Identifying Copeland Winners in Dueling Bandits with Indifferences

Viktor Bengs
LMU Munich
MCML

Björn Haddenhorst
Paderborn University

Eyke Hüllermeier
LMU Munich
MCML

Abstract

We consider the task of identifying the Copeland winner(s) in a dueling bandits problem with ternary feedback. This is an underexplored but practically relevant variant of the conventional dueling bandits problem, in which, in addition to strict preference between two arms, one may observe feedback in the form of an indifference. We provide a lower bound on the sample complexity for any learning algorithm finding the Copeland winner(s) with a fixed error probability. Moreover, we propose POCOWISTA, an algorithm with a sample complexity that almost matches this lower bound, and which shows excellent empirical performance, even for the conventional dueling bandits problem. For the case where the preference probabilities satisfy a specific type of stochastic transitivity, we provide a refined version with an improved worst case sample complexity.

1 INTRODUCTION

Dueling bandits (Yue and Joachims, 2009) or the more general problem class of preference-based bandits (Bengs et al., 2021) is a practically relevant variant of the standard reward-based multi-armed bandits (Lattimore and Szepesvári, 2020), in which a learner seeks to find in a sequential decision making process an optimal arm (choice alternative) by selecting two (or more) arms as its action and obtaining feedback in the form of a noisy preference over the selected arms. The setting is motivated by a broad range of applications, where no numerical rewards for the actions are obtained and only comparisons of arms (choice alternatives) are possible as actions. In information retrieval systems with human preference judgments (Clarke et al., 2021), for example, humans choose their most preferred choice alternative among the two (or more) retrieved choice alternatives (e.g., text passages, movies, etc.). Another example is the analysis of voting behavior, in which voters express their preferences over pairs of political parties or candidates (Brady and Ansolabehere, 1989). The rationale for these types of applications is that humans are generally better at coherently expressing their preference for two choice alternatives than at reliably assessing those two on a numerical scale (Carterette et al., 2008; Li et al., 2021).

Even though there is a large body of literature on dueling or preference-based bandits (see Sui et al. (2018); Bengs et al. (2021)) covering various variants or aspects of the initial setting, little attention has been paid to the variant, in which a learner might observe an indifference between the selected arms for comparison as the explicit feedback. In practice, however, this type of feedback is quite common, especially when the preference feedback is provided by a human. In the human preference judgment example above, the human might be indifferent between the two retrieved choice alternatives, as the two are considered equally good/mediocre/bad, so neither is chosen. Similarly, voters might be indifferent between two political parties or candidates, or two athletes resp. sport teams competing against each other might draw.

In several areas of preference-based learning, extensions of existing models or methods have been considered to appropriately incorporate such indifferences. Notable examples are the recent extensions of established probabilistic ranking models (Firth et al., 2019; Turner et al., 2020; Henderson, 2022), the field of partial label ranking (Alfaro et al., 2022, 2023), an extension of the established label ranking problem, or preference-based Bayesian optimization (Dewancker et al., 2018; Nguyen et al., 2021). However, the field of preference-based bandits lags behind in this regard, as the only work appears to be Gajane et al. (2015)

Proceedings of the 27$^{\text{th}}$ International Conference on Artificial Intelligence and Statistics (AISTATS) 2024, Valencia, Spain. PMLR: Volume 238.

that considers an adversarial learning scenario for regret minimization.

Motivated by this gap in the literature on preference-based bandits, we consider the *stochastic* dueling bandits problem with indifferences (or ternary feedback) for the task of identifying an optimal arm as quickly as possible, i.e., with as few queried feedback observations as possible. Similarly as in the conventional dueling bandits problem with binary feedback (i.e., strict preferences) specifying the notion of optimality of an arm raises some issues. Indeed, the most natural notion of an optimal arm would be an arm, which is non-dominated by any other arm in terms of the probability of being strictly preferred or indifferent. This notion corresponds to the Condorcet winner (CW) in the conventional setting, where the probability of observing indifferences is zero. However, it is well known that such an arm may not exist in general in the conventional setting, an issue obviously shared by the adopted CW for our considered setting. In contrast to the conventional CW, the adopted CW does not even guarantee uniqueness of the optimal arm[1].

The non-existence issue of the CW has led several authors to consider alternative notions for the optimality of arms guaranteed to exist in any case. Most of them have their roots in tournament solutions used in social choice and voting theory (Brandt et al., 2015, 2016) or game theory (Owen, 2013). One popular alternative is the Copeland set (Copeland, 1951) defined as the set of choice alternatives (arms) with the highest Copeland score. In the absence of indifferent preferences, the Copeland score of a choice alternative is the number of choice alternatives it dominates in terms of the pairwise preference probability. In settings like ours, i.e., where indifferences might be present, the Copeland score of a choice alternative is the sum of (i) the number of choice alternatives it *strictly* dominates, and (ii) half the number of choice alternatives it is most likely indifferent to. Again, both definitions coincide for the conventional setting, where the probability of observing an indifference is zero.

As the term itself suggests, there can be several arms in the Copeland set, each of which is called a Copeland winner (COWI). Despite this non-uniqueness issue, the main advantages of considering the Copeland set are that (i) it is guaranteed to exist and (ii) that it consists only of the Condorcet winner(s) in case of its (their) existence. Moreover, the majority of alternative optimality notions from tournament solutions are in fact supersets of the Copeland set (see Ramamohan et al. (2016)).

[1]As an example, consider the case of three arms, where each arm has a probability of 1/2 of being preferred over or indifferent to another arm, respectively.

Outline and Contributions
In this paper, we make the following contributions:

Introduction of the problem (Sec. 2.1): We are the first to consider the problem of finding a Copeland winner in a dueling bandits problem with possible indifference observations. This problem variant is of practical relevance, especially for applications involving human feedback.

Lower bounds (Sec. 2.2): We provide lower bounds on the sample complexity for any learning algorithm to find a Copeland winner with a fixed confidence in this problem variant. Our lower bounds imply as a special case a long-missing lower bound for the conventional dueling bandits problem.

