$, we exchange $S_{1|Q}$ with $S_{(Q)|Q}$. And if $j \in Q \not\ni 1$, we exchange $S_{j|Q}$ with $S_{(1)|Q}$.

In comparison to this, the lower bound just shown reveals that any (possibly deterministic) solution to $\mathcal{P}_{\mathrm{GBW}}(\mathbf{S},\boldsymbol{\gamma})$ fulfills

$$\mathbb{E}[B(\mathrm{Alg},\mathbf{s})] \geq \left(1-\frac{1}{n}\right)\frac{k}{k+n-2}\binom{n}{k}B_2 \quad \textit{with} \quad B_2 \coloneqq \overline{\gamma}^{-1}\left(\frac{S_{(1)}-S_{(|Q|)}}{2}\right).$$

Consequently, the optimality-gap between the upper and lower bound is of the order

$$B_1^{-1}B_2\left(1-\frac{1}{n}\right)\frac{k}{k+n-2}.$$

(ii) In the proof of Theorem 3.1 (ii), where we showed that $|\mathcal{W}'| \geq \frac{(1-1/n)k}{k+n-2}\binom{n}{k}$ leads to a contradiction, we have constructed an instance $\mathbf{s}' \in \mathfrak{S}(\mathbf{S},\boldsymbol{\gamma})$ with $\mathrm{GBW}(\mathbf{s}) = 1 \notin \mathrm{GBW}(\mathbf{s}')$ such that Alg observes on $\mathbf{s}$ the same feedback as on $\mathbf{s}'$. To finish the proof, we have only used that Alg is correct for $\mathbf{s}$ and for $\mathbf{s}'$, but we did not require correctness of Alg on any instance $\mathbf{s}'' \in \mathfrak{S}(\mathbf{S},\boldsymbol{\gamma}) \setminus \{\mathbf{s},\mathbf{s}'\}$. The construction of $\mathbf{s}'$ therein dependeds on the behaviour of Alg only by means of the choices of $\mathcal{W}$ and j in the proof, i.e., we have the dependence $\mathbf{s}' = \mathbf{s}'(\mathcal{W},j)$. Recall that for constructing $\mathbf{s}'$ we used that $|\mathcal{W}| = |\mathcal{Q}_{=k}| - |\mathcal{W}'| \geq \binom{n}{k} - \frac{(1-1/n)k}{k+n-2}\binom{n}{k}$, so that for $j \in [n]\setminus\{1\}$, the set

$$\left\{\mathbf{s}'(\mathcal{W},j)\,\middle|\,\mathcal{W} \subseteq \mathcal{Q}_{=k} \textit{ with } |\mathcal{W}| \geq \binom{n}{k} - \frac{(1-1/n)k}{k+n-2}\binom{n}{k} \textit{ and } j \in [n]\setminus\{1\}\right\}$$

of possible choices for $\mathbf{s}'$ has at most

$$N \coloneqq (n-1)\sum\nolimits_{l=\left\lceil\binom{n}{k}-\frac{(1-1/n)k}{k+n-2}\binom{n}{k}\right\rceil}^{\binom{n}{k}}\binom{\binom{n}{k}}{l}$$

elements, say $\mathbf{s}'_1,\ldots,\mathbf{s}'_N$. Thus, the formulation of the theorem may be strengthened in the following way:
If $\mathbf{S}$ is homogeneous and $\boldsymbol{\gamma}$ fixed, then there exist $N+1$ instances $\mathbf{s},\mathbf{s}'_1,\ldots,\mathbf{s}'_N$ with the following property: Whenever a (possibly probabilistic) sequential testing algorithm Alg correctly identifies the GBW for any of these $N+1$ instances, then

$$\mathbb{E}\left[B(\mathcal{A},\mathbf{s})\right] \geq \left(1-\frac{1}{n}\right)\frac{k}{k+n-2}\binom{n}{k}\overline{\gamma}^{-1}\left(\frac{S_{(1)}-S_{(|Q|)}}{2}\right).$$

B.3 Proof of Theorem 3.1 (iii): Lower Bound for GCopeW Identification

Recall that $\mathrm{GCopeW}(\mathbf{s})$ is the set of elements $i \in [n]$, for which the limits $S_{i|Q} = \lim_{t\to\infty} s_{i|Q}(t)$ have the highest Copeland score

$$S_i^{\mathcal{C}} = \frac{\sum_{Q\in\mathcal{Q}_{=k}(i)}\mathbf{1}\{S_{i|Q} = S_{(1)|Q}\}}{|\mathcal{Q}_{=k}(i)|} = \frac{\sum_{Q\in\mathcal{Q}_{=k}(i)}\mathbf{1}\{S_{i|Q} = S_{(1)|Q}\}}{\binom{n-1}{k-1}}.$$

Proof of Theorem 3.1.(iii). Similarly as in the proofs (i) and (ii) Theorem 3.1, we proceed in two steps.

Part 1: The statement holds in case Alg is deterministic.
Abbreviate $B' \coloneqq \min_{Q\in\mathcal{Q}_{\leq k}}\min_{i\in Q}\ \gamma_{i|Q}^{-1}\left(\frac{S_{(1)|Q}-S_{(|Q|)|Q}}{2}\right)$ and fix a family of permutations $(\pi_Q)_{Q\in\mathcal{Q}_{\leq k}}$ with $S_{(1)|Q} = S_{\pi_Q(1)|Q}$ for all $Q \in \mathcal{Q}_{\leq k}(1)$. Exactly as in the proofs of the lower bounds for GCW and GBW identification, we define $\mathbf{s} = (s_{i|Q}(t))_{Q\in\mathcal{Q}_{\leq k},i\in Q,t\in\mathbb{N}}$ via

$$s_{i|Q}(t) \coloneqq \begin{cases}\frac{S_{(1)|Q}+S_{(|Q|)|Q}}{2}, & \text{if } t < B'\\ S_{\pi_Q(i)|Q}, & \text{if } t \geq B'.\end{cases}$$

In the proof of the lower bound of GCW identification we have already verified $\mathbf{s} \in \mathfrak{S}(\mathbf{S},\boldsymbol{\gamma})$. For any $j \in \{2,\ldots,m\}$ and $Q \in \mathcal{Q}_{=k}(1)\cap\mathcal{Q}_{=k}(j)$ we have $S_{1|Q} > S_{j|Q}$, and using that $|\mathcal{Q}_{=k}(i')\setminus\mathcal{Q}_{=k}(j')|$

is the same for every distinct $i', j' \in [n]$ we thus have

$$\begin{aligned}
&\sum_{Q \in \mathcal{Q}_{=k}(1)} \mathbb{1}\{S_{1|Q} = S_{(1)|Q}\} \\
&\quad = \sum_{Q \in \mathcal{Q}_{=k}(1) \cap \mathcal{Q}_{=k}(j)} \mathbb{1}\{S_{1|Q} = S_{(1)|Q}\} + \sum_{Q \in \mathcal{Q}_{=k}(1) \setminus \mathcal{Q}_{=k}(j)} \mathbb{1}\{S_{1|Q} = S_{(1)|Q}\} \\
&\quad = \sum_{Q \in \mathcal{Q}_{=k}(1) \cap \mathcal{Q}_{=k}(j)} \mathbb{1}\{S_{1|Q} = S_{(1)|Q}\} + |\mathcal{Q}_{=k}(1) \setminus \mathcal{Q}_{=k}(j)| \\
&\quad > \sum_{Q \in \mathcal{Q}_{=k}(1) \cap \mathcal{Q}_{=k}(j)} \mathbb{1}\{S_{j|Q} = S_{(1)|Q}\} + |\mathcal{Q}_{=k}(j) \setminus \mathcal{Q}_{=k}(1)| \\
&\quad \geq \sum_{Q \in \mathcal{Q}_{=k}(1) \cap \mathcal{Q}_{=k}(j)} \mathbb{1}\{S_{j|Q} = S_{(1)|Q}\} + \sum_{Q \in \mathcal{Q}_{=k}(j) \setminus \mathcal{Q}_{=k}(1)} \mathbb{1}\{S_{j|Q} = S_{(1)|Q}\} \\
&\quad = \sum_{Q \in \mathcal{Q}_{=k}(j)} \mathbb{1}\{S_{j|Q} = S_{(1)|Q}\}.
\end{aligned}$$

As $|\mathcal{Q}_{=k}(1)| = |\mathcal{Q}_{=k}(j)|$, this shows $\mathrm{GCopeW}(\mathbf{s}) = 1$.
Similarly as in the proof of (ii), we will show indirectly that

$$\mathcal{W}' \coloneqq \{Q \in \mathcal{Q}_{=k} \,:\, \text{Alg started on } \mathbf{s} \text{ queries } Q \text{ at least } B' \text{ times}\}$$

contains at least $\frac{(1-1/n)k}{k+n-2}\binom{n}{k}$ elements. For this purpose, let us assume on the contrary $|\mathcal{W}'| < \frac{(1-1/n)k}{k+n-2}\binom{n}{k}$ and write $\mathcal{W} \coloneqq \mathcal{Q}_{=k} \setminus \mathcal{W}'$. Lemma B.5 allows us to fix a $j \in [n] \setminus \{1\}$ with $|\mathcal{Q}_{=k}(j) \cap \mathcal{W}| > |\mathcal{Q}_{=k}(1) \cap \mathcal{W}'|$. Now, define $\mathbf{s}' = (s'_{i|Q}(t))_{Q \in \mathcal{Q}_{\leq k}, i \in Q, t \in \mathbb{N}}$ analogously as in the proof of (ii), i.e., via $s'_{\cdot|Q}(\cdot) = s_{\cdot|Q}(\cdot)$ for any $Q \in (\mathcal{Q}_{\leq k} \setminus (\mathcal{Q}_{=k}(1) \cup \mathcal{Q}_{=k}(j))) \cup \mathcal{W}'$ and

$$s'_{i|Q}(t) \coloneqq \begin{cases} s_{i|Q}(t), & \text{if } t < B' \text{ or } \{1, j\} \not\subseteq Q, \\ S_{(1)|Q}, & \text{if } i = j \in Q \text{ and } t \geq B', \\ S_{(|Q|)|Q}, & \text{if } i = 1 \in Q \text{ and } t \geq B', \\ S_{1|Q}, & \text{if } t \geq B', i = \arg\min_{l' \in Q} S_{l'|Q} \text{ and } 1 \in Q \not\ni j, \\ S_{j|Q}, & \text{if } t \geq B', i = \mathrm{argmax}_{l' \in Q} S_{l'|Q} \text{ and } j \in Q \not\ni 1, \\ S_{i|Q}, & \text{otherwise,} \end{cases}$$

for $Q \in (\mathcal{Q}_{=k}(1) \cup \mathcal{Q}_{=k}(j)) \cap \mathcal{W}$. Similarly as for $\mathbf{s}$, we see $\mathbf{s}' \in \mathfrak{S}(\mathbf{S}, \boldsymbol{\gamma})$. The corresponding limit values $S'_{i|Q} = \lim_{t \to \infty} s'_{i|Q}(t)$ fulfill

$$\forall Q \in \mathcal{Q}_{=k}(1) \cap \mathcal{W} : S'_{1|Q} = S_{(|Q|)|Q} \quad \text{and} \quad \forall Q \in \mathcal{Q}_{=k}(j) \cap \mathcal{W} : S'_{j|Q} = S_{(1)|Q},$$

and trivially also $S_{(|Q|)|Q} \leq S'_{i|Q} \leq S_{(1)|Q}$ for any $Q \in \mathcal{Q}_{=k}, i \in Q$. Therefore, by choice of j, the corresponding Copeland scores $(S')^{\mathcal{C}}_i$ for $\mathbf{s}'$ fulfill

$$\begin{aligned}
\binom{n-1}{k-1}(S')^{\mathcal{C}}_1 &= \sum\nolimits_{Q\in\mathcal{Q}_{=k}(1)} \mathbb{1}\{S'_{1|Q} = S'_{(1)|Q}\} \\
&= \sum_{Q\in\mathcal{Q}_{=k}(1)\cap\mathcal{W}'} \mathbb{1}\{S'_{1|Q} = S'_{(1)|Q}\} + \sum_{Q\in\mathcal{Q}_{=k}(1)\cap\mathcal{W}} \mathbb{1}\{S'_{1|Q} = S'_{(1)|Q}\} \\
&= \sum_{Q\in\mathcal{Q}_{=k}(1)\cap\mathcal{W}'} \mathbb{1}\{S_{1|Q} = S_{(1)|Q}\} + \sum_{Q\in\mathcal{Q}_{=k}(1)\cap\mathcal{W}} \mathbb{1}\{S_{(|Q|)|Q} = S_{(1)|Q}\} \\
&= |\mathcal{Q}_{=k}(1)\cap\mathcal{W}'| \\
&< |\mathcal{Q}_{=k}(j)\cap\mathcal{W}| \\
&= \sum_{\mathcal{Q}_{=k}(j)\cap\mathcal{W}} \mathbb{1}\{S_{(1)|Q} = S_{(1)|Q}\} \\
&= \sum_{\mathcal{Q}_{=k}(j)\cap\mathcal{W}} \mathbb{1}\{S'_{j|Q} = S'_{(1)|Q}\} \\
&\leq \sum_{\mathcal{Q}_{=k}(j)\cap\mathcal{W}} \mathbb{1}\{S'_{j|Q} = S'_{(1)|Q}\} + \sum_{\mathcal{Q}_{=k}(j)\cap\mathcal{W}'} \mathbb{1}\{S'_{j|Q} = S'_{(1)|Q}\} \\
&= \sum_{Q\in\mathcal{Q}_{=k}(j)} \mathbb{1}\{S'_{j|Q} = S'_{(1)|Q}\} \\
&= \binom{n-1}{k-1}(S')^{\mathcal{C}}_j,
\end{aligned}$$

where we used that $S_{(1)|Q} = S'_{(1)|Q}$. This shows $1 \notin \mathrm{GCopeW}(\mathbf{s}')$. But since $s_{\cdot|\cdot}(\cdot) = s'_{\cdot|\cdot}(\cdot)$ holds on $\{t < B'\}$ as well as on $\mathcal{W}'$, Alg observes for $\mathbf{s}$ until termination exactly the same feedback as for $\mathbf{s}'$. Consequently, it outputs for both instances the same decision. Since $\mathrm{GCopeW}(\mathbf{s}) = 1 \notin \mathrm{GCopeW}(\mathbf{s}')$, it makes on at least one of the instances a mistake, which contradicts the correctness of Alg.
Thus, $|\mathcal{W}'| \geq \frac{(1-1/n)k}{k+n-2}\binom{n}{k}$ has to hold and we conclude

$$B(\mathrm{Alg}, \mathbf{s}) \geq \sum\nolimits_{Q\in\mathcal{W}'} B_Q(\mathrm{Alg}, \mathbf{s}) \geq |\mathcal{W}'| \cdot B' \geq \left(1 - \frac{1}{n}\right)\frac{k}{k+n-2}\binom{n}{k}B'.$$