A practically useful (Sec. 5) and near-optimal algorithm (Sec. 3): We construct a learning algorithm, POCOWISTA, which selects pairs of arms in a challenge-tournament-like fashion and exploits prior-posterior-ratio martingale confidence sequences for determining the Copeland scores of the underlying arms in an asymptotically optimal way. In numerical simulations, we find that it performs quite well even for conventional dueling bandits.

Relaxing quadratic dependency (Sec. 4): We show that in the case of an underlying transitivity of the preference relations, the update formula of POCOWISTA can be modified (called TRA-POCOWISTA) such that the quadratic dependency w.r.t. the number of arms n of the worst-case sample complexity can be relaxed to a log-linear dependency.

2 PROBLEM FORMULATION

We consider a set $\mathcal{A}$ of $n \in \mathbb{N}_{\geq 2}$ available choice alternatives that we refer to as arms and simply identify them by their index: $\mathcal{A} = \{1, \ldots, n\}$. The learning process consists of consequential rounds, in which the learner performs an action leading to some feedback for its action. More precisely, the learner's action in round t corresponds to choosing a pair of arms $(i_t, j_t) \in \mathcal{A} \times \mathcal{A}$, for which it observes noisy feedback o_t with three possible realizations:

- $i_t \succ j_t$, i.e., arm i_t is strictly preferred over arm j_t,
- $i_t \prec j_t$, i.e., arm j_t is strictly preferred over arm i_t,
- $i_t \cong j_t$, i.e., neither i_t is strictly preferred over j_t nor the opposite (indifference between i_t and j_t).

Each of the three possible explicit observations is determined by one of the following matrices $P^{\succ}, P^{\prec}, P^{\cong} \in [0,1]^{n \times n}$. Here, the (i,j)-th entry of $P^{\succ}$ (or $P^{\prec}$) denoted by $P^{\succ}_{i,j}$ (or $P^{\prec}_{i,j}$) specifies the probability of observing a strict preference of i over j (or j over i), while the (i,j)-th entry of $P^{\cong}$ denoted

by $P^{\cong}_{i,j}$ specifies the probability of observing an indifference between i and j. Apparently, it holds that $P^{\succ}_{i,j} + P^{\cong}_{i,j} + P^{\prec}_{i,j} = 1$ for any $i, j \in \mathcal{A}$, and consequently any dueling bandits problem with indifferences is uniquely determined by one of its strict preference probability matrices, since $P^{\succ}_{j,i} = P^{\prec}_{i,j}$. The Copeland score of arm $i \in \mathcal{A}$ is

$$\begin{aligned} \mathrm{CP}(i) = &\sum\nolimits_{j \neq i} 1_{[\![P^{\succ}_{i,j} > \max\{P^{\prec}_{i,j}, P^{\cong}_{i,j}\}]\!]} \\ &+ \tfrac{1}{2} \sum\nolimits_{j \neq i} 1_{[\![P^{\cong}_{i,j} > \max\{P^{\succ}_{i,j}, P^{\prec}_{i,j}\}]\!]}, \end{aligned} \tag{1}$$

where $1_{[\![\cdot]\!]}$ denotes the indicator function. Thus, an arm gets a score of one for each arm it dominates and half a point for each arm it is indifferent to. The Copeland set (or the set of Copeland winners) consists of all arms with maximal Copeland score, denoted by

$$\mathcal{C} = \{i \in \mathcal{A} \,|\, \mathrm{CP}(i) = \max\nolimits_j \mathrm{CP}(j)\}. \tag{2}$$

The goal of the learner is to find an element of the Copeland set, i.e., a Copeland winner (COWI), by performing as few actions as possible. To this end, the learner may decide to stop the learning process at some round τ and output an arm $\hat{i} \in \mathcal{A}$ deemed to be a COWI. Because of the stochasticity of the observed feedback, the learner can make mistakes, so any reasonable learner should meet a theoretical guarantee that its output is correct. Thus, if $\delta \in (0, 1)$ is the desired bound on the error probability, it should hold that $\mathbb{P}(\hat{i} \notin \mathcal{C}) \leq \delta$. Additionally, the learner's stopping time τ should be as small as possible (in expectation or with high probability), while still guaranteeing the latter error probability bound.

2.1 Related Work

The conventional dueling bandits problem, i.e., without indifferences, has been introduced as a practical variant of the classical multi-armed bandit (MAB) problem by Yue and Joachims (2009). Initially, the problem has been studied intensively for the task of regret minimization under the assumption of an existing CW (Yue et al., 2012; Zoghi et al., 2014, 2015b; Komiyama et al., 2015) to specify a target arm. Due to the non-existence issue of the CW, several works have considered alternative optimality notions for the target arm such as the COWI (Zoghi et al., 2015a; Komiyama et al., 2016; Wu and Liu, 2016) or other tournament solutions (Ramamohan et al., 2016).

Similar to the classical MAB problem, there has been also much research interest in the pure exploration task in the conventional dueling bandits problem, where the goal is to identify the target arm as quickly as possible. Again, the majority of works have used the CW to specify the target arm (Karnin, 2016; Mohajer et al., 2017; Ren et al., 2019, 2020) or a generalized variant for multi-dueling settings (Haddenhorst et al., 2021a; Brandt et al., 2022). Although the identification of a COWI has been in fact studied before the aforementioned works (Busa-Fekete et al., 2013; Urvoy et al., 2013; Busa-Fekete et al., 2014), none of these considered the case where indifferences are observed as explicit feedback.

In practical use cases, observing indifferences as pairwise preference feedback plays an important role, as has been highlighted by a number of papers with applications ranging from sports (Tiwisina and Külpmann, 2019), medicine (Li et al., 2021), crowdsourcing tasks (Asudeh et al., 2015; Clarke et al., 2021) to information retrieval (Yan et al., 2022). In addition, there is recent work in preference-based Bayesian optimization (González et al., 2017) that incorporates indifference feedback into the optimization procedure (Dewancker et al., 2018; Nguyen et al., 2021).

In the dueling resp. preference-based bandit literature, however, the only work which has considered indifferences is Gajane et al. (2015), which addresses an adversarial learning scenario for the task of regret minimization. To this end, the Sparring algorithm (Ailon et al., 2014), which uses two bandit algorithms for the classic MAB setting (each of which selects one arm of the pair to be dueled), is used with two instantiations of Exp3 (Auer et al., 2002) suitably modified to account for preference feedback.

2.2 Lower bounds

For convenience, write $\mathbf{P} = ((P^{\succ}_{i,j}, P^{\cong}_{i,j}, P^{\prec}_{i,j}))_{i<j}$, and denote by $P^{(1)}_{i,j}, P^{(2)}_{i,j}, P^{(3)}_{i,j}$ the order statistics of $P^{\succ}_{i,j}, P^{\cong}_{i,j}$ and $P^{\prec}_{i,j}$. For any learning algorithm A for the dueling bandits problem with indifferences let $\tau^{\mathrm{A}}(\mathbf{P})$ denote its number of samples (or actions), when started on a dueling bandits problem characterized by $\mathbf{P}$. If $\mathbf{P}$ is clear from the context, we simply write τ^{A}. We denote by

$$\begin{aligned} I(j) &= \{i \in \mathcal{A} \setminus \{j\} \,|\, P^{\cong}_{i,j} > \max\{P^{\prec}_{i,j}, P^{\succ}_{i,j}\}\} \\ L(j) &= \{i \in \mathcal{A} \setminus \{j\} \,|\, P^{\succ}_{i,j} > \max\{P^{\prec}_{i,j}, P^{\cong}_{i,j}\}\} \end{aligned}$$

the set of all indifferent resp. superior arms to some arm $j \in \mathcal{A}$. We write $\mathcal{C}(\mathbf{P})$ for the Copeland set in (2) to highlight its dependence on the underlying dueling bandits problem with indifferences. If $\mathbf{P}$ is fixed, let $d_j = \max_i \mathrm{CP}(i) - \mathrm{CP}(j)$ be the difference between the largest Copeland score and arm j's Copeland score. Finally, $\mathrm{KL}((p_1, p_2, p_3), (q_1, q_2, q_3))$ is the Kullback-Leibler divergence between two categorical random variables with parameters (p_1, p_2, p_3) and (q_1, q_2, q_3), while we use the common notation $\mathrm{kl}(p, q)$

for the Kullback-Leibler divergence between two categorical random variables with parameters $(p, 1-p)$ and $(q, 1-q)$, i.e., two Bernoulli distributions with success parameters p and q.

The following theorem provides a lower bound on the sample complexity of any learner for the dueling bandits problem (i) without, or (ii) with indifferences. We provide the proof as well as a more sophisticated but more technical variant of (ii) in the appendix (Sec. B), which is also non-trivial on some exceptional instances where this simpler bound fails to be positive.

Theorem 2.1. *If* A *correctly identifies the COWI with confidence* $1-\delta$*, then*

$$\mathbb{E}[\tau^{\mathrm{A}}(\mathbf{P})] \geq \ln \tfrac{1}{2.4\delta} \sum\nolimits_{j \in \mathcal{A} \setminus \{i^*\}} C_j \min_{k \in L(j) \cup I(j)} \tfrac{1}{D_{j,k}(\mathbf{P})},$$

where $\mathcal{C}(\mathbf{P}) = \{i^*\}$ *and*

(i) $D_{j,k}(\mathbf{P}) \coloneqq \mathrm{kl}(P^{\succ}_{j,k}, 1 - P^{\succ}_{j,k})$ *and* $C_j \coloneqq \max\left\{ \frac{|L(j)| 1_{[\![|L(j)| \geq d_j + 1]\!]}}{d_j + 1}, \frac{(|L(j)|-1) 1_{[\![i^* \in L(j)]\!]}}{|L(j)| + d_j - 2} \right\}$ *for all* $\mathbf{P}$ *without indifferences* ($P^{\cong} = \mathbf{0} \in [0,1]^{n \times n}$) *and* $\min_{i<j} \min\{P^{\succ}_{i,j}, 1 - P^{\succ}_{i,j}\} > 0$.

(ii) $D_{j,k}(\mathbf{P}) \coloneqq \max\{\mathrm{KL}^{(1)}_{j,k}, \mathrm{KL}^{(2)}_{j,k}\}$ *with*
$\mathrm{KL}^{(1)}_{j,k} = \mathrm{KL}((P^{\succ}_{j,k}, P^{\cong}_{j,k}, P^{\prec}_{j,k}), (P^{\cong}_{j,k}, P^{\succ}_{j,k}, P^{\prec}_{j,k}))$,
$\mathrm{KL}^{(2)}_{j,k} = \mathrm{KL}((P^{\succ}_{j,k}, P^{\cong}_{j,k}, P^{\prec}_{j,k}), (P^{\prec}_{j,k}, P^{\cong}_{j,k}, P^{\succ}_{j,k}))$,

$$C_j = \max_{(i,l) \in \Psi(j)} \frac{\binom{|I(j)|}{i}\binom{|L(j)|}{l}}{\binom{|I(j)|-1}{i-1}\binom{|L(j)|}{l} 1_{[\![i \geq 1]\!]} + \binom{|I(j)|}{i}\binom{|L(j)|-1}{l-1} 1_{[\![l \geq 1]\!]}},$$

$$\Psi(j) \coloneqq \big\{(i,l) \in \{0, \ldots, |I(j)|\} \times \{0, \ldots, |L(j)|\} \mid i + 2l \geq 2d_j + 1\big\}$$

for any $\mathbf{P}$ *with* $\min_{j,k} \min\{P^{\succ}_{j,k}, P^{\cong}_{j,k}, P^{\prec}_{j,k}\} > 0$.