Since $1 - \frac{1}{n} \geq 1/2$ and $k \leq n+2$ hold by assumption, we have in particular

$$\begin{aligned}
B(\mathrm{Alg}, \mathbf{s}) &\geq \frac{k}{4n}\binom{n}{k} \min_{Q\in\mathcal{Q}_{\leq k}} \min_{i\in Q} \gamma^{-1}_{i|Q}\left(\frac{S_{(1)|Q} - S_{(|Q|)|Q}}{2}\right) \\
&= \frac{1}{4}\binom{n-1}{k-1} \min_{Q\in\mathcal{Q}_{\leq k}} \min_{i\in Q} \gamma^{-1}_{i|Q}\left(\frac{S_{(1)|Q} - S_{(|Q|)|Q}}{2}\right) \in \Omega\left(\binom{n-1}{k-1}\right).
\end{aligned}$$

Part 2: The statement holds for arbitrary Alg**.**
Similarly as for the proofs of the lower bound of (i) and (ii) of this theorem, the proof follows by means of Yao's minimax principle. □

C Generalized Borda Winner Identification

Let ROUNDROBIN be the algorithm, which enumerates all possible subsets of the fixed subset size k, chooses each subset in a round-robin fashion and returns the arm with the highest empirical Borda score $s^{\mathcal{B}}_i$ after the available budget is exhausted. It is a straightforward baseline method, which we analyze theoretically in terms of the sufficient and necessary budget to return a generalized Borda winner (GBW) $i^*_{\mathcal{B}}$. For this purpose, let $\hat{\gamma}_i(t) = \frac{1}{|\mathcal{Q}_{=k}(i)|}\sum_{Q\in\mathcal{Q}_{=k}(i)} \gamma_{i|Q}(t)$ and $\hat{\gamma}^{\max}_{i,j}(t) = \max\{\hat{\gamma}_i(t), \hat{\gamma}_j(t)\}$.

Theorem C.1. ROUNDROBIN *returns* $i^*_{\mathcal{B}}$ *if it is executed with a budget* $B \geq z_{\mathrm{RR}}$*, where*

$$z_{\mathrm{RR}} := \binom{n}{k} \max_{\rho \in \mathcal{A}, \rho \neq i^*_{\mathcal{B}}} \left(\hat{\gamma}^{\max}_{i^*_{\mathcal{B}},\rho}\right)^{-1} \left(\frac{S^{\mathcal{B}}_{i^*_{\mathcal{B}}} - S^{\mathcal{B}}_{\rho}}{2}\right).$$

The latter bound is tight in a worst-case scenario, as the following result shows (cf. Sec. D.1 for the proofs).

Theorem C.2. *For any asymptotical Borda scores* $S^{\mathcal{B}}_1, \dots, S^{\mathcal{B}}_n$*, there exists a corresponding instance* $\mathbf{s}$ *such that if* $B < z_{\mathrm{RR}}$ *then* ROUNDROBIN *will not return* $i^*_{\mathcal{B}}$*.*

Thus, ROUNDROBIN is already nearly-optimal (up to a factor $\mathcal{O}(n/k)$) with respect to worst-case scenarios due to Theorem 3.1 (see Rem. B.6 for a more detailed discussion.).

D Proofs of Section 4

In this section we provide the detailed proofs of Section 4. We assume throughout that $\frac{B}{\binom{n}{k}}$ is a natural number, i.e., the budget is a multiple of $\binom{n}{k.}$

D.1 Proof of Theorems C.1 and C.2

Proof of Theorem C.1. After relabeling the arms in round r we may assume w.l.o.g. $i^*_{\mathcal{B}} = 1$. We will prove the theorem by contradiction and therefore assume

$$\begin{aligned}
&\rho = \mathrm{argmax}_{i \in \mathcal{A}}\, s^{\mathcal{B}}_i \left(\frac{B}{\binom{n}{k}}\right) \neq 1 \\
&\Rightarrow s^{\mathcal{B}}_1 \left(\frac{B}{\binom{n}{k}}\right) < \max_{j=2,\dots n} s^{\mathcal{B}}_j \left(\frac{B}{\binom{n}{k}}\right) = s^{\mathcal{B}}_{\rho} \left(\frac{B}{\binom{n}{k}}\right) \\
&\Rightarrow S^{\mathcal{B}}_1 - S^{\mathcal{B}}_{\rho} < s^{B}_{\rho} \left(\frac{B}{\binom{n}{k}}\right) - S^{\mathcal{B}}_{\rho} + S^{\mathcal{B}}_1 - s^{\mathcal{B}}_1 \left(\frac{B}{\binom{n}{k}}\right) \\
&\quad = \frac{1}{|\mathcal{Q}_{=k}(\rho)|} \sum_{Q \in \mathcal{Q}_{=k}(\rho)} \left(s_{\rho|Q}\left(\frac{B}{\binom{n}{k}}\right) - S_{\rho|Q}\right) + \frac{1}{|\mathcal{Q}_{=k}(1)|} \sum_{Q \in \mathcal{Q}_{=k}(1)} \left(S_{1|Q} - s_{1|Q}\left(\frac{B}{\binom{n}{k}}\right)\right) \\
&\Rightarrow S^{\mathcal{B}}_1 - S^{\mathcal{B}}_{\rho} < \hat{\gamma}_{\rho}\left(\frac{B}{\binom{n}{k}}\right) + \hat{\gamma}_1\left(\frac{B}{\binom{n}{k}}\right) \\
&\Rightarrow S^{\mathcal{B}}_1 - S^{\mathcal{B}}_{\rho} < 2 \cdot \hat{\gamma}^{\max}_{1,\rho}\left(\frac{B}{\binom{n}{k}}\right),
\end{aligned}$$

where $\hat{\gamma}_i(t) = \frac{1}{|\mathcal{Q}_{=k}(i)|} \sum_{Q \in \mathcal{Q}_{=k}(i)} \gamma_{i|Q}(t)$ and $\hat{\gamma}^{\max}_{i,j}(t) = \max\{\hat{\gamma}_i(t), \hat{\gamma}_j(t)\}$. With this, however, we can derive

$$\Rightarrow\ z_{RR} = \left(\hat{\gamma}^{\max}_{1,\rho}\right)^{-1} \left(\frac{S^{\mathcal{B}}_1 - S^{\mathcal{B}}_{\rho}}{2}\right) \binom{n}{k} \geq B,$$

which contradicts the assumption we make on the budget B. Thus, it holds that the returned arm is $\rho = 1$. □

Proof of Theorem C.2. Let $\beta(t)$ be an arbitrary, monotonically decreasing function of t with $\lim_{t\to\infty} \beta(t) = 0$. We define for all $j \in \mathcal{A}$ with $j \neq i^*_{\mathcal{B}}$ the empirical Borda scores to be $s^{\mathcal{B}}_j(t) = S^{\mathcal{B}}_j + \beta(t)$ and $s^{\mathcal{B}}_{i^*_{\mathcal{B}}}(t) = S^{\mathcal{B}}_{i^*_{\mathcal{B}}} - \beta(t)$, where $(S^{\mathcal{B}}_i)_{i \in [n]}$ are arbitrary real values such that $S^{\mathcal{B}}_{i^*_{\mathcal{B}}}$ is the unique maximum for some $i^*_{\mathcal{B}} \in [n]$. We can again assume after relabeling all arms

that w.l.o.g. that $i^*_{\mathcal{B}} = 1$ and $\operatorname{argmax}_{j=2,\dots,n} S^{\mathcal{B}}_j = 2$. Note that $\hat{\gamma}_i(t) = \beta(t)$ for all $i \in \mathcal{A}$. In light of these considerations, ROUNDROBIN returns 1 as the best arm if and only if

$$
\begin{aligned}
s^{\mathcal{B}}_1\left(\frac{B}{\binom{n}{k}}\right) > \max_{j=2,\dots,n} s^{\mathcal{B}}_j\left(\frac{B}{\binom{n}{k}}\right) \quad & \Leftrightarrow S^{\mathcal{B}}_1 - \hat{\gamma}_1\left(\frac{B}{\binom{n}{k}}\right) > \max_{j=2,\dots,n} S^{\mathcal{B}}_j + \hat{\gamma}_j\left(\frac{B}{\binom{n}{k}}\right) \\
& \Leftrightarrow S^{\mathcal{B}}_1 - \hat{\gamma}_1\left(\frac{B}{\binom{n}{k}}\right) > S^{\mathcal{B}}_2 + \hat{\gamma}_2\left(\frac{B}{\binom{n}{k}}\right) \\
& \Leftrightarrow \hat{\gamma}_1\left(\frac{B}{\binom{n}{k}}\right) + \hat{\gamma}_2\left(\frac{B}{\binom{n}{k}}\right) < S^{\mathcal{B}}_1 - S^{\mathcal{B}}_2 \\
& \Leftrightarrow 2 \cdot \hat{\gamma}^{\max}_{1,2}\left(\frac{B}{\binom{n}{k}}\right) < S^{\mathcal{B}}_1 - S^{\mathcal{B}}_2 \\
& \Leftrightarrow B \geq \binom{n}{k}\left(\hat{\gamma}^{\max}_{1,2}\right)^{-1}\left(\frac{S^{\mathcal{B}}_1 - S^{\mathcal{B}}_2}{2}\right).
\end{aligned}
$$

Thus, the necessary budget is z_{RR} in this case concluding the claim. □

D.2 Proofs of Theorem 4.1 and 4.2

Proof of Theorem 4.1. For the sake of convenience, let us abbreviate $[R] \coloneqq \{1, \dots, R\}$ and $\mathbb{A}_{rj} \coloneqq \mathbb{A}_{r,j}$ in the following. By possibly relabeling the arms and query sets queried by the algorithm, we can assume w.l.o.g. $i^* = 1$ and $\mathbb{A}_r(1) = \mathbb{A}_{r1}$ for all $r \in [R]$ in the following. In particular, we have $S_{1|\mathbb{A}_{r1}} = S_{(1)|\mathbb{A}_{r1}}$ for all $r \in [R]$. We prove the correctness of the algorithm indirectly. Thus, we start by assuming that the best arm is not contained in the last partition (i.e., the remaining active arm):

$$
\begin{aligned}
& \mathbb{A}_{R+1} \neq \{1\} \\
\Leftrightarrow & \exists r \in [R] : 1 \notin \mathbb{A}_{r+1} \wedge 1 \in \mathbb{A}_r \\
\Rightarrow & \exists r \in [R] : \sum_{i \in \mathbb{A}_{r1}} \mathbf{1}\{s_{i|\mathbb{A}_{r1}}(b_r) \geq s_{1|\mathbb{A}_{r1}}(b_r)\} > f(|\mathbb{A}_{r1}|) \\
\Rightarrow & \exists r \in [R] : \sum_{i \in \mathbb{A}_{r1}} \mathbf{1}\{S_{1|\mathbb{A}_{r1}} - S_{i|\mathbb{A}_{r1}} \leq S_{1|\mathbb{A}_{r1}} - s_{1|\mathbb{A}_{r1}}(b_r) - S_{i|\mathbb{A}_{r1}} + s_{i|\mathbb{A}_{r1}}(b_r)\} > f(|\mathbb{A}_{r1}|) \\
\Rightarrow & \exists r \in [R] : \sum_{i \in \mathbb{A}_{r1}} \mathbf{1}\{S_{1|\mathbb{A}_{r1}} - S_{i|\mathbb{A}_{r1}} \leq |S_{1|\mathbb{A}_{r1}} - s_{1|\mathbb{A}_{r1}}(b_r)| + |S_{i|\mathbb{A}_{r1}} - s_{i|\mathbb{A}_{r1}}(b_r)|\} > f(|\mathbb{A}_{r1}|) \\
\Rightarrow & \exists r \in [R] : \sum_{i \in \mathbb{A}_{r1}} \mathbf{1}\{S_{1|\mathbb{A}_{r1}} - S_{i|\mathbb{A}_{r1}} \leq 2\bar{\gamma}_{\mathbb{A}_{r1}}(b_r)\} > f(|\mathbb{A}_{r1}|) \\
\Rightarrow & \exists r \in [R] : S_{1|\mathbb{A}_{r1}} - S_{(f(\mathbb{A}_{r1})+1)|\mathbb{A}_{r1}} \leq 2\bar{\gamma}_{\mathbb{A}_{r1}}(b_r) \\
\Rightarrow & \exists r \in [R] : \left\lfloor \frac{B}{P_r R} \right\rfloor = b_r \leq \bar{\gamma}^{-1}_{\mathbb{A}_{r1}}\left(\frac{S_{1|\mathbb{A}_{r1}} - S_{(f(\mathbb{A}_{r1})+1)|\mathbb{A}_{r1}}}{2}\right) \\
\Rightarrow & \exists r \in [R] : B \leq P_r R \left\lceil \bar{\gamma}^{-1}_{\mathbb{A}_{r1}}\left(\frac{S_{1|\mathbb{A}_{r1}} - S_{(f(|\mathbb{A}_{r1}|)+1)|\mathbb{A}_{r1}}}{2}\right)\right\rceil \\
\Rightarrow & B \leq R \max_{r \in [R]} P_r \left\lceil \bar{\gamma}^{-1}_{\mathbb{A}_{r1}}\left(\frac{S_{1|\mathbb{A}_{r1}} - S_{(f(|\mathbb{A}_{r1}|)+1)|\mathbb{A}_{r1}}}{2}\right)\right\rceil = z\left(f, R, \{P_r\}_{1 \leq r \leq R}\right),
\end{aligned}
$$

which contradicts the assumption we make on the budget B. Thus, it holds that the remaining active arm in round $R+1$ is $i^* = 1$. □

Remark D.1. *Using the definition of $\bar{\gamma}_Q(t)$ and $\bar{\gamma}(t)$ we can derive the following more coarser bounds on the sufficient budget:*

$$
\begin{aligned}
z_1\left(f, R, \{P_r\}_{1 \leq r \leq R}\right) &= R \max_{r \in [R]} P_r \left\lceil \bar{\gamma}^{-1}\left(\frac{S_{1|\mathbb{A}_{r1}} - S_{(f(|\mathbb{A}_{r1}|)+1)|\mathbb{A}_{r1}}}{2}\right)\right\rceil, \\
z_2\left(f, R, \{P_r\}_{1 \leq r \leq R}\right) &= R\left(\max_{r \in [R]} P_r\right) \max_{Q \in \mathcal{Q}_{\leq k}} \left\lceil \bar{\gamma}^{-1}\left(\frac{S_{1|Q} - S_{(f(|Q|)+1)|Q}}{2}\right)\right\rceil.
\end{aligned}
$$

Proof of Theorem 4.2. After relabeling, we may suppose w.l.o.g. $i^* = 1$. Let $\beta : \mathbb{N} \to (0, \infty)$ be an arbitrary strictly decreasing function with $\beta(t) \to 0$ as $t \to \infty$ and

$$\left\{ \frac{S_{1|Q} - S_{j|Q}}{2} : Q \in \mathcal{Q}_{\leq k}, j \in Q \right\} \subseteq \beta(\mathbb{N}).$$

Then, β is invertible on $\beta(\mathbb{N})$ and its inverse function $\beta^{-1} : \beta(\mathbb{N}) \to \mathbb{N}$ trivially fulfills $\beta^{-1}(\alpha) = \min\{t \in \mathbb{N} : \beta(t) \leq \alpha\}$ for all $\alpha \in \beta(\mathbb{N})$. Define for any $Q \in \mathcal{Q}_{\leq k}$ and $i \in Q$ the family of statistics by means of

$$s_{i|Q}(t) := \begin{cases} S_{i|Q} - \beta(t), & \text{if } i = \operatorname{argmax}_{j \in Q} S_{j|Q}, \\ S_{i|Q} + \beta(t), & \text{otherwise,} \end{cases}$$

and note that $\overline{\gamma}_Q(t) = \beta(t)$ for all $Q \in \mathcal{Q}_{\leq k}$ and $t \in \mathbb{N}$. Writing $b_r = \left\lfloor \frac{B}{RP_r} \right\rfloor$ we obtain due to the choice of β that

$$\begin{aligned}
&B < R \max_{r \in [R]} P_r \bar{\gamma}^{-1}_{\mathbb{A}_{r1}} \left(\frac{S_{1|\mathbb{A}_{r1}} - S_{(f(|\mathbb{A}_{r1}|)+1)|\mathbb{A}_{r1}}}{2} \right) \\
\Rightarrow &\exists r \in [R] : B < RP_r \min \left\{ t \in \mathbb{N} : \bar{\gamma}_{\mathbb{A}_{r1}}(t) \leq \frac{S_{1|\mathbb{A}_{r1}} - S_{(f(|\mathbb{A}_{r1}|)+1)|\mathbb{A}_{r1}}}{2} \right\} \\
\Rightarrow &\exists r \in [R] : b_r < \min \left\{ t \in \mathbb{N} : \beta(t) \leq \frac{S_{1|\mathbb{A}_{r1}} - S_{(f(|\mathbb{A}_{r1}|)+1)|\mathbb{A}_{r1}}}{2} \right\} = \beta^{-1} \left(\frac{S_{1|\mathbb{A}_{r1}} - S_{(f(|\mathbb{A}_{r1}|)+1)|\mathbb{A}_{r1}}}{2} \right) \\
\Rightarrow &\exists r \in [R] : 2\beta(b_r) > S_{1|\mathbb{A}_{r1}} - S_{(f(|\mathbb{A}_{r1}|)+1)|\mathbb{A}_{r1}} = s_{1|\mathbb{A}_{r1}}(b_r) + \beta(b_r) - \left(s_{(f(|\mathbb{A}_{r1}|)+1)|\mathbb{A}_{r1}}(b_r) - \beta(b_r) \right) \\
\Rightarrow &\exists r \in [R] : s_{1|\mathbb{A}_{r1}}(b_r) < s_{(f(|\mathbb{A}_{r1}|)+1)|\mathbb{A}_{r1}}(b_r) \\
\Rightarrow &\exists r \in [R] : 1 \notin \mathbb{A}_{r+1} \\
\Rightarrow &1 \notin \mathbb{A}_{R+1}.
\end{aligned}$$

This shows that $z(f, R, \{P_r\}_{1 \leq r \leq R})$ is the necessary budget for returning the best arm i^* in this scenario. □

D.3 Proof of Corollary 4.3

For sake of convenience, we provide the entire pseudo-code of CSWS in Algorithm 3, which results by using $f(x) = x - 1$ as well as P_r^{CSWS} and R^{CSWS} as defined in Section 4.1 in Algorithm 1.

Proof of Corollary 4.3 (CSWS case). Suppose $B > 0$ to be arbitrary but fixed. First, note that there are at most $\lceil \log_k(n) \rceil$ rounds within the first while-loop and at most 1 in the second, so that we have at most $\lceil \log_k(n) \rceil + 1$ many rounds in total. The total number of partitions in round $r \in \{1, \ldots, \lceil \log_k(n) \rceil + 1\}$ is at most $\left\lceil \frac{n}{k^r} \right\rceil$. Abbreviating $R := R^{\text{CSWS}}$ and $P_r := P_r^{\text{CSWS}}$ for the moment, the budget allocated to a partition in round r is by definition $b_r = \lfloor \frac{B}{RP_r} \rfloor = \left\lfloor \frac{B}{\lceil \frac{n}{k^r} \rceil \cdot (\lceil \log_k(n) \rceil + 1)} \right\rfloor$. Hence, the total budget used by CSWS is

$$\sum_{r=1}^{\lceil \log_k(n) \rceil + 1} \#\{\text{partitions in round r}\} \cdot b_r = \sum_{r=1}^{\lceil \log_k(n) \rceil + 1} \left\lceil \frac{n}{k^r} \right\rceil \left\lfloor \frac{B}{\lceil \frac{n}{k^r} \rceil \cdot (\lceil \log_k(n) \rceil + 1)} \right\rfloor \leq B.$$

Thus, the stated correctness of CSWS follows directly from Theorem 4.1.

□

For sake of convenience, we provide the entire pseudo-code of CSR in Algorithm 4, which results by using $f(x) = 1$ as well as P_r^{CSR} and R^{CSR} as defined in Section 4.1 in Algorithm 1.

Proof of Corollary 4.3 (CSR case). Suppose $B > 0$ to be arbitrary but fixed. First, note that there are at most $\left\lceil \log_{1-\frac{1}{k}} \left(\frac{1}{n} \right) \right\rceil$ rounds within the first while-loop and at most $k - 1$ in the second, so that we have at most $\left\lceil \log_{1-\frac{1}{k}} \left(\frac{1}{n} \right) \right\rceil + k - 1$ many rounds in total. The total number of partitions in round

Algorithm 3 Combinatorial Successive Winner Stays (CSWS)

Input: set of arms $[n]$, subset size $k \leq n$, sampling budget B

Initialization: For each $r \in \{1, \dots, \lceil \log_k(n) \rceil + 1\}$ let $b_r := \left\lfloor \frac{B}{\lceil \frac{n}{k^r} \rceil \cdot (\lceil \log_k(n) \rceil + 1)} \right\rfloor$, $\mathbb{A} \leftarrow [n]$, $r \leftarrow 1$

while $|\mathbb{A}_r| \geq k$ **do**
  $J = \lceil \frac{n}{k^r} \rceil$
  $\mathbb{A}_{r1}, \mathbb{A}_{r2}, \dots, \mathbb{A}_{r,J} \leftarrow Partition(\mathbb{A}_r, k)$
  if $|\mathbb{A}_{r,J}| < k$ **then**
    $\mathcal{R} \leftarrow \mathbb{A}_{r,J}$, $J \leftarrow J - 1$
  else
    $\mathcal{R} \leftarrow \emptyset$
  end if
  $\mathbb{A}_{r+1} \leftarrow \emptyset$
  for $j \in [J]$ **do**
    Play the set $\mathbb{A}_{r,j}$ for b_r times
    For all $i \in \mathbb{A}_{r,j}$, update $s_{i|\mathbb{A}_{r,j}}(b_r)$
    Let $w \in \text{argmax}_i \; s_{i|\mathbb{A}_{r,j}}(b_r)$
    $\mathbb{A}_{r+1} \leftarrow \mathbb{A}_{r+1} \cup \{w\}$
  end for
  $\mathbb{A}_{r+1} \leftarrow \mathbb{A}_{r+1} \cup \mathcal{R}$
  $r \leftarrow r + 1$
end while
$\mathbb{A}_{r+1} \leftarrow \emptyset$
while $|\mathbb{A}_r| > 1$ **do**
  Play the set $\mathbb{A}_r$ for b_r times
  For all $i \in \mathbb{A}_r$, update $s_{i|\mathbb{A}_r}(b_r)$
  Let $w \in \text{argmax}_i \; s_{i|\mathbb{A}_r}(b_r)$
  $\mathbb{A}_{r+1} \leftarrow \mathbb{A}_{r+1} \cup \{w\}$
  $r \leftarrow r + 1$
end while

Output: The remaining item in $\mathbb{A}_r$

$r \in \{1, \dots, \left\lceil \log_{1-\frac{1}{k}}\left(\frac{1}{n}\right) \right\rceil + k - 1\}$ is at most $\left\lceil \frac{n(1-\frac{1}{k})^{r-1}}{k} \right\rceil$. The budget allocated to a partition in round r (i.e., b_r) is by definition given by

$$b_r = \lfloor B/(R^{\text{CSR}} P_r^{\text{CSR}}) \rfloor = \left\lfloor \frac{B}{\left\lceil \frac{n(1-\frac{1}{k})^{r-1}}{k} \right\rceil \left(\left\lceil \log_{1-\frac{1}{k}}\left(\frac{1}{n}\right) \right\rceil + k - 1 \right)} \right\rfloor .$$

Consequently, the total budget used by CSR is

$$\begin{aligned} & \sum_{r=1}^{\left\lceil \log_{1-\frac{1}{k}}\left(\frac{1}{n}\right) \right\rceil + k - 1} \#\{\text{partitions in round r}\} \cdot b_r \\ = & \sum_{r=1}^{\left\lceil \log_{1-\frac{1}{k}}\left(\frac{1}{n}\right) \right\rceil + k - 1} \left\lceil \frac{n(1-\frac{1}{k})^{r-1}}{k} \right\rceil \left\lfloor \frac{B}{\left\lceil \frac{n(1-\frac{1}{k})^{r-1}}{k} \right\rceil \left(\left\lceil \log_{1-\frac{1}{k}}\left(\frac{1}{n}\right) \right\rceil + k - 1 \right)} \right\rfloor \\ \leq & B. \end{aligned}$$

Therefore, the statement follows from Theorem 4.1.

□

For sake of convenience, we provide the entire pseudo-code of CSH in Algorithm 5, which results by using $f(x) = \lceil x/2 \rceil$ as well as P_r^{CSH} and R^{CSH} as defined in Section 4.1 in Algorithm 1.

Proof of Corollary 4.3 (CSH case). Suppose $B > 0$ to be arbitrary but fixed. First, note that there are at most $\lceil \log_2(n) \rceil$ rounds within the first while-loop and at most $\lceil \log_2(k) \rceil$ in the second, so that

Algorithm 4 Combinatorial Successive Reject (CSR)

Input: set of arms $[n]$, subset size $k \leq n$, sampling budget B

Initialization: For each $r \in \{0, \dots, \lceil \log_{1-\frac{1}{k}}\left(\frac{1}{n}\right)\rceil\}$ let $b_r := \left\lfloor \frac{B}{\left\lceil \frac{n(1-\frac{1}{k})^{r-1}}{k} \right\rceil \left(\left\lceil \log_{1-\frac{1}{k}}\left(\frac{1}{n}\right)\right\rceil + k - 1\right)} \right\rfloor$,

$\mathbb{A} \leftarrow [n]$, $r \leftarrow 1$

while $|\mathbb{A}_r| \geq k$ **do**
  $J = \lceil \frac{n(1-\frac{1}{k})^{r-1}}{k} \rceil$
  $\mathbb{A}_{r1}, \mathbb{A}_{r2}, \dots, \mathbb{A}_{r,J} \leftarrow Partition(\mathbb{A}_r, k)$
  if $|\mathbb{A}_{r,J}| < k$ **then**
    $\mathcal{R} \leftarrow \mathbb{A}_{r,J}$, $J \leftarrow J - 1$
  else
    $\mathcal{R} \leftarrow \emptyset$
  end if
  $\mathbb{A}_{r+1} \leftarrow \mathbb{A}_r$
  for $j \in [J]$ **do**
    Play the set $\mathbb{A}_{r,j}$ for b_r times
    For all $i \in \mathbb{A}_{r,j}$, update $s_{i|\mathbb{A}_{r,j}}(b_r)$
    Let $w \in \arg\min_i \; s_{i|\mathbb{A}_{r,j}}(b_r)$
    $\mathbb{A}_{r+1} = \mathbb{A}_{r+1} \backslash \{w\}$
  end for
  $\mathbb{A}_{r+1} \leftarrow \mathbb{A}_{r+1} \cup \mathcal{R}$
  $r \leftarrow r + 1$
end while
$\mathbb{A}_{r+1} \leftarrow \mathbb{A}_r$
while $|\mathbb{A}_r| > 1$ **do**
  Play the set $\mathbb{A}_r$ for b_r times
  For all $i \in \mathbb{A}_r$, update $s_{i|\mathbb{A}_r}(b_r)$
  Let $w \in \arg\min_i \; s_{i|\mathbb{A}_r}(b_r)$
  $\mathbb{A}_{r+1} = \mathbb{A}_{r+1} \backslash \{w\}$
  $r \leftarrow r + 1$
end while

Output: The remaining item in $\mathbb{A}_r$

we have at most $\lceil \log_2(n) \rceil + \lceil \log_2(k) \rceil$ many rounds in total. The total number of partitions in round $r = 1, \dots, \lceil \log_2(n) \rceil + \lceil \log_2(k) \rceil$ is at most $\left\lceil \frac{n}{2^{r-1}k} \right\rceil$. The budget allocated to a partition in round r is

$$b_r = \lfloor B/(R^{\mathrm{CSH}} P_r^{\mathrm{CSH}}) \rfloor = \left\lfloor \frac{B}{\left\lceil \frac{n}{2^{r-1}k} \right\rceil \cdot (\lceil \log_2(n) \rceil + \lceil \log_2(k) \rceil)} \right\rfloor.$$