An easier-to-interpret form of the bounds would be

$$\ln \tfrac{1}{2.4\delta} \sum\nolimits_{j \in \mathcal{A} \setminus \{i^*\}} \sum\nolimits_{k \in L(j) \cup I(j)} \tfrac{1}{D_{j,k}(\mathbf{P})}, \qquad (3)$$

that resembles the well-known lower bounds in the standard multi-armed bandits literature (e.g., Theorem 4 in Kaufmann et al. (2016). However, note that we get a lower bound for the latter as follows:

$$\sum_{k \in L(j) \cup I(j)} \tfrac{1}{D_{j,k}(\mathbf{P})} \geq \min_{k \in L(j) \cup I(j)} \tfrac{1}{D_{j,k}(\mathbf{P})} \cdot |L(j) \cup I(j)|.$$

In light of this, the C_j terms on the right-hand side of Theorem 2.1 can be seen as lower bounds for the $|L(j) \cup I(j)|$ terms. Even though the bounds in Theorem 2.1 are "only" lower bounds for the "natural" lower bounds in (3), they are sufficient to derive, for example, the expected $\Omega(n^2)$ worst-case bound. Lower bounds of that type are common in the dueling bandit literature for best arm identification tasks, due to the combinatorial nature of the problem, e.g., see Theorem 5.2 in Haddenhorst et al. (2021a).

Moreover, note that the lower bound in Theorem 2.1 (i) is non-trivial (i.e., positive) for any admissible $\mathbf{P}$, and in a worst-case sense on instances $\mathbf{P}$ with $\min_{i<j} |P^{\succ}_{i,j} - 1/2| > \Delta$ and $\mathrm{CP}(i^*) = n/2 + o(n)$ of order $\Omega(n^2/\Delta^2 \ln 1/\delta)$. Thus, existing approaches for COWI identification in a conventional dueling bandits setting are nearly optimal (cf. Table 7 in Bengs et al. (2021)). The bounds in (i) and (ii) are consistent in the sense that if $\max_{j \neq i^*} |I(j)| = 0$, the factor C_j appearing in (ii) is exactly the maximum term appearing in (i), and similarly $\Omega(n^2/\Delta^2 \ln 1/\delta)$ samples might be necessary in expectation to identify the COWI of an indifferent $\mathbf{P}$ with $\min_{i<j} |P^{(1)}_{i,j} - P^{(2)}_{i,j}| > \Delta$.

3 LEARNING ALGORITHM

The key to designing an efficient learning algorithm for the COWI identification task in general is to determine as quickly as possible which arms are potential COWIs and then restrict the sampling mechanism to the set of potential COWIs. In light of this, an important component to ensure the efficiency of this sampling procedure is to decide quickly and reliably the allocation of the Copeland scores (cf. (1)). The latter can be seen in fact as finding the mode of a ternary distribution: For a fixed arm pair, say i, j, the dueling feedback is governed by a ternary distribution $P_{i,j}$ with probabilities $P^{\succ}_{i,j}, P^{\cong}_{i,j}$ and $P^{\prec}_{i,j}$ for $i \succ j, i \cong j$ and $i \prec j$. The Copeland score of arm i is then the number of ternary distributions $P_{i,j}$ (for varying $j \neq i$) for which $P^{\succ}_{i,j}$ is the mode and half the number of these for which $P^{\cong}_{i,j}$ is the mode. Thus, it seems reasonable to use an efficient estimation procedure for correctly identifying the mode of a discrete distribution.

3.1 Mode Identification

In Jain et al. (2022) the PPR-1v1 algorithm is proposed for mode identification by combining the 1-versus-1-principle from multiclass classification with prior-posterior-ratio (PPR) martingale confidence sequences (Waudby-Smith and Ramdas, 2020), which provide anytime confidence sequences on a specific parameter of a distribution.

The prior-posterior-ratio martingale. Let $(P_\theta)_{\theta \in \Theta}$ be a family of distributions with parameter space Θ. Let π_0 be a prior distribution on Θ and π_t be the posterior distribution after observing $t \in \mathbb{N}$ many i.i.d. observations $X_1, \ldots, X_t$ according to P_{θ^*} for some (unknown) $\theta^* \in \Theta$. The prior-posterior ratio (PPR) is given by $R_t(\theta) = \pi_0(\theta)/\pi_t(\theta)$ for $\theta \in \Theta$. If π_0

assigns non-zero mass everywhere on Θ, then

$$C_t = \{\theta \mid R_t(\theta) < 1/\delta\} = \left\{\theta \,\middle|\, \delta < {}^{\pi_t(\theta)}/_{\pi_0(\theta)}\right\} \quad (4)$$

is a $(1-\delta)$-confidence sequence for θ^*, i.e., it holds that $\mathbb{P}(\exists t \in \mathbb{N} : \theta^* \notin C_t) \leq \delta$ (see Waudby-Smith and Ramdas (2020)). The name PPR martingale stems from the fact that $(R_t(\theta^*))_{t=1}^T$ is a martingale w.r.t. the canonical filtration of $X_1, X_2, \ldots, X_T$ for any $T \in \mathbb{N}$.

PPR-Bernoulli test. As an exemplary application of this result, consider the case of Bernoulli distributions for P_θ, i.e., $P_\theta = \mathrm{Ber}(\theta)$ and $\Theta = [0,1]$, for which one seeks to determine as quickly as possible (i.e., in a sequential manner) whether $\theta^* > 1/2$ or $\theta^* < 1/2$ holds. Note that this is equivalent to identifying the mode of the Bernoulli distribution $\mathrm{Ber}(\theta^*)$. Using as the (conjugate) prior a Beta distribution with both parameters being 1, one obtains the uniform distribution on Θ, which fulfills the requirements for C_t in (4) to be an anytime confidence sequence for θ^*. Further, the posterior distribution after observing t many i.i.d. Bernoulli samples is a Beta distribution with parameters $(s_t(1)+1, s_t(0)+1)$, where $s_t(x)$ is the number of observed $x \in \{0,1\}$. Thus, one can stop the sampling process as soon as $1/2$ is not contained in C_t and declaring the x with the most observations as the mode. Formally, the PPR-Bernoulli test is to declare x as the mode if $f_{\mathrm{Beta}}(1/2; s_t(x)+1, s_t(\neg x)+1) \leq \delta$ and $s_t(x) \geq s_t(\neg x)$, where $f_{\mathrm{Beta}}(\cdot; \alpha, \beta)$ is the probability density function of an (α, β)-Beta distribution.