In particular, the total budget used by CSH is

$$\begin{aligned} & \sum_{r=1}^{\lceil \log_2(n) \rceil + \lceil \log_2(k) \rceil} \#\{\text{partitions in round r}\} \cdot b_r \\ = & \sum_{r=1}^{\lceil \log_2(n) \rceil + \lceil \log_2(k) \rceil} \left\lceil \frac{n}{2^{r-1}k} \right\rceil \cdot \left\lfloor \frac{Bk}{\lceil \frac{n}{2^{r-1}} \rceil (\lceil \log_2(n) \rceil + \lceil \log_2(k) \rceil)} \right\rfloor \\ \leq & B. \end{aligned}$$

Once again, Theorem 4.1 allows us to conclude the proof. □

E Proofs of Section 5

E.1 Stochastic Numerical Feedback: Proof of Corollary 5.1

A rich class of statistics can be obtained by applying a linear functional $U(F) = \int r(x) \mathrm{d}F(x)$, where F is a cumulative distribution function and $r : \mathbb{R} \to \mathbb{R}$ some measurable function, on the empirical distribution function [49], i.e., for any $x \in \mathbb{R}$ and any multiset of (reward) observations O

Algorithm 5 Combinatorial Successive Halving (CSH)

Input: set of arms $[n]$, subset size $k \leq n$, sampling budget B

Initialization: For each $r \in \{0, \dots, \lceil \log_2(n) \rceil + \lceil \log_2(k) \rceil\}$ let $b_r := \left\lfloor \frac{Bk}{\lceil \frac{n}{2^{r-1}} \rceil (\lceil \log_2(n) \rceil + \lceil \log_2(k) \rceil)} \right\rfloor$,

$\mathbb{A} \leftarrow [n]$, $r \leftarrow 1$

while $|\mathbb{A}_r| \geq k$ **do**
  $J = \lceil \frac{n}{2^{r-1}k} \rceil$
  $\mathbb{A}_{r1}, \mathbb{A}_{r2}, \dots, \mathbb{A}_{r,J} \leftarrow Partition(\mathbb{A}_r, k)$
  if $|\mathbb{A}_{r,j}| < k$ **then**
    $\mathcal{R} \leftarrow \mathbb{A}_{r,j}$, $J \leftarrow J - 1$
  else
    $\mathcal{R} \leftarrow \emptyset$
  end if
  for $j \in [J]$ **do**
    Play the set $\mathbb{A}_{r,j}$ for b_r times
    For all $i \in \mathbb{A}_{r,j}$, update $s_{i|\mathbb{A}_{r,j}}(b_r)$
    Define $\bar{s} \leftarrow$ Median$(\{s_{i|\mathbb{A}_{r,j}}(b_r)\}_{i \in \mathbb{A}_{r,j}})$
    $\mathbb{A}_{r+1} \leftarrow \{i \in \mathbb{A}_{r,j} | s_{i|\mathbb{A}_{r,j}}(b_r) \leq \bar{s}\}$
  end for
  $\mathbb{A}_{r+1} \leftarrow \mathbb{A}_{r+1} \cup \mathcal{R}$
  $r \leftarrow r + 1$
end while
$\mathbb{A}_r \leftarrow \mathbb{A}_r \cup \{k - |\mathbb{A}_r|$ random elements from $[n] \backslash \mathbb{A}_r\}$
while $|\mathbb{A}_r| > 1$ **do**
  Play the set $\mathbb{A}_r$ for b_r times
  For all $i \in \mathbb{A}_r$, update $s_{i|\mathbb{A}_r}(b_r)$
  Define $\bar{s} \leftarrow$ Median$(\{s_{i|\mathbb{A}_r}(b_r)\}_{i \in \mathbb{A}_r})$
  $\mathbb{A}_{r+1} \leftarrow \{i \in \mathbb{A}_r | s_{i|\mathbb{A}_r}(b_r) \leq \bar{s}\}$
  $r \leftarrow r + 1$
end while

Output: The remaining item in $\mathbb{A}_r$

$$\tilde{s}(O, x) = \tfrac{1}{|O|} \textstyle\sum_{o \in O} \mathbb{1}\{x \leq o\}.$$

This leads to the statistics

$$s_{i|Q}(t) = U(\tilde{s}(o_{i|Q}(1), \dots, o_{i|Q}(t), \cdot)) = \sum_{s=1}^{t} \frac{r(o_{i|Q}(s))}{t},$$

which converge to $S_{i|Q} = \mathbb{E}_{X \sim \nu_{i|Q}}[r(X)]$ by the law of large numbers, provided these expected values exist. In this section we show the following result which generalizes Corollary 5.1 for statistics of the above kind.

Corollary E.1. *Let f, R and $\{P_r\}_{r \in [R]}$ be as in Theorem 4.1 and suppose that $r(o_{i|Q}(t))$ are σ-sub-Gaussian and such that their means $S_{i|Q} := \mathbb{E}_{X \sim \nu_{i|Q}}[r(X)]$ satisfy (A2). Then, there is a function*

$$C(\delta, \varepsilon, k, R, \sigma) \in \mathcal{O}\left(\sigma^2 \varepsilon^{-2} \ln\left({}^{kR}/_{\delta} \ln\left({}^{kR\sigma}/_{\varepsilon\delta}\right)\right)\right)$$

with the following property: If i^* is the GCW and $\sup_{Q \in \mathcal{Q}_{\leq k}(i^*)} \Delta_{(f(|Q|)+1)|Q} \leq \varepsilon$, then Algorithm 1 used with a budget B larger than $C(\delta, \varepsilon, k, R, \sigma) \cdot R \max_{r \in [R]} P_r$ returns i^* with probability at least $1 - \delta$.

Note that we immediately obtain the proof for Corollary 5.1 as a special case of Corollary E.1 by using the the identity function $r(x) = x$.

The following two lemmata serve as a preparation for the proof of Corollary E.1. The proof of Lemma E.2 is an adaptation of the proof of Lemma 3 in [24].

Lemma E.2. *Let $X_1, X_2, \dots \sim \mathcal{X}$ be iid real-valued random variables and $r : \mathbb{R} \to \mathbb{R}$ such that $r(\mathcal{X})$ is σ^2-sub-Gaussian. For any $\epsilon \in (0, 1)$ and $\delta \in (0, \log(1 + \epsilon)/e)$ one has with probability at*

least $1 - \frac{(2+\epsilon)}{\epsilon}\left(\frac{\delta}{\log(1+\epsilon)}\right)^{(1+\epsilon)}$ *for any* $t \geq 1$

$$\sum_{i=1}^{t} r(X_i) - t \cdot \mathbb{E}_{X\sim\mathcal{X}}[r(X)] \leq (1+\sqrt{\epsilon})\sqrt{2\sigma^2(1+\epsilon)t\log\left(\frac{\log((1+\epsilon)t)}{\delta}\right)}.$$

Moreover, the same concentration inequality holds for $-\left(\sum_{i=1}^{t} r(X_i) - t \cdot \mathbb{E}_{X\sim\mathcal{X}}[r(X)]\right)$ *as well.*

Proof. We denote in the following $\psi(x) = \sqrt{2\sigma^2 x \log\left(\frac{\log(x)}{\delta}\right)}$ and $R_t = \sum_{i=1}^{t} r(X_i) - t \cdot \mathbb{E}_{X\sim\mathcal{X}}[r(X)]$ and define a sequence of integers (u_k) as $u_0 = 1$ and $u_{k+1} = \lceil(1+\epsilon)u_k\rceil$. The maximal *Azuma-Hoeffding Inequality* states that for any martingale difference sequence $S_1, S_2, \ldots$ with each element being σ^2-sub-Gaussian, it holds that for any $\alpha > 0$, $n \geq 1$:

$$\mathbb{P}\left(\max_{i\in[n]} S_i - S_0 \geq \alpha\right) \leq \exp\left(-\frac{\alpha^2}{2\sum_{j=1}^{n}\sigma_j^2}\right).$$

In the following let $\mathcal{F}_0 = \{\emptyset, \Omega\}$ be the trivial σ-algebra and for $k \in \{1, \ldots, n\}$ let $\mathcal{F}_k = \sigma(X_1, \ldots, X_k)$ be the σ-algebra generated by the observations $X_1, \ldots, X_k$. Then

$$\begin{aligned}
\mathbb{E}[R_{t+1}|\mathcal{F}_t] &= \mathbb{E}[r(X_{t+1}) - \mathbb{E}_{X\sim\mathcal{X}}[r(X)] + R_t|\mathcal{F}_t] \\
&= \mathbb{E}[r(X_{t+1})|\mathcal{F}_t] - \mathbb{E}_{X\sim\mathcal{X}}[r(X)] + \mathbb{E}[R_t|\mathcal{F}_t] \\
&= R_t
\end{aligned}$$

which shows the martingale property of R_t. Note, that $R_0 = 0$ and $R_{t+1} - R_t = r(X_{t+1}) - \mathbb{E}_{X\sim\mathcal{X}}[r(X)]$, which is according to the assumption σ^2-sub-Gaussian and has zero mean, for any $t \in \mathbb{N}$. Thus, we can apply the maximal Azuma-Hoeffding inequality for $R_1, R_2, \ldots, R_t$.

Step 1.
In the first step of the proof we derive a bound for the probability of a lower bound of R_{u_k} for $k \geq 1$. For this we use the union bound, the maximal Azuma-Hoeffding inequality, the fact that $u_k \geq (1+\epsilon)^k$, a sum-integral comparison and some simple transformations and obtain

$$\begin{aligned}
&\mathbb{P}\left(\exists k \geq 1 \,:\, R_{u_k} \geq \sqrt{1+\epsilon}\psi(u_k)\right) \\
&\leq \sum_{k=1}^{\infty} \mathbb{P}\left(R_{u_k} \geq \sqrt{1+\epsilon}\psi(u_k)\right) \\
&\leq \sum_{k=1}^{\infty} \exp\left(-\frac{(1+\epsilon)\psi(u_k)^2}{2u_k\sigma^2}\right) \\
&= \sum_{k=1}^{\infty} \exp\left(-(1+\epsilon)\log\left(\frac{\log(u_k)}{\delta}\right)\right) \\
&\leq \sum_{k=1}^{\infty} \exp\left(-(1+\epsilon)\log\left(\frac{\log((1+\epsilon)^k)}{\delta}\right)\right) \\
&= \sum_{k=1}^{\infty} \left(\frac{\delta}{k\log((1+\epsilon))}\right)^{(1+\epsilon)} \\
&= \left(\frac{2\delta}{\log((1+\epsilon))}\right)^{(1+\epsilon)} \sum_{k=1}^{\infty}\left(\frac{1}{k}\right)^{(1+\epsilon)} \\
&= \left(\frac{\delta}{\log((1+\epsilon))}\right)^{(1+\epsilon)} \left(1 + \sum_{k=2}^{\infty}\left(\frac{1}{k}\right)^{(1+\epsilon)}\right) \\
&\leq \left(\frac{\delta}{\log((1+\epsilon))}\right)^{(1+\epsilon)} \left(1 + \int_{k=1}^{\infty}\left(\frac{1}{k}\right)^{(1+\epsilon)}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \left(\frac{\delta}{\log((1+\epsilon))}\right)^{(1+\epsilon)} \left(1 + \left[-\frac{1}{\epsilon}\left(\frac{1}{k}\right)^{\epsilon}\right]_1^{\infty}\right) \\
&= \left(\frac{\delta}{\log((1+\epsilon))}\right)^{(1+\epsilon)} \left(1 + \frac{1}{\epsilon}\right).
\end{aligned}
$$

Step 2.
Next, we bound the probability that the difference between some R_s and R_t exceeds a lower bound for some $s = u_k$, $k \in \mathbb{N}$ and $s \leq t \leq u_{k+1}$. Note that $R_t - R_{u_k}$ and R_{t-u_k} have the same distribution, such that we obtain

$$
\begin{aligned}
&\mathbb{P}\left(\exists t \in \{u_k+1, \dots, u_{k+1}-1\} \,:\, R_t - R_{u_k} \geq \sqrt{\epsilon}\psi(u_{k+1})\right) \\
&= \mathbb{P}\left(\exists t \in [u_{k+1} - u_k - 1] \,:\, R_t \geq \sqrt{\epsilon}\psi(u_{k+1})\right) \\
&\leq \exp\left(-\frac{\epsilon\psi(u_{k+1})^2}{2\sigma^2(u_{k+1}-u_k-1)}\right) \\
&= \exp\left(-\frac{\epsilon u_{k+1}}{u_{k+1}-u_k-1}\log\left(\frac{\log(u_{k+1})}{\delta}\right)\right) \\
&\leq \exp\left(-\frac{\epsilon u_{k+1}}{(1+\epsilon)u_k+1-u_k-1}\log\left(\frac{\log(u_{k+1})}{\delta}\right)\right) \\
&= \exp\left(-\frac{u_{k+1}}{u_k}\log\left(\frac{\log(u_{k+1})}{\delta}\right)\right) \\
&\leq \exp\left(-(1+\epsilon)\log\left(\frac{\log(u_{k+1})}{\delta}\right)\right) \\
&\leq \left(\frac{\delta}{(k+1)\log(1+\epsilon)}\right)^{1+\epsilon},
\end{aligned}
$$

where we used once again the maximal Azuma-Hoeffding inequality and that $u_{k+1} \geq (1+\epsilon)u_k$ as well as that $\frac{u_{k+1}}{u_k} \geq 1+\epsilon$. For all possible $k \in \mathbb{N}$ we get with the union bound and a similar sum-integral comparison as above

$$
\begin{aligned}
&\mathbb{P}\left(\exists k \in \mathbb{N},\ \exists t \in \{u_k+1, \dots, u_{k+1}-1\} \,:\, R_t - R_{u_k} \geq \sqrt{\epsilon}\psi(u_{k+1})\right) \\
&\leq \sum_{k=1}^{\infty}\left(\frac{\delta}{(k+1)\log(1+\epsilon)}\right)^{1+\epsilon} \\
&= \sum_{k=2}^{\infty}\left(\frac{\delta}{k\log(1+\epsilon)}\right)^{1+\epsilon} \\
&\leq \int_{k=1}^{\infty}\left(\frac{\delta}{k\log(1+\epsilon)}\right)^{1+\epsilon} \\
&= \left(\frac{\delta}{\log(1+\epsilon)}\right)^{1+\epsilon}\frac{1}{\epsilon}
\end{aligned}
$$

Step 3.
Finally, by combining Step 1 and 2 we can infer that for any $k \geq 0$ and $t \in \{u_k+1, \dots, u_{k+1}-1\}$ it holds

$$
\begin{aligned}
R_t &= R_t - R_{u_k} + R_{u_k} \\
&\leq \sqrt{\epsilon}\psi(u_{k+1}) + \sqrt{1+\epsilon}\psi(u_k) \\
&\leq \sqrt{\epsilon}\psi((1+\epsilon)t) + \sqrt{1+\epsilon}\psi(t) \\
&\leq (1+\sqrt{\epsilon})\psi((1+\epsilon)t),
\end{aligned}
$$

with probability at least $1 - \frac{2+\epsilon}{\epsilon}\left(\frac{\delta}{\log(1+\epsilon)}\right)^{1+\epsilon}$ leading to the first claim of the lemma.