PPR-1-versus-1 test. At first sight, it is tempting to instantiate the PPR martingale approach with the Dirichlet distribution as the conjugate prior for a categorical distribution to identify the latter's mode by stopping the sampling process similarly as for the PPR-Bernoulli test. However, if the categorical distribution has more than two categories, say $c_1, c_2, \ldots, c_K$ with $K \in \mathbb{N}_{\geq 2}$, then it is difficult to obtain a closed-form criterion as in the PPR-Bernoulli test, so that costly numerical computations are needed.

In light of this, in Jain et al. (2022) it is proposed to reduce the mode identification problem to multiple PPR-Bernoulli tests using the 1-versus-1-principle from multiclass classification. Thus, a PPR-Bernoulli test is simultaneously conducted for each pair of categories (c_i, c_j) with $i \neq j$ and an error probability of $\delta/(K-1)$, each of which uses only the number of occurrences of c_i and c_j and ignoring the remaining ones. If there exists a category that has won all of its tests, it is declared to be the mode. This procedure is equivalent to monitoring only the PPR-Bernoulli test for the pair of categories $(c_{t(1)}, c_{t(2)})$, where $c_{t(1)}$ resp. $c_{t(2)}$ is the category that has the most resp. second most occurrences after observing t samples. Consequently, the prior-posterior-ratio-1-versus-1 (PPR-1v1) test decision is to declare $c_{t(1)}$ as the mode if $f_{\mathrm{Beta}}(1/2; s_{t(1)}+1, s_{t(2)}+1) \leq \delta/(K-1)$, where $s_{t(x)}$ denotes the number of occurrences of $c_{t(x)}$.

The probability of making an error with this test procedure, i.e., not declaring the true mode as the mode of the underlying categorical distribution, is bounded by means of a union bound by δ. Moreover, the above PPR-1v1 test is asymptotically optimal in the sense that the ratio of its expected stopping time and the lower bound on the expected stopping time for a fixed categorical distribution tends to one for the error probability δ tending to zero (Jain et al., 2022).

We can transfer this test procedure to the case of identifying the mode of a ternary distribution $P_{i,j}$, which is equivalent to finding the mode of a categorical distribution with three categories $c_1 := \text{“}i \succ j\text{”}, c_2 := \text{“}i \cong j\text{”}$ and $c_3 := \text{“}i \prec j\text{”}$. The explicit PPR-1v1 procedure for this case is given in Algo. 1, where $S_{(1)} \geq S_{(2)} \geq S_{(3)}$ is the order statistics of S_1, S_2, S_3.

Algorithm 1 PPR-1v1

Input: Arms i and j, error prob. $\delta \in (0,1)$
Initialization: $S = (S_1, S_2, S_3) \leftarrow (0,0,0)$
while TRUE do
  Compare i and j
  Observe $o \in \{i \succ j, i \cong j, i \prec j\}$
  if $o = i \succ j$ **then**
    $S_1 \leftarrow S_1 + 1$
  else if $o = i \cong j$ **then**
    $S_2 \leftarrow S_2 + 1$
  else
    $S_3 \leftarrow S_3 + 1$
  end if
  if $f_{\mathrm{Beta}}(1/2; S_{(1)}+1, S_{(2)}+1) \leq \delta/2$ **then**
    return $\mathrm{argmax}_{k=1,2,3} S_k$
  end if
end while

3.2 POCOWISTA

Guided by the design idea above, i.e., determining as quickly as possible which arms are potential COWIs and then restricting the sampling mechanism to the set of potential COWIs, we propose the POCOWISTA (POtential COpeland WInner STays Algorithm) in Algo. 2. For each arm two Copeland scores are maintained: (i) the current Copeland score $\widehat{CP}(\cdot)$, which is determined (using the PPR-1v1 algorithm) by the duels already contested, and (ii) the potential Copeland score $\overline{CP}(\cdot)$, which is determined by both the duels already contested and the duels not yet contested, which add an optimistic bonus to the current Copeland score. For the calculation of this optimistic bonus, a set of arms $D(\cdot)$ is maintained for each arm, which includes

the arms that have already been compared to (dueled with) the respective arm as well as the arm itself. The optimistic bonus for an arm, say i, is then the size of the set of arms not compared to i so far, i.e., $|\mathcal{A} \setminus D(i)| = n - |D(i)|$.

Algorithm 2 POCOWISTA

1: **Input:** Set of arms $\mathcal{A}$, error prob. $\delta \in (0,1)$
2: **Initialization:** $e \leftarrow 1$ and for each $i \in \mathcal{A}$ set
   $D(i) \leftarrow \{i\}$ (set of already compared arms)
   $\widehat{CP}(i) \leftarrow 0$ (current Copeland score)
   $\overline{CP}(i) \leftarrow n-1$ (potential Copeland score)
3: **while** $\nexists i$ s.t. $\widehat{CP}(i) \geq \overline{CP}(j)\ \forall j \in \mathcal{A} \setminus \{i\}$ **do**
4: $\quad i_e = \operatorname{argmax}_{i \in \mathcal{A}} \overline{CP}(i)$
5: $\quad j_e = \operatorname{argmax}_{j \in \mathcal{A} \setminus D(i_e)} \widehat{CP}(j)$
6: $\quad k \leftarrow$ PPR-1v1$(i_e, j_e, \delta/\binom{n}{2})$
7: $\quad$ Scores-Update(i_e, j_e, k)
8: $\quad e \leftarrow e+1$
9: **end while**
10: **return** $\operatorname{argmax}_{i \in \mathcal{A}} \widehat{CP}(i)$