Step 4.

Note that $\tilde{R}_t = t \cdot \mathbb{E}_{X\sim\mathcal{X}}[r(X)] - \sum_{i=1}^t r(X_i)$ is a martingale difference sequence with $\tilde{R}_{t+1} - \tilde{R}_t = -R_t + R_{t+1} = \mathbb{E}_{X\sim\mathcal{X}}[r(X)] - r(X_{t+1})$, which is according to the assumption σ^2-sub-Gaussian and has zero mean, for any $t \in \mathbb{N}$. Thus, repeating Step 1–3 for $\tilde{R}_1, \tilde{R}_2, \ldots, \tilde{R}_t$ shows the second claim of the lemma. □

Lemma E.3. *Let $X_1, X_2, \ldots \sim \mathcal{X}$ be iid real-valued random variables and $r : \mathbb{R} \to \mathbb{R}$ such that $r(\mathcal{X})$ is σ^2-sub-Gaussian. For any $\gamma \in (0,1)$ we have*

$$\mathbb{P}\left(\exists t \in \mathbb{N} : \left|\sum_{i=1}^t r(X_i) - t \cdot \mathbb{E}_{X\sim\mathcal{X}}[r(X)]\right| > (1+\sqrt{1/2})\sqrt{3\sigma^2 t \ln\left(\frac{10^{2/3}\ln(3t/2)}{\gamma^{2/3}\ln(3/2)}\right)}\right) \le \gamma.$$

Proof. Let $\gamma \in (0,1)$ be fixed and $\varepsilon := 1/2$. Then, $\gamma' := \left(\frac{\gamma}{10}\right)^{2/3}\ln(3/2)$ fulfills

$$\frac{2+\varepsilon}{\varepsilon}\left(\frac{\gamma'}{\ln(1+\varepsilon)}\right)^{1+\varepsilon} = 5\left((\gamma/10)^{2/3}\right)^{3/2} = \gamma/2$$

and moreover $\gamma' < (1/10)^{2/3}\ln(3/2) < e^{-1}\ln(3/2)$. Consequently, Lemma E.2 yields with

$$\tilde{c}_\gamma(t) := (1+\sqrt{\varepsilon})\sqrt{2\sigma^2(1+\varepsilon)t\ln\left(\frac{\ln((1+\varepsilon)t)}{\gamma'}\right)} = (1+\sqrt{1/2})\sqrt{3\sigma^2 t\ln\left(\frac{10^{2/3}\ln(3t/2)}{\gamma^{2/3}\ln(3/2)}\right)}$$

that

$$\mathbb{P}\left(\exists t \in \mathbb{N} : \sum_{i=1}^t r(X_i) - t \cdot \mathbb{E}_{X\sim\mathcal{X}}[r(X)] > \tilde{c}_\gamma(t)\right) \le \gamma/2.$$

as well as

$$\mathbb{P}\left(\exists t \in \mathbb{N} : -\left(\sum_{i=1}^t r(X_i) - t \cdot \mathbb{E}_{X\sim\mathcal{X}}[r(X)]\right) > \tilde{c}_\gamma(t)\right) \le \gamma/2.$$

Thus, we obtain

$$\begin{aligned}
&\mathbb{P}\left(\exists t \in \mathbb{N} : \left|\sum_{i=1}^t r(X_i) - t \cdot \mathbb{E}_{X\sim\mathcal{X}}[r(X)]\right| > \tilde{c}_\gamma(t)\right) \\
&\quad \le \mathbb{P}\left(\exists t \in \mathbb{N} : \sum_{i=1}^t r(X_i) - t \cdot \mathbb{E}_{X\sim\mathcal{X}}[r(X)] > \tilde{c}_\gamma(t)\right) \\
&\quad + \mathbb{P}\left(\exists t \in \mathbb{N} : -\left(\sum_{i=1}^t r(X_i) - t \cdot \mathbb{E}_{X\sim\mathcal{X}}[r(X)]\right) > \tilde{c}_\gamma(t)\right) \\
&\quad \le \gamma/2 + \gamma/2 = \gamma.
\end{aligned}$$

□

We are now ready to prove Corollary E.1.

Proof of Corollary E.1. Recall the definition of $\tilde{c}_\gamma(t)$ from the proof of Lemma E.3 and let

$$c_\gamma(t) := \frac{2}{t}\tilde{c}_\gamma(t) = 2(1+\sqrt{1/2})\sqrt{\frac{3\sigma^2}{t}\ln\left(\frac{10^{2/3}\ln(3t/2)}{\gamma^{2/3}\ln(3/2)}\right)}$$

for any $\gamma \in (0,1)$, $t \in \mathbb{N}$. For any fixed γ, $c_\gamma : \mathbb{N} \to (0,\infty), t \mapsto c_\gamma(t)$ is strictly monotonically decreasing with $\lim_{t\to\infty} c_\gamma(t) = 0$. Contraposition of (1) in [24] states

$$t > \frac{1}{c}\ln\left(\frac{2\ln((1+\varepsilon)/(c\omega))}{\omega}\right) \Rightarrow c > \frac{1}{t}\ln\left(\frac{\ln((1+\varepsilon)t)}{\omega}\right) \qquad \forall t \ge 1, \varepsilon \in (0,1), c > 0, \omega \le 1.$$

For any $\alpha > 0$ and $\gamma \in (0,1)$, using this with $\omega = \frac{\gamma^{2/3}\ln(3/2)}{10^{2/3}}, c = \frac{\alpha^2}{12(1+\sqrt{1/2})^2\sigma^2}$ and $\varepsilon = 1/2$ reveals

$$\begin{aligned} c_\gamma^{-1}(\alpha) &= \min\{t \in \mathbb{N} : c_\gamma(t) \le \alpha\} \\ &= \min\left\{t \in \mathbb{N} : \frac{1}{t}\ln\left(\frac{10^{2/3}\ln(3t/2)}{\gamma^{2/3}\ln(3/2)}\right) \le \frac{\alpha^2}{12(1+\sqrt{1/2})^2\sigma^2}\right\}. \end{aligned}$$

Thus, we have $c \ge \frac{1}{t}\ln\left(\frac{\ln((1+\varepsilon)t)}{\omega}\right)$ and we know, that this statement is true if $t \ge \frac{1}{c}\ln\left(\frac{2\ln((1+\varepsilon)/(c\omega))}{\omega}\right)$. In particular also for the smallest such t, for which holds $t \le \left\lceil \frac{1}{c}\ln\left(\frac{2\ln((1+\varepsilon)/(c\omega))}{\omega}\right)\right\rceil + 1$. It follows

$$c_\gamma^{-1}(\alpha) \le \left\lceil \frac{12(1+\sqrt{1/2})^2\sigma^2}{\alpha^2}\ln\left(\frac{2\cdot 10^{2/3}}{\gamma^{2/3}\ln(3/2)}\ln\left(\frac{18\cdot 10^{2/3}(1+\sqrt{1/2})^2\sigma^2}{\gamma^{2/3}\ln(3/2)\,\alpha^2}\right)\right)\right\rceil + 1,$$

which is of the order $\mathcal{O}(\sigma^2\alpha^{-2}\ln\ln(\alpha^{-1}\sigma)\ln\gamma^{-1})$.

Now, suppose $\max_{Q\in\mathcal{Q}_{\le k}(i^*)} \Delta_{(f(|Q|)+1)|Q} \le \varepsilon$ and that Algorithm 1 is started with a budget B larger than

$$c_{\delta/(kR)}^{-1}(\varepsilon/2)\cdot R\max_{r\in[R]} P_r.$$

Recall that $\gamma_{i|Q}(t) = |s_{i|Q}(t) - S_{i|Q}|$, $s_{i|Q}(t) = \frac{1}{t}\sum_{s=1}^t r(o_{i|Q}(s))$ and $S_{i|Q} = \mathbb{E}_{X\sim\nu_{i|Q}}[r(X)]$. With this, we obtain for any possible sequence of partitions $(E_r)_{r\in[R]} \in (\mathcal{Q}_{\le k})^R$ with $\mathbb{P}(\mathbb{A}_r(i^*) = E_r\,\forall r\in[R]) > 0$ that

$$\begin{aligned} &\mathbb{P}\left(\exists t\in\mathbb{N}, r\in[R], i\in E_r : \gamma_{i|E_r}(t) \ge c_{\delta/(kR)}(t) \,\middle|\, \mathbb{A}_r(i^*) = E_r\,\forall r\in[R]\right) \\ &\le \sum_{r\in[R]}\sum_{i\in E_r} \mathbb{P}\left(\exists t\in\mathbb{N} : \gamma_{i|E_r}(t) \ge c_{\delta/(kR)}(t) \,\middle|\, \mathbb{A}_r(i^*) = E_r\,\forall r\in[R]\right) \\ &= \sum_{r\in[R]}\sum_{i\in E_r} \mathbb{P}\left(\exists t\in\mathbb{N} : \left|\frac{1}{t}\sum_{t'=1}^t r(o_{i|E_r}(t')) - \mathbb{E}_{X\sim\nu_{i|E_r}}[r(X)]\right| \ge c_{\delta/(kR)}(t) \,\middle|\, \mathbb{A}_r(i^*) = E_r\,\forall r\in[R]\right) \\ &= \sum_{r\in[R]}\sum_{i\in E_r} \mathbb{P}\left(\exists t\in\mathbb{N} : \frac{1}{t}\left|\sum_{t'=1}^t r(o_{i|E_r}(t')) - t\cdot\mathbb{E}_{X\sim\nu_{i|E_r}}[r(X)]\right| \ge c_{\delta/(kR)}(t) \,\middle|\, \mathbb{A}_r(i^*) = E_r\,\forall r\in[R]\right) \\ &= \sum_{r\in[R]}\sum_{i\in E_r} \mathbb{P}\left(\exists t\in\mathbb{N} : \left|\sum_{t'=1}^t r(o_{i|E_r}(t')) - t\cdot\mathbb{E}_{X\sim\nu_{i|E_r}}[r(X)]\right| \ge \tilde{c}_{\delta/(kR)}(t) \,\middle|\, \mathbb{A}_r(i^*) = E_r\,\forall r\in[R]\right) \\ &\le \sum_{r\in[R]}\sum_{i\in E_r} \frac{\delta}{kR} \le \delta, \end{aligned}$$

where we used Lemma E.3 for the second last inequality. Using the law of total probability for all possible sequences of partitions $(E_r)_{r\in[R]}$, we see that the event

$$\mathcal{E} := \left\{\exists t\in\mathbb{N}, r\in[R], i\in\mathbb{A}_r(i^*) : \gamma_{i|\mathbb{A}_r(i^*)}(t) \ge c_{\delta/(kR)}(t)\right\}$$

occurs with probability

$$\begin{aligned} &\mathbb{P}(\mathcal{E}) \\ &= \sum_{(E_r)_{r\in[R]}} \mathbb{P}\left(\exists t\in\mathbb{N}, r\in[R], i\in E_r : \gamma_{i|E_r}(t) \ge c_{\delta/(kR)}(t) \,\middle|\, \mathbb{A}_r(i^*) = E_r\,\forall r\in[R]\right) \\ &\quad\times\ \mathbb{P}(\mathbb{A}_r(i^*) = E_r\,\forall r\in[R]) \\ &\le \delta\sum\nolimits_{(E_r)_{r\in[R]}} \mathbb{P}(\mathbb{A}_r(i^*) = E_r\,\forall r\in[R]) = \delta. \end{aligned}$$

On $\mathcal{E}^c$ we have $\overline{\gamma}_{\mathbb{A}_r(i^*)}(t) < c_{\delta/(kR)}(t)$ for all $t\in\mathbb{N}, r\in[R]$ and thus in particular $\overline{\gamma}_{\mathbb{A}_r(i^*)}^{-1}(\alpha) \ge c_{\delta/(kR)}^{-1}(\alpha)$ for any $\alpha\in(0,\infty)$. Since $\max_{Q\in\mathcal{Q}_{\le k}(i^*)} \Delta_{(f(|Q|)+1)|Q} \le \varepsilon$, Theorem 4.1 thus lets us

conclude

$$\begin{aligned}\mathbb{P}\left(\text{Alg. 1 returns } i^*\right) &\geq \mathbb{P}\left(B > R\max_{r\in[R]} P_r \overline{\gamma}^{-1}_{\mathbb{A}_r(i^*)}\left(\frac{\Delta_{(f(|\mathbb{A}_r(i^*)|)+1)|\mathbb{A}_r(i^*)}}{2}\right)\right)\\ &\geq \mathbb{P}\left(\left\{B > R\max_{r\in[R]} P_r \overline{\gamma}^{-1}_{\mathbb{A}_r(i^*)}(\varepsilon/2)\right\}\cap\mathcal{E}^c\right)\\ &\geq \mathbb{P}\left(\left\{B > R\max_{r\in[R]} P_r c^{-1}_{\delta/(kR)}(\varepsilon/2)\right\}\cap\mathcal{E}^c\right)\\ &= \mathbb{P}(\mathcal{E}^c) \geq 1-\delta,\end{aligned}$$

where the equality holds due to the assumption on B. Consequently, we can conclude the proof by defining

$$\begin{aligned}C(\delta,\varepsilon,k,R) &\coloneqq c^{-1}_{\delta/(kR)}(\varepsilon/2)\\ &\leq \left\lceil \frac{48(1+\sqrt{1/2})^2\sigma^2}{\varepsilon^2}\ln\left(\frac{2(10kR)^{2/3}}{\delta^{2/3}\ln(3/2)}\ln\left(\frac{72\cdot(10kR)^{2/3}(1+\sqrt{1/2})^2\sigma^2}{\delta^{2/3}\varepsilon^2\ln(3/2)}\right)\right)\right\rceil + 1\\ &\in \mathcal{O}\left(\frac{\sigma^2}{\varepsilon^2}\ln\left(\frac{kR}{\delta}\ln\left(\frac{kR\sigma}{\varepsilon\delta}\right)\right)\right).\end{aligned}$$