Algorithm 3 Scores-Update

1: **Input:** Arms i, j, ternary decision $k \in \{1,2,3\}$
2: **if** $k=1$ **then**
3: $\quad \widehat{CP}(i) \leftarrow \widehat{CP}(i) + 1$
4: **else if** $k=2$ **then**
5: $\quad \widehat{CP}(i) \leftarrow \widehat{CP}(i) + 1/2$, $\widehat{CP}(j) \leftarrow \widehat{CP}(j) + 1/2$
6: **else**
7: $\quad \widehat{CP}(j) \leftarrow \widehat{CP}(j) + 1$
8: **end if**
9: $D(i) \leftarrow D(i) \cup \{j\}$, $D(j) \leftarrow D(j) \cup \{i\}$
10: $\overline{CP}(i) \leftarrow n - |D(i)| + \widehat{CP}(i)$
11: $\overline{CP}(j) \leftarrow n - |D(j)| + \widehat{CP}(j)$

The algorithm proceeds in epochs, in each of which it uses as the "first" arm the current incumbent in terms of the potential Copeland scores (line 4, Algo. 2) and as the "second" arm the one with the highest current Copeland score among those not yet compared to the first (line 5, Algo. 2). These two arms are successively dueled against each other until the mode of their feedback distribution is identified by means of the PPR-1v1 algorithm (line 6, Algo. 2), which leads to an update of their Copeland scores (line 7, Algo. 2): The current Copeland score of the dominating arm is increased by one (lines 2–3,6–7, Algo. 3), while in case of an indifference both obtain half a point (lines 4–5, Algo. 3). In addition, the set of already compared arms of the two arms is extended by the other one (line 9, Algo. 3) and the potential Copeland scores are updated as well (lines 10–11, Algo. 3).

The update procedure corresponds to the end of an epoch, which leads to the start of a new epoch, unless there is an arm whose current Copeland score is not smaller than any other potential Copeland score (line 3, Algo. 2). In such a case the arm is a COWI and returned by the algorithm (line 10).

For the sampling complexity of POCOWISTA we derive the following result (proof in Sec. C).

Theorem 3.1. *Let* A := POCOWISTA. *For any dueling bandits problem with indifferences characterized by* $\mathbf{P} = ((P^{\succ}_{i,j}, P^{\cong}_{i,j}, P^{\prec}_{i,j}))_{i<j}$, *such that there exists no pair* $i, j \in \mathcal{A}$ *with* $i \neq j$ *and* $P^{\succ}_{i,j} = P^{\succ}_{j,i} = 1/3$, *it holds*

$$\mathbb{P}\big(\hat{i}_{\mathrm{A}} \in \mathcal{C}(\mathbf{P}) \text{ and } \tau^{\mathrm{A}}(\mathbf{P}) \leq t(\mathbf{P}, \delta)\big) \geq 1 - \delta,$$

where $t(\mathbf{P}, \delta) \leq \sum_{i<j} t_0\big((P^{\succ}_{i,j}, P^{\cong}_{i,j}, P^{\prec}_{i,j}), \delta/\binom{n}{2}\big)$,

$$t_0\big((p_1,p_2,p_3),\delta\big) = \tfrac{c_1 p_{(1)}}{(p_{(1)}-p_{(2)})^2} \ln\Big(\tfrac{\sqrt{2\,c_2} p_{(1)}}{\sqrt{\delta}(p_{(1)}-p_{(2)})}\Big), \quad (5)$$

$p_{(1)} \geq p_{(2)} \geq p_{(3)}$ *is the order statistic of* $p_1, p_2, p_3 \in [0,1]$, $c_1 = 194.07$, *and* $c_2 = 79.86$.

Here, $t_0\big((p_1,p_2,p_3),\delta\big)$ is the sample complexity of PPR-1v1 to identify the mode of a (categorical) distribution with probabilities p_1, p_2, p_3 with an error probability of at most δ (see Theorem 9 in Jain et al. (2022)). Since in the worst case, one needs to ensure for all pairs of arms i, j that the mode of the ternary distribution $P_{i,j}$ is correctly identified, the PPR-1v1 algorithm is used with $\delta/\binom{n}{2}$ for its error probability (line 6 of Algo. 2) to ensure that the overall error probability δ is not exceeded. If $\mathbf{P}$ is such that $\min_{i<j} |P^{(1)}_{i,j} - P^{(2)}_{i,j}| > \Delta$, we see that POCOWISTA's sample complexity is in $O(n^2 \ln(n)/\Delta^2 \ln 1/\delta)$ which almost matches the worst-case lower bound in Theorem 2.1 by Pinsker's inequality (see Theorem F.4 in Haddenhorst et al. (2021b)). Thus, for the conventional dueling bandit case it has the same worst-case sample complexity as existing algorithms such as SAVAGE (Urvoy et al., 2013) or PBR-CCSO (Busa-Fekete et al., 2013).

A comparison of the sample complexity per pair, say i and j, of all the three algorithms *only* with respect to the number of arms n, error probability δ and the gap between the arms $\Delta_{i,j} = |P^{\succ}_{i,j} - 1/2|$ for the conventional dueling bandits is given in the following table:

POCOWISTA	SAVAGE	PBR-CCSO
$\frac{1}{\Delta^2_{i,j}} \ln\big(\frac{n}{\sqrt{\delta}} \cdot \frac{1}{\Delta_{i,j}}\big)$	$\frac{1}{\Delta^2_{i,j}} \ln\big(\frac{n}{\delta} \cdot \frac{1}{\Delta_{i,j}}\big)$	$\frac{1}{\Delta^2_{i,j}} \ln\big(\frac{n^2}{\delta} \cdot \frac{1}{\Delta_{i,j}}\big)$

There, we see that POCOWISTA and SAVAGE have a better dependence on n compared to PBR-CCSO, while POCOWISTA additionally has a better dependence on δ than the other two. Accordingly, POCOWISTA is expected to have a better sample complexity in practical applications than SAVAGE, which in turn has a better sample complexity than PBR-CCSO. This is supported by our experimental results in Sec. 5.