□

E.2 Stochastic Preference Winner Feedback: Proof of Corollary 5.2

The following two lemmata serve as a preparation for the proof of Corollary 5.2. But first let us introduce the $(k-1)$-simplex

$$\mathcal{S}_k = \left\{(p_i)_{i\in[k]} \in [0,1]^k \,:\, \sum\nolimits_{i=1}^k p_i = 1 \wedge \forall i : p_i \geq 0\right\}.$$

Lemma E.4 (Dvoretzky-Kiefer-Wolfowitz inequality for categorical random variables)**.** *Let $\{X_t\}_{t\in\mathbb{N}}$ be a sequence of iid random variables $X_t \sim \mathrm{Cat}(\mathbf{p})$ for some $\mathbf{p} \in \mathcal{S}_k$. For $t \in \mathbb{N}$ let $\hat{\mathbf{p}}^t$ be the corresponding empirical distribution after the t observations $X_1,\ldots,X_t$, i.e., $\hat{p}^t_i = \frac{1}{t}\sum_{s=1}^t \mathbf{1}_{\{X_s=i\}}$ for all $i \in [k]$. Then, we have for any $\varepsilon > 0$ and $t \in \mathbb{N}$ the estimate*

$$\mathbb{P}\left(\left|\left|\hat{\mathbf{p}}^t - \mathbf{p}\right|\right|_\infty > \varepsilon\right) \leq 4e^{-t\varepsilon^2/2}.$$

Proof. Confer [19, 35] as well as Theorem 11.6 in [29]. Moreover, note that the cumulative distribution functions F resp. $\hat{F}^t$ of $X_1 \sim \mathrm{Cat}(\mathbf{p})$ resp. $\hat{\mathbf{p}}^t$ fulfill $p_j = F(j) - F(j-1)$ and $\hat{p}^t_j = \hat{F}^t(j) - \hat{F}^t(j-1)$ and thus

$$|\hat{p}^t_j - p_j| \leq |\hat{F}^t(j) - F(j)| + |\hat{F}^t(j-1) - F(j-1)|.$$

for each $j \in [k]$. □

Lemma E.5. *For every $\beta \in [1, e/2]$, $c_1, c_2 > 0$ the number*

$$x \coloneqq \frac{\beta}{c_1}\left(\ln\left(\frac{c_2 e}{c_1^\beta}\right) + \ln\ln\left(\frac{c_2}{c_1^\beta}\right)\right)$$

fulfills $c_1 x \geq \ln(c_2 x^\beta)$.

Proof. This is Lemma 18 in [20]. □

Proof of Corollary 5.2. For $t \in \mathbb{N}$ and $\gamma \in (0,1)$ define

$$c_\gamma(t) \coloneqq \sqrt{\frac{4\ln(2\pi^2 t^2/(3\gamma))}{t}}$$

and note that, for any fixed γ, the function $c_\gamma : \mathbb{N} \to (0,\infty), t \mapsto c_\gamma(t)$ is strictly monotonically decreasing with $\lim_{t\to\infty} c_\gamma(t) = 0$. For any $\alpha > 0$, $\gamma \in (0,1)$, we obtain via Lemma E.5 with the choices $\beta = 1$, $c_1 = \frac{\alpha^2}{8}$ and $c_2 = \sqrt{2/(3\gamma)}\pi$ the estimate

$$\begin{aligned} c_\gamma^{-1}(\alpha) &= \min\left\{t \in \mathbb{N} : 4\ln(2\pi^2 t^2/(3\gamma)) \le t\alpha^2\right\} \\ &= \min\left\{t \in \mathbb{N} : \ln\left(\sqrt{2/(3\gamma)}\pi t\right) \le \frac{\alpha^2}{8} t\right\} \\ &\le \left\lceil \frac{8}{\alpha^2}\left(\ln\left(\frac{8\sqrt{2/(3\gamma)}\pi e}{\alpha^2}\right) + \ln\ln\left(\frac{8\sqrt{2/(3\gamma)}\pi}{\alpha^2}\right)\right)\right\rceil + 1. \end{aligned}$$

Now, suppose $\max_{Q\in\mathcal{Q}_{\le k}(i^*)} \Delta_{(f(|Q|)+1)|Q} \le \varepsilon$ and that Algorithm 1 is started with a budget B larger than

$$c_{\delta/R}^{-1}(\varepsilon/2) \cdot R \max_{r\in[R]} P_r.$$

Recall that in this preference-based setting we use as the statistic the empirical mean of the (winner) observations we obtained for arm i after querying Q (with $i \in Q$) for t many times. In particular, we set

$$s_{i|Q}(t) = \frac{w_{i|Q}(t)}{t} = \frac{1}{t}\sum\nolimits_{t'=1}^{t} o_{i|Q}(t'),$$

where $o_{i|Q}(t') = 1$ if arm i is the preferred (or winning) arm among the arms in Q, if Q is queried for the t'-th time, and 0 otherwise. Thus, $w_{i|Q}(t)$ is the total number of times arm i has won in the query set Q after t queries. Moreover, $\gamma_{i|Q}(t) = |s_{i|Q}(t) - S_{i|Q}|$, where $S_{i|Q} = p_{i|Q}$ and $o_{i|Q}(t') \sim \mathrm{Cat}(\mathbf{p_Q})$. With this, we obtain for any $t \in \mathbb{N}$ and any possible sequence of partitions $(E_r)_{r\in[R]} \in (\mathcal{Q}_{\le k})^R$ with $\mathbb{P}(\mathbb{A}_r(i^*) = E_r\, \forall r \in [R]) > 0$ that

$$\begin{aligned} &\mathbb{P}\left(\exists r \in [R] : \gamma_{i|E_r}(t) \ge c_{\delta/R}(t) \,\middle|\, \mathbb{A}_r(i^*) = E_r\, \forall r \in [R]\right) \\ &\le \sum_{r\in[R]} \mathbb{P}\left(\gamma_{i|E_r}(t) \ge c_{\delta/R}(t) \,\middle|\, \mathbb{A}_r(i^*) = E_r\, \forall r \in [R]\right) \\ &= \sum_{r\in[R]} \mathbb{P}\left(\max\nolimits_{i\in E_r} \left|\frac{1}{t}\sum_{t'=1}^{t} o_{i|E_r}(t') - S_{i|E_r}\right| > \sqrt{\frac{4\ln(2\pi^2 t^2/(3\gamma))}{t}} \,\middle|\, \mathbb{A}_r(i^*) = E_r\, \forall r \in [R]\right) \\ &\le \frac{6\delta}{\pi^2 t^2}, \end{aligned}$$

where we used Lemma E.4 in the last inequality. Using the law of total probability for all possible sequences of partitions $(E_r)_{r\in[R]}$, we see that the event

$$\mathcal{E} := \left\{\exists t \in \mathbb{N}, r \in [R] : \overline{\gamma}_{\mathbb{A}_r(i^*)}(t) \ge c_{\delta/R}(t)\right\}$$

occurs with probability

$$\begin{aligned} \mathbb{P}(\mathcal{E}) &\le \sum_{t\in\mathbb{N}} \sum_{(E_r)_{r\in[R]}} \mathbb{P}\left(\exists r \in [R] : \overline{\gamma}_{E_r}(t) \ge c_{\delta/R}(t) \,\middle|\, \mathbb{A}_r(i^*) = E_r\, \forall r \in [R]\right) \\ &\quad \times\ \mathbb{P}(\mathbb{A}_r(i^*) = E_r\, \forall r \in [R]) \\ &\le \sum\nolimits_{t\in\mathbb{N}} \frac{6\delta}{\pi^2 t^2} \sum\nolimits_{(E_r)_{r\in[R]}} \mathbb{P}(\mathbb{A}_r(i^*) = E_r\, \forall r \in [R]) \le \delta. \end{aligned}$$

On $\mathcal{E}^c$ we have $\overline{\gamma}_{\mathbb{A}_r(i^*)}(t) < c_{\delta/R}(t)$ for all $t \in \mathbb{N}, r \in [R]$ and thus in particular $\overline{\gamma}^{-1}_{\mathbb{A}_r(i^*)}(\alpha) \ge c^{-1}_{\delta/R}(\alpha)$ for any $\alpha \in (0,\infty)$. Since $\max_{Q\in\mathcal{Q}_{\le k}(i^*)} \Delta_{(f(|Q|)+1)|Q} \le \varepsilon$, Theorem 4.1 thus lets us conclude

$$\begin{aligned} \mathbb{P}\left(\text{Alg. 1 returns } i^*\right) &\ge \mathbb{P}\left(B > R\max\nolimits_{r\in[R]} P_r \overline{\gamma}^{-1}_{\mathbb{A}_r(i^*)}\left(\frac{\Delta_{(f(|\mathbb{A}_r(i^*)|)+1)|\mathbb{A}_r(i^*)}}{2}\right)\right) \\ &\ge \mathbb{P}\left(\left\{B > R\max\nolimits_{r\in[R]} P_r \overline{\gamma}^{-1}_{\mathbb{A}_r(i^*)}(\varepsilon/2)\right\} \cap \mathcal{E}^c\right) \\ &\ge \mathbb{P}\left(\left\{B > R\max\nolimits_{r\in[R]} P_r c^{-1}_{\delta/R}(\varepsilon/2)\right\} \cap \mathcal{E}^c\right) \\ &= \mathbb{P}(\mathcal{E}^c) \ge 1 - \delta, \end{aligned}$$

where the equality holds due to the assumption on B. Consequently, the statement holds with

$$\begin{aligned} C(\delta, \varepsilon, k, R) &\coloneqq c_{\delta/R}^{-1}(\varepsilon/2) \\ &\leq \left\lceil \frac{32}{\varepsilon^2} \left(\ln \left(\frac{32\sqrt{2R/(3\delta)}\pi e}{\varepsilon^2} \right) + \ln \ln \left(\frac{32\sqrt{2R/(3\delta)}\pi}{\varepsilon^2} \right) \right) \right\rceil + 1 \\ &\in \mathcal{O}\left(\frac{1}{\varepsilon^2} \ln \left(\frac{R}{\delta \varepsilon^4} \right) \right). \end{aligned}$$

□

F Comparisons of the Algorithms

In the following we summarize the theoretical results obtained for our proposed algorithms in a concise way. First of all, we give an overview of the individual key quantities of each algorithm in Table 2, where we assume w.l.o.g. that $\binom{n}{k}$ is a divisor of B in ROUNDROBIN to make the assignments of R, P_r and $f(s)$ for ROUNDROBIN well-defined. The maximal number of different query sets is derived in Section F.1.

Table 2: Comparison of the maximal number of rounds, the maximal number of partitions per round, the amount of retained arms from each partition and the maximal number of query sets for ROUNDROBIN and our proposed algorithms CSWS, CSR and CSH.

Alg.	R	P_r	$f(x)$	max #query_sets
ROUNDROBIN	1	$\binom{n}{k}$	$x \mapsto x$	$\binom{n}{k}$
CSWS	$\lceil \log_k(n) \rceil + 1$	$\left\lceil \frac{n}{k^r} \right\rceil$	$x \mapsto 1$	$R^{CSWS} + n \cdot \left(\frac{1 - 1/k^{\lceil \log_k(n) \rceil + 1}}{k-1} \right)$
CSR	$\left\lceil \log_{1-\frac{1}{k}} \left(\frac{1}{n}\right) \right\rceil + k - 1$	$\left\lceil \frac{n(1-\frac{1}{k})^{(r-1)}}{k} \right\rceil$	$x \mapsto x - 1$	$R^{CSR} + n \left(1 - \left(1 - \frac{1}{k}\right)^{\lceil \log_{1-\frac{1}{k}} \left(\frac{1}{n}\right) \rceil + k - 1} \right)$
CSH	$\lceil \log_2(n) \rceil + \lceil \log_2(k) \rceil$	$\left\lceil \frac{n}{2^{r-1} k} \right\rceil$	$x \mapsto \left\lceil \frac{x}{2} \right\rceil$	$R^{CSH} + \frac{2n}{k} \left(1 - 1/2^{\lceil \log_2(n) \rceil + \lceil \log_2(k) \rceil}\right)$

Using Remark D.1 we can derive the following sufficient budgets of the algorithms summarized in the following table, where $\pi(Q) \in Q$ be the $\lfloor \frac{|Q|}{2} \rfloor + 1$- th best arm with respect to $(S_{i|Q})_{i \in Q}$.

Table 3: Comparison of the sufficient budget for ROUNDROBIN and our proposed algorithms CSWS, CSR and CSH.

Algorithm	Sufficient budget
ROUNDROBIN	$\binom{n}{k} \max_{i \in \mathcal{A}, i \neq i^*_{\mathcal{B}}} \left(\hat{\gamma}^{\max}_{i^*_{\mathcal{B}}, i} \right)^{-1} \left(\frac{S^{\mathcal{B}}_{i^*_{\mathcal{B}}} - S^{\mathcal{B}}_i}{2} \right)$
CSWS	$\left\lceil \frac{n}{k} \right\rceil \left(\lceil \log_k(n) \rceil + 1 \right) \cdot \max_{Q \in \mathcal{Q}_{\leq k} : i^* \in Q} \max_{i \in Q \setminus \{i^*\}} \left\lceil \bar{\gamma}^{-1} \left(\frac{S_{i^*\|Q} - S_{i\|Q}}{2} \right) \right\rceil$
CSR	$\left\lceil \frac{n}{k} \right\rceil \left(\left\lceil \log_{1-\frac{1}{k}} \left(\frac{1}{n}\right) \right\rceil + k - 1 \right) \cdot \max_{Q \in \mathcal{Q}_{\leq k} : i^* \in Q} \min_{i \in Q \setminus \{i^*\}} \left\lceil \bar{\gamma}^{-1} \left(\frac{S_{i^*\|Q} - S_{i\|Q}}{2} \right) \right\rceil$
CSH	$\left\lceil \frac{n}{k} \right\rceil \left(\lceil \log_2(n) \rceil + \lceil \log_2(k) \rceil \right) \cdot \max_{Q \in \mathcal{Q}_{\leq k} : i^* \in Q} \left\lceil \bar{\gamma}^{-1} \left(\frac{S_{i^*\|Q} - S_{\pi(Q)\|Q}}{2} \right) \right\rceil$

In Section F.2 we compare these quantities for the special case, in which the gaps $\Delta_{i|Q} = S_{i^*|Q} - S_{i|Q}$ are all equal to some $\Delta > 0$, while in Section F.3 we derive the sufficient budgets resulting from Corollaries 5.1 and 5.2 for the reward setting and preference-based setting, respectively, to return the best arm with high probability in the stochastic setting. Note that if $\gamma_{i|Q}(t) = \gamma(t)$ and $S_{(2)|Q} = \cdots = S_{(|Q|)|Q}$ are fulfilled for all $Q \in \mathcal{Q}_{\leq k}, i \in Q$ and $t \in \mathbb{N}$, then the lower bound in Theorem 3.1 (i) matches the above upper bound for CSWS up to a factor $C = \lceil \log_k(n) \rceil + 1$.