3.3 Reduction to Conventional Dueling Bandits

Another tempting idea would be to modify existing algorithms for the conventional dueling bandit setting to accept indifference feedback as follows. Whenever an indifference is observed for the two chosen arms i_t and j_t, this feedback is changed to either $i_t \succ j_t$ or $i_t \prec j_t$, by flipping a (fair) coin. However, this reduction approach has two key issues:

The reduction can change the target arm. Assume the preference probabilities for the initial problem (with indifferences) for three arms to be as follows:

$$\begin{array}{lll} P(1 \succ 2) = 0.5 & P(1 \sim 2) = 0.1 & P(1 \prec 2) = 0.4 \\ P(1 \succ 3) = 0.1 & P(1 \sim 3) = 0.75 & P(1 \prec 3) = 0.15 \\ P(2 \succ 3) = 0.5 & P(2 \sim 3) = 0.1 & P(2 \prec 3) = 0.4 \end{array}$$

The reduction transforms the probabilities to be:

$$\begin{array}{ll} P(1 \succ 2) = 0.55 & P(1 \prec 2) = 0.45 \\ P(1 \succ 3) = 0.475 & P(1 \prec 3) = 0.525 \\ P(2 \succ 3) = 0.55 & P(2 \prec 3) = 0.45 \end{array}$$

Thus, the Copeland scores for the initial problem are 1.5 for arm 1, 1 for 2, and 0.5 for 3, so the Copeland winner set consists of only arm 1. However, for the reduced problem, all three arms have a Copeland score of 1, so all of them will be considered to be Copeland winners. Thus, any modified algorithm using this reduction will, roughly speaking, err in 2 out of 3 cases.

The reduction can change the gap. Using the previous example, we can see that learning in principle can be made harder when using the reduction: The minimal gap for the reduced problem is 0.05, while it was 0.1 initially.

4 TRANSITIVE PREFERENCES

Although the sample complexity of POCOWISTA almost matches the lower bound, it is in some sense unsatisfactory as it is log-linear with respect to the number of actions (the number of arm pairs) in the worst case, i.e., $O(n^2 \ln(n))$. In light of this, we consider in this section structural properties on the preference probabilities such that the dependency of POCOWISTA's sample complexity with respect to n is reduced to $O(n \ln(n))$ in the worst case. To this end, we leverage the commonly used stochastic transitivity properties in conventional dueling bandits, which in essence assume that if an arm i is preferred over arm j and arm j over arm k, then i is also preferred over arm k. As this notion of transitivity merely considers the part of the feedback regarding the strict preferences, we augment it with the notion of IP-transitivity and PI-transitivity as well as transitivity of indifferences (Hansson and Grüne-Yanoff, 2022) in order to account for the possibility of observing indifferences.

Definition 4.1. A dueling bandits problem with indifferences characterized by $\mathbf{P} = ((P^{\succ}_{i,j}, P^{\cong}_{i,j}, P^{\prec}_{i,j}))_{i<j}$ is called *transitive* if for any distinct $i, j, k \in \mathcal{A}$ it holds:

1. Transitivity of strict preference.
   If $P^{\succ}_{i,j} > \max(P^{\prec}_{i,j}, P^{\cong}_{i,j})$ and $P^{\succ}_{j,k} > \max(P^{\prec}_{j,k}, P^{\cong}_{j,k})$, then $P^{\succ}_{i,k} > \max(P^{\prec}_{i,k}, P^{\cong}_{i,k})$.
2. IP-transitivity.
   If $P^{\cong}_{i,j} > \max(P^{\prec}_{i,j}, P^{\succ}_{i,j})$ and $P^{\succ}_{j,k} > \max(P^{\prec}_{j,k}, P^{\cong}_{j,k})$, then $P^{\succ}_{i,k} > \max(P^{\prec}_{i,k}, P^{\cong}_{i,k})$.
3. PI-transitivity.
   If $P^{\succ}_{i,j} > \max\left(P^{\prec}_{i,j}, P^{\cong}_{i,j}\right)$and $P^{\cong}_{j,k} > \max\left(P^{\prec}_{j,k}, P^{\succ}_{j,k}\right)$, then $P^{\succ}_{i,k} > \max\left(P^{\prec}_{i,k}, P^{\cong}_{i,k}\right)$.
4. Transitivity of indifference.
   If $P^{\cong}_{i,j} > \max(P^{\prec}_{i,j}, P^{\succ}_{i,j})$ and $P^{\cong}_{j,k} > \max(P^{\prec}_{j,k}, P^{\succ}_{j,k})$, then $P^{\cong}_{i,k} > \max(P^{\prec}_{i,k}, P^{\succ}_{i,k})$.

For the conventional dueling bandit setting, i.e., where $P^{\cong} = \mathbf{0} \in [0,1]^{n \times n}$, the transitivity property in Def. 4.1 is equivalent to strict weak stochastic transitivity.

Under the assumption of a transitive dueling bandits problem with indifferences as stipulated by Def. 4.1, one can modify POCOWISTA and in particular the update rule to take the transitivity property into account. The resulting algorithm will be called TRA-POCOWISTA and for sake of completeness, we state its pseudo-code in Algo.4. The main difference to POCOWISTA is to maintain three additional sets for each arm: the set of defeated arms $W(\cdot)$, the set of indifferent arms $I(\cdot)$, and the set of inferior arms $L(\cdot)$ (line 2). These sets are then updated at the end of each round according to the implications 1-4. of transitivity and in turn used to update the current and potential Copeland score of the two arms involved by using Algo. 5 in line 7.

Algorithm 4 TRA-POCOWISTA

1: **Input:** Set of arms $\mathcal{A}$, error prob. $\delta \in (0,1)$
2: **Initialization:** $e \leftarrow 1$ and for each $i \in \mathcal{A}$
   $D(i) \leftarrow \{i\}$, $\widehat{CP}(i) \leftarrow 0$, $\overline{CP}(i) \leftarrow n-1$
   $W(i) \leftarrow \emptyset$ (set of defeated arms)
   $I(i) \leftarrow \emptyset$ (set of indifferent arms)
   $L(i) \leftarrow \emptyset$ (set of superior arms)
3: **while** $\nexists i$ s.t. $\widehat{CP}(i) \geq \overline{CP}(j)\, \forall j \in \mathcal{A} \setminus \{i\}$ **do**
4: $i_e = \operatorname{argmax}_{i \in \mathcal{A}} \overline{CP}(i)$
5: $j_e = \operatorname{argmax}_{j \in \mathcal{A} \setminus D(i_e)} \widehat{CP}(j)$
6: $k \leftarrow$ PPR-1v1$(i_e, j_e, \delta/n)$
7: TRANSITIVE-SCORE-UPDATE(i_e, j_e, k)
8: $e \leftarrow e+1$
9: **end while**
10: **return** $\operatorname{argmax}_{i \in \mathcal{A}} \widehat{CP}(i)$

More specifically, the transitivity of strict preferences and the IP-transitivity imply that once it is ensured that an arm i either dominates another one j or is

indifferent to it, then i will also dominate all arms dominated by j. PI transitivity implies that if i dominates j, all arms that are indifferent to j are also dominated by i. Thus, i's current Copeland score can be increased, in addition to the update due to the domination of j (increasing by one), by the number of arms that are dominated/indifferent by/to j (line 3). In case of indifference between i and j the current Copeland scores can be increased, in addition to the update due to the indifference (increasing both by one half), by the number of arms that are dominated/indifferent to the other (lines 9–10). Conversely, the potential Copeland score of the dominated arm j can be decreased, in addition to the update due to the domination by i, by the number of arms that are dominating/indifferent to i (line 18)

Algorithm 5 TRANSITIVE-SCORE-UPDATE

1: **Input:** Arms i, j, $k \in \{1, 2, 3\}$
2: **if** $k = 1$ **then**
3: $\quad \widehat{CP}(i) \leftarrow \widehat{CP}(i) + |W(j) \cup I(j)| + 1$
4: $\quad W(i) \leftarrow W(i) \cup W(j) \cup I(j) \cup \{j\}$
5: $\quad D(i) \leftarrow D(i) \cup W(j) \cup I(j) \cup \{j\}$
6: $\quad L(j) \leftarrow L(j) \cup L(i) \cup I(i) \cup \{i\}$
7: $\quad D(j) \leftarrow D(j) \cup L(i) \cup I(i) \cup \{i\}$
8: **else if** $k = 2$ **then**
9: $\quad \widehat{CP}(i) \leftarrow \widehat{CP}(i) + |W(j)| + 1/2(1 + |I(j)|)$
10: $\quad \widehat{CP}(j) \leftarrow \widehat{CP}(j) + |W(i)| + 1/2(1 + |I(i)|)$
11: $\quad W(i) \leftarrow W(i) \cup W(j), W(j) \leftarrow W(i)$
12: $\quad L(i) \leftarrow L(i) \cup L(j), L(j) \leftarrow L(i)$
13: $\quad I(i) \leftarrow I(i) \cup I(j) \cup \{j\}, I(j) \leftarrow I(i) \cup I(j) \cup \{i\}$
14: $\quad D(i) \leftarrow D(i) \cup D(j),\ D(j) \leftarrow D(i)$
15: **else**
16: $\quad$ Same as for $k = 1$ with i and j reversed
17: **end if**
18: Same steps as line 10 and 11 in Algo. 3

Although there are three additional data structures to manage, the memory complexity is still linear w.r.t. the number of arms n, as in POCOWISTA. In addition, we obtain the following improved bound on the sample complexity w.r.t. n (proof in Sec. D).

Theorem 4.2. *Let* A := TRA-POCOWISTA. *For any dueling bandits problem with indifferences as in Theorem 3.1 which in addition is transitive according to Def. 4.1, it holds that*

$$\mathbb{P}\big(\hat{i}_{\mathrm{A}} \in \mathcal{C}(\mathbf{P}) \text{ and } \tau^{\mathrm{A}}(\mathbf{P}) \leq \tilde{t}(\mathbf{P}, \delta)\big) \geq 1 - \delta,$$

where $\tilde{t}(\mathbf{P}, \delta) = \sum_{e=1}^{E} t_0\big((P^{\succ}_{i_e,j_e}, P^{\cong}_{i_e,j_e}, P^{\prec}_{i_e,j_e}), \delta/n\big)$, t_0 *is as in* (5) *and* $E \leq n$.

5 EXPERIMENTS

Since, to the best of our knowledge, there are no algorithms for identification tasks in dueling bandits problems with indifferences, we resort in the following experiments to the conventional dueling bandits problem, i.e., where $P^{\cong} = \mathbf{0}$. First, we compare POCOWISTA with SAVAGE (Urvoy et al., 2013) and PBR-CCSO (Busa-Fekete et al., 2013), which are the only available methods for the task of identifying a Copeland winner. Since the Copeland set boils down to a singleton set consisting of the Condorcet winner (CW) in case of the latter's existence, we compare POCOWISTA also with the state-of-the-art algorithm SELECT (Mohajer et al., 2017) and DKWT (Haddenhorst et al., 2021a) for identifying a CW.

Copeland Winner Identification. Given a strict preference probability matrix $P^{\succ} \in \mathcal{P}(n)$ for n arms, with $\mathcal{P}(n) = \big\{P^{\succ} \in [0,1]^{n \times n} \,\big|\, P^{\succ}_{i,j} + P^{\succ}_{j,i} = 1, \forall i \neq j\big\}$, we consider first the setting of identifying a COWI of $P^{\succ}$. We distinguish between two classes of a (conventional) dueling bandits problem:

$$\mathcal{P}_1(n) := \big\{P^{\succ} \in \mathcal{P}(n) \,\big|\, |P^{\succ}_{i,j} - 1/2| \geq 0.1\, \forall i \neq j\big\},$$
$$\mathcal{P}_2(n) := \big\{P^{\succ} \in \mathcal{P}(n) \,\big|\, |P^{\succ}_{i,j} - 1/2| \geq 0.05, \; |P^{\succ}_{i,j} - 1| \leq 0.3,\ \forall i \neq j\big\}.$$

The class $\mathcal{P}_1(n)$ are easy problems, as the gap parameters are quite large, while $\mathcal{P}_2(n)$ consists of more difficult problem instances, where the pairwise probabilities are close to $1/2$ making it more difficult to identify whether one arm dominates the other or vice versa.

For both classes, a strict preference probability matrix is repeatedly selected uniformly at random, and then used to generate the feedback for the learning algorithms for a total of 100 repetitions