F.1 Maximal Number of Different Query Sets

The maximal number of required query sets for each algorithm is $\sum_{r=1}^{R} P_r$. Note that this is a geometric series and thus the partial sum can easily be computed for each of our proposed algorithms.

CSWS By using the specified valued of R and P_r for CSWS, we obtain that the number of different query set is at most

$$\begin{aligned}
\sum_{r=1}^{R^{CSWS}} P_r^{CSWS} &= \sum_{r=1}^{\lceil \log_k(n) \rceil + 1} \left\lceil \frac{n}{k^r} \right\rceil \\
&\leq \lceil \log_k(n) \rceil + 1 + \sum_{r=1}^{\lceil \log_k(n) \rceil + 1} \frac{n}{k^r} \\
&= \lceil \log_k(n) \rceil + 1 + n \cdot \left(\sum_{r=0}^{\lceil \log_k(n) \rceil + 1} \left(\frac{1}{k}\right)^r - 1 \right) \\
&= \lceil \log_k(n) \rceil + 1 + n \cdot \left(\frac{1 - 1/k^{\lceil \log_k(n) \rceil + 2}}{1 - 1/k} - 1 \right) \\
&= \lceil \log_k(n) \rceil + 1 + n \cdot \left(\frac{1 - 1/k^{\lceil \log_k(n) \rceil + 1}}{k - 1} \right),
\end{aligned}$$

where we used for the inequality that $\lceil x \rceil \leq x + 1$ for any $x \in \mathbb{R}$.

CSR For CSR we get as an upper bound on the number of different query sets:

$$\begin{aligned}
\sum_{r=1}^{R^{CSR}} P_r^{CSR} &= \sum_{r=1}^{\lceil \log_{1-\frac{1}{k}}\left(\frac{1}{n}\right) \rceil + k - 1} \left\lceil \frac{n\left(1 - \frac{1}{k}\right)^{r-1}}{k} \right\rceil \\
&\leq \left\lceil \log_{1-\frac{1}{k}}\left(\frac{1}{n}\right) \right\rceil + k - 1 + \sum_{r=1}^{\lceil \log_{1-\frac{1}{k}}\left(\frac{1}{n}\right) \rceil + k - 1} \frac{n\left(1 - \frac{1}{k}\right)^{r-1}}{k} \\
&= \left\lceil \log_{1-\frac{1}{k}}\left(\frac{1}{n}\right) \right\rceil + k - 1 + \frac{n}{k} \sum_{r=0}^{\lceil \log_{1-\frac{1}{k}}\left(\frac{1}{n}\right) \rceil + k - 2} \left(1 - \frac{1}{k}\right)^r \\
&= \left\lceil \log_{1-\frac{1}{k}}\left(\frac{1}{n}\right) \right\rceil + k - 1 + \frac{n}{k} \frac{\left(1 - \left(1 - \frac{1}{k}\right)^{\lceil \log_{1-\frac{1}{k}}\left(\frac{1}{n}\right) \rceil + k - 1}\right)}{\left(1 - \left(1 - \frac{1}{k}\right)\right)} \\
&= \left\lceil \log_{1-\frac{1}{k}}\left(\frac{1}{n}\right) \right\rceil + k - 1 + n \left(1 - \left(1 - \frac{1}{k}\right)^{\lceil \log_{1-\frac{1}{k}}\left(\frac{1}{n}\right) \rceil + k - 1}\right).
\end{aligned}$$

CSH Similarly, we can obtain for CSH the following maximum number of different query sets:

$$\begin{aligned}
\sum_{r=1}^{R^{CSH}} P_r^{CSH} &= \sum_{r=1}^{\lceil \log_2(n) \rceil + \lceil \log_2(k) \rceil} \left\lceil \frac{n}{2^{r-1}k} \right\rceil \\
&\leq \lceil \log_2(n) \rceil + \lceil \log_2(k) \rceil + \sum_{r=1}^{\lceil \log_2(n) \rceil + \lceil \log_2(k) \rceil} \frac{n}{2^{r-1}k} \\
&= \lceil \log_2(n) \rceil + \lceil \log_2(k) \rceil + \frac{n}{k} \sum_{r=0}^{\lceil \log_2(n) \rceil + \lceil \log_2(k) \rceil - 1} \left(\frac{1}{2}\right)^r \\
&= \lceil \log_2(n) \rceil + \lceil \log_2(k) \rceil + \frac{2n}{k} \left(1 - 1/2^{\lceil \log_2(n) \rceil + \lceil \log_2(k) \rceil}\right).
\end{aligned}$$

These upper bounds on the maximum number of different query sets are summarized in Table 2. Note that ROUNDROBIN by design queries the possible query sets of $\mathcal{Q}_{=k}$ in a round-robin fashion, so that the number of different query sets is indeed $|\mathcal{Q}_{=k}| = \binom{n}{k}$.

F.2 Comparison of Sufficient Budgets

In order to compare the derived sufficient budgets of the different algorithms (see Table 3), we consider in the following the setting where the generalized Condorcet winner coincides with the generalized Borda winner. In addition we assume that the limit statistic $S_{i|Q}$ for each arm $i \in \mathcal{A}$ has always the same difference to the limit of the optimal arm $S_{i^*|Q}$ if $i^* \in Q$. More precisely, for each arms $i \in \mathcal{A}$ and each query set $Q \in \mathcal{Q}_{\leq k}$ we have $\Delta_{i|Q} = \Delta$ for some fixed $\Delta > 0$. In this way, the γ-dependent term present in the sufficient budget for each algorithm is simply $\left\lceil \bar{\gamma}^{-1}\left(\frac{\Delta}{2}\right)\right\rceil$. As a consequence, we can neglect this term as it has no influence on the differences in the desired budgets for the various algorithms and the remaining term based on the product of the number of rounds, i.e. R, and the number of partitions in round 1, i.e. P_1, is driving the (rough) sufficient budget bounds (see Table 2). However, the number of partitions in round 1 is the same for all algorithms, so that we can neglect this term as well. With a slight abuse of denotation, we refer to this remainder term simply as the sufficient budget in the following. With these considerations, it is easy to see that RoundRobin requires the highest sufficient budget even for moderate sizes of n if k is sufficiently lower than n. To get an impression how the sufficient budget behaves for the more sophisticated algorithms based on the successive elimination strategy, we plot these in Figure 2 as curves depending on the number of arms n for different subset sizes k. Note, that in contrast to CSWS and CSH, the sufficient budget of CSR is higher for bigger subset sizes k, since only a smaller proportion of all arms is discarded after each round. In the case $k = 2$ the number of rounds are all the same, so that consequently the sufficient budget is the same for all three algorithms.

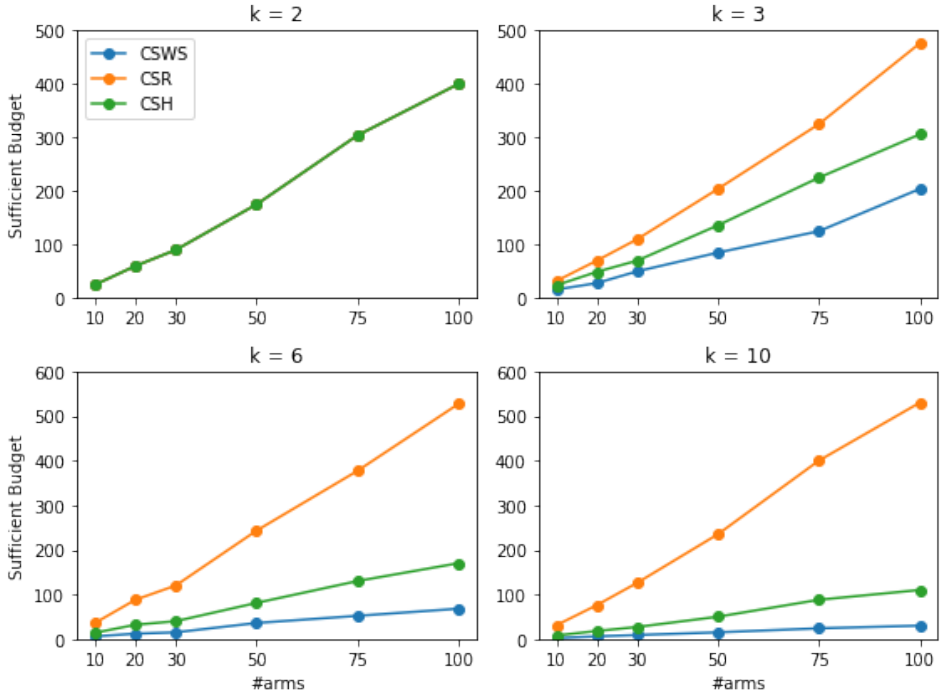

Figure 2: Comparison of required budget for our proposed algorithms for different values of the number of arms n and the subset size k.

F.3 Applications to Stochastic Settings

In Table 4 the sufficient budgets for our proposed algorithms in the stochastic setting with reward feedback and preference-based feedback are listed. Note, that these results are simply derived by applying Corollary 5.1 and resp. Corollary 5.2 with the specific instantiations of R and P_r for our algorithms (see Tables 2 and 3).

Table 4: Comparison of the sufficient budgets for our proposed algorithms CSWS, CSR and CSH in the reward and preference-based setting.

Alg.	Budget in reward setting
CSWS	$\frac{1}{\epsilon^2}\ln\left(\frac{k(\lceil\log_k(n)\rceil+1)}{\delta}\ln\left(\frac{k(\lceil\log_k(n)\rceil+1)}{\epsilon\delta}\right)\right)\cdot(\lceil\log_k(n)\rceil+1)\left\lceil\frac{n}{k}\right\rceil$
CSR	$\frac{1}{\epsilon^2}\ln\left(\frac{k\left(\left\lceil\log_{1-\frac{1}{k}}\left(\frac{1}{n}\right)\right\rceil+k-1\right)}{\delta}\ln\left(\frac{k\left(\left\lceil\log_{1-\frac{1}{k}}\left(\frac{1}{n}\right)\right\rceil+k-1\right)}{\epsilon\delta}\right)\right)\cdot\left(\left\lceil\log_{1-\frac{1}{k}}\left(\frac{1}{n}\right)\right\rceil+k-1\right)\left\lceil\frac{n}{k}\right\rceil$
CSH	$\frac{1}{\epsilon^2}\ln\left(\frac{k(\lceil\log_2(n)\rceil+\lceil\log_2(k)\rceil)}{\delta}\ln\left(\frac{k(\lceil\log_2(n)\rceil+\lceil\log_2(k)\rceil)}{\epsilon\delta}\right)\right)\cdot(\lceil\log_2(n)\rceil+\lceil\log_2(k)\rceil)\left\lceil\frac{n}{k}\right\rceil$
Alg.	**Budget in preference-based setting**
CSWS	$\frac{1}{\epsilon^2}\ln\left(\frac{\lceil\log_k(n)\rceil+1}{\delta\epsilon^4}\right)\cdot(\lceil\log_k(n)\rceil+1)\left\lceil\frac{n}{k}\right\rceil$
CSR	$\frac{1}{\epsilon^2}\ln\left(\frac{\left\lceil\log_{1-\frac{1}{k}}\left(\frac{1}{n}\right)\right\rceil+k-1}{\delta\epsilon^4}\right)\cdot\left(\left\lceil\log_{1-\frac{1}{k}}\left(\frac{1}{n}\right)\right\rceil+k-1\right)\left\lceil\frac{n}{k}\right\rceil$
CSH	$\frac{1}{\epsilon^2}\ln\left(\frac{\lceil\log_2(n)\rceil+\lceil\log_2(k)\rceil}{\delta\epsilon^4}\right)\cdot(\lceil\log_2(n)\rceil+\lceil\log_2(k)\rceil)\left\lceil\frac{n}{k}\right\rceil$

G Further Experiments

In the following, we present some further experiments comparing our proposed algorithms with each other on synthetic data including a detailed description of the data generation and the experiment setting.

G.1 Synthetic Data

For each $Q \in \mathcal{Q}_{\leq k}$ with $Q = \{i_1, \dots, i_{|Q|}\}$ we consider the case where the observation vector $\mathbf{o}_Q$ is a random sample from a multivariate Gaussian distribution with mean $\mu_Q = (\mu_{i_1|Q}, \dots, \mu_{i_{|Q|}|Q})^\top$ and a diagonal covariance matrix $\mathrm{diag}(\sigma_{i_1|Q}, \dots, \sigma_{i_{|Q|}|Q})$. Here, $\mu_{i_j|Q}$ are values in $[0, 1]$ for $i_j \neq i^*$ and $\sigma_{i|Q}$ in $[0.05, 0.2]$ (all randomly sampled). For any Q with $i^* \in Q$ we set $\mu_{i^*|Q} = \max_{j \in Q, j \neq i^*} \mu_{j|Q} + \varepsilon$ for some $\varepsilon > 0$, which ensures (A2) to hold for the expected values. In our experiments we always use a value of $\varepsilon = 0.1$. In the following we vary the values of $n \in \{50, 100\}$, $k \in \{2, 4, 6, 8, 10\}$ and $B \in \{50, 100, 200, 300, 500\}$.

We consider a reward setting and use the empirical mean as the statistic (see Section 5). We do not force the generalized Borda winner to be the same as the generalized Condorcet winner, but they naturally coincidence in most of the runs by sampling the observation vector as defined above.

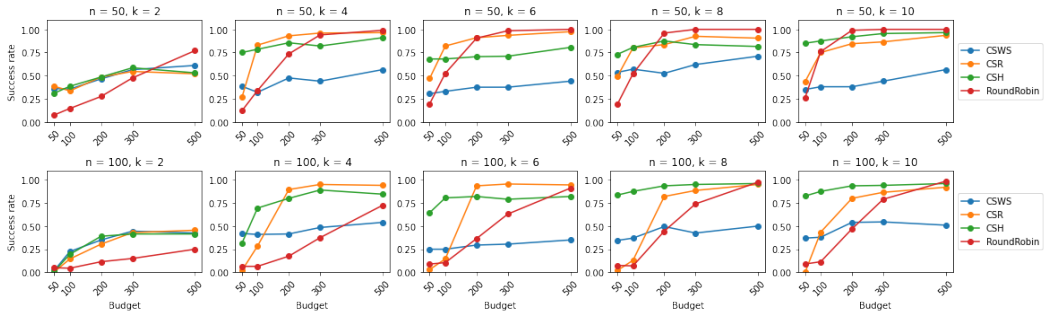

Figure 3: Success rates of our proposed algorithms for varying n, k and budget B in the reward setting.

The success rates of our proposed algorithms for identifying i^* given a budget B are shown in Figure 3. It is visible, that in particular for the challenging scenario, where the budget B and the subset size k are small and the number of arms n is large, both CSH and CSR perform well. Especially CSH has overall a solid performance.

Reward setting. In contrast to the experiments with reward feedback shown in the main paper, we try in the following experiments to force the generalized Borda winner to be different from the generalized Condorcet winner. For this purpose, we fix one random arm $i^*_{\mathcal{B}} \in [n] \backslash \{i^*\}$ as the prospective generalized Borda winner and set its expected value to $\mu_{i^*_{\mathcal{B}}|Q} = \max_{j \in Q, j \neq i^*_{\mathcal{B}}} \mu_{j|Q} + 2\epsilon$ for any $Q \in \mathcal{Q}_{\leq k}$ with $i^*_{\mathcal{B}} \in Q$ and $i^* \notin Q$. Thus, $i^*_{\mathcal{B}}$ is likely the generalized Borda winner and is different from the generalized Condorcet winner. Since our goal is to find the generalized Condorcet winner i^*, ROUNDROBIN will probably fail most of the times in finding i^*. This is due to the fact that ROUNDROBIN focuses on identifying $i^*_{\mathcal{B}}$, i.e., the the generalized Borda winner, which, however, does not coincidence with the generalized Condorcet winner i^*.

This suspicion is confirmed by the results of the experiments shown in Figure 4 illustrating the empirical success rates for finding the generalized Condorcet winner in the setting described above. Except for some cases where the subset size k is relatively large in comparison to the total number of arms, such that the generalized Condorcet winner is already contained in most of the seen subsets and hence is automatically also the generalized Borda winner, ROUNDROBIN performs poorly in finding the generalized Condorcet winner and is always outperformed by the algorithms based on the combinatorial successive elimination strategy in Section 4.1.

Preference-based setting with different GCW and GBW. In the preference-based setting we ignore the explicit numerical values of the observation vector and only use the information which

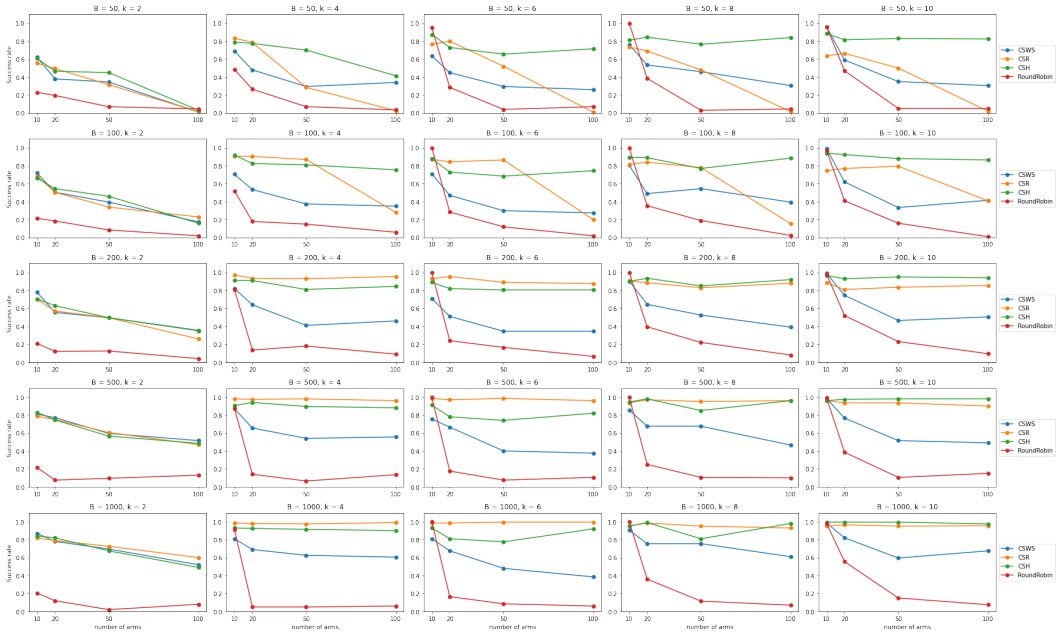

Figure 4: Success rates of our proposed algorithms for varying n, k and budget B in the reward setting with different generalized Condorcet winner and generalized Borda winner.

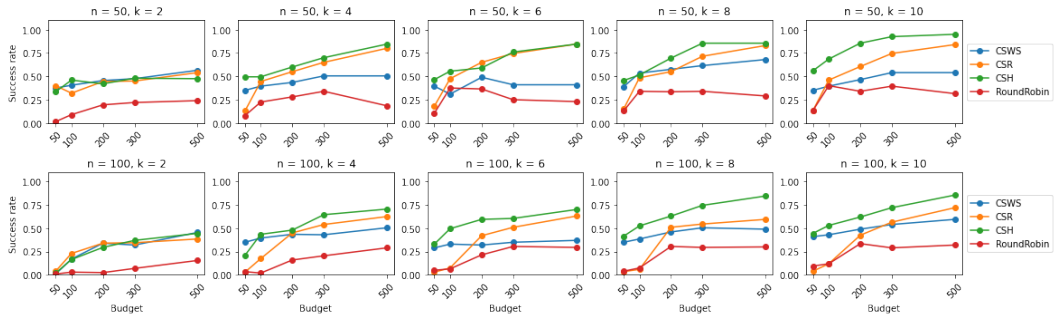

Figure 5: Success rates of our proposed algorithms for varying n, k and budget B in the preference-based setting with different generalized Condorcet winner and generalized Borda winner.

arm was (not) the winner, i.e., which had (not) the highest observation value in the query set used , formally $s_{i_j|Q}(t) = \frac{1}{t}\sum_{s=1}^{t} \mathbb{1}\{o_{i_j|Q}(s) = \max_{i=i_1,\dots i_{|Q|}} o_{i|Q}(s)\}$. Additionally, we fix one arm $i^*_{\mathcal{B}} \in [n]\backslash\{i^*\}$ and set $\mu_{i^*_{\mathcal{B}}|Q} = \max_{j\in Q, j\neq i^*_{\mathcal{B}}} \mu_{j|Q} + 2\varepsilon$ for any Q with $i^*_{\mathcal{B}} \in Q$ and $i^* \notin Q$. In this way, $i^*_{\mathcal{B}}$ is the generalized Borda winner and different from i^*.

The success rates of our proposed algorithms for identifying i^* in this setting are shown in Figure 5. As expected our methods outperform ROUNDROBIN in all scenarios.

Preference-based setting. We now investigate the case, in which we do not force the generalized Borda winner and the generalized Condorcet winner to be different, thus they will naturally coincidence in most of the cases. This is achieved by considering the problem configuration as in the reward setting specified in Section 6, and ignoring the explicit numerical values (as in the preference-setting above).

The resulting success rates for finding the generalized Condorcet winner illustrated in Figure 6 are similar to the results in the reward setting for matching generalized Condorcet winner and generalized

Borda winner. This means that, in particular, when the budget is small, the number of arms is large and the subset size is small, the algorithms following the combinatorial successive elimination strategy outperform ROUNDROBIN. Note that this setting is arguably the most relevant setup for practical applications. Moreover, Figure 6 illustrates the natural effect one would expect for the number of arms n on success rates, namely that success rates decrease with a larger number of arms.

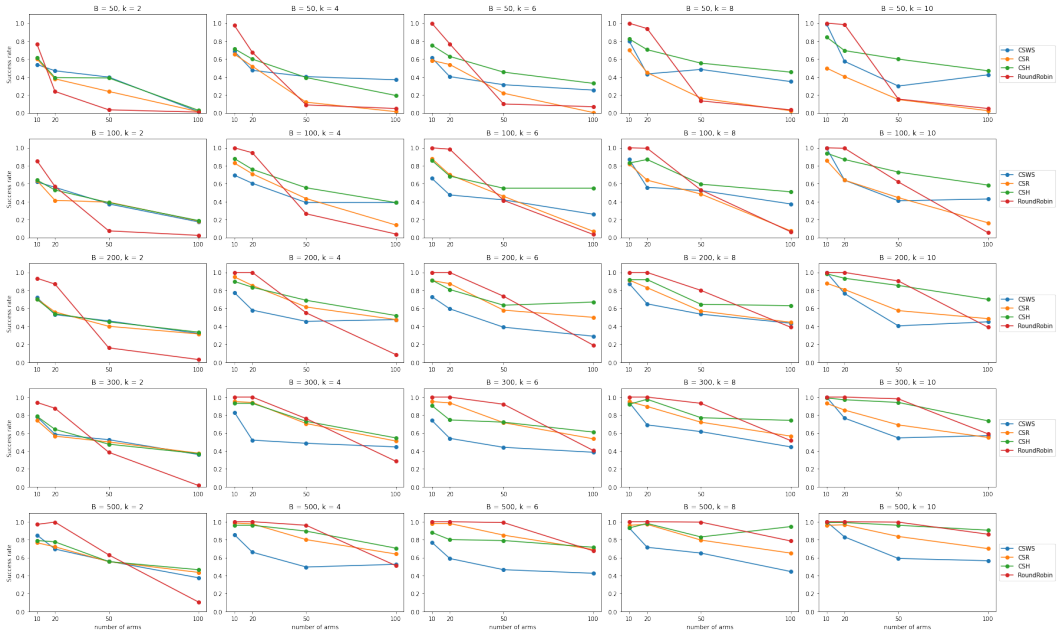

Figure 6: Success rates of our proposed algorithms for varying n, k and budget B in the preference-based setting with (mostly) matching generalized Condorcet winner and generalized Borda winner.

G.2 Statistics beyond the Arithmetic Mean

We consider in the following the reward setting, where each observation is random sampled from the following distribution

$$\boldsymbol{o}_Q(t) \sim \mathcal{N}\left(\begin{pmatrix} \mu_{1|Q} \\ \vdots \\ \mu_{|Q||Q} \end{pmatrix}, \begin{pmatrix} \sigma_{1|Q} \\ \vdots \\ \sigma_{|Q||Q} \end{pmatrix}\right)$$

for $\mu_{i|Q}$ is sampled randomly from $[0, 1]$ and $\sigma_{i|Q}$ from $[0.05, 0.2]$ for each arm $i \in Q$.

Median An alternative to the arithmetic mean would be to measure the quality of the arms by the median of the seen observations. In particular, when the observations are prone to outliers, the median provides a more robust statistic: $s_{i|Q}(t) = \textsc{Median}(o_{i|Q}(1), \ldots, o_{i|Q}(t))$ for each arm $i \in Q$. The results for this setting are illustrated in Figure 7.

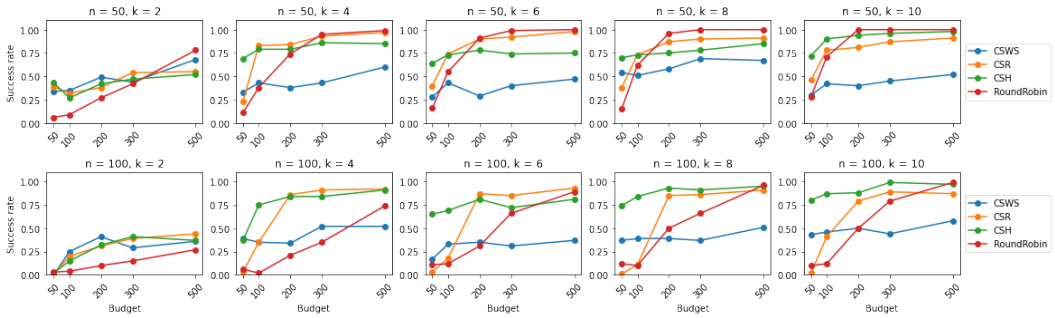

Figure 7: Success rates of our proposed algorithms for varying n, k and budget B in the rewards setting with (mostly) matching generalized Condorcet winner and generalized Borda winner and using the median as the statistic.

Power-Mean Another possibility is to use the so called power-mean, which is a compromise between the maximum and the arithmetic mean for a (multi)set of observations. Since the arithmetic mean is known to underestimate the true quality of an arm, while the maximum overestimates it, the power mean is often a good compromise, as it lies between the two. It is defined by $s_{i|Q}(t) = \left(\frac{1}{t}\sum_{t'=1}^{t} o_{i|Q}(t')^q\right)^{1/q}$ for each arm $i \in Q$ and a fixed $q \in \mathbb{N}$. We use in the following $q = 2$. The results for this setting are illustrated in Figure 8.

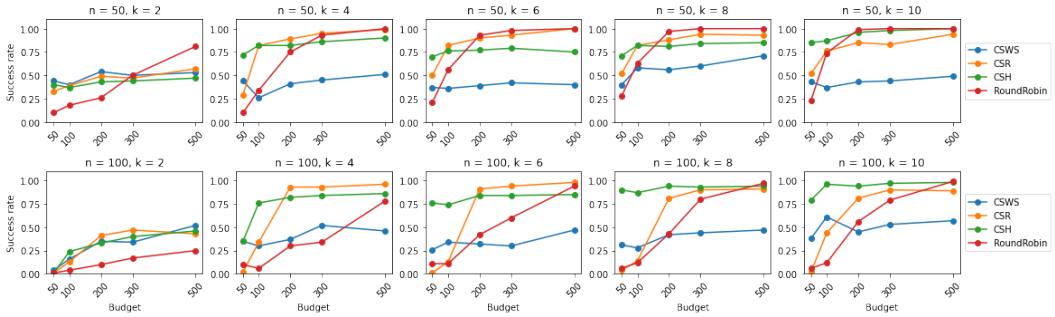

Figure 8: Success rates of our proposed algorithms for varying n, k and budget B in the rewards setting with (mostly) matching generalized Condorcet winner and generalized Borda winner and using the power mean as the statistic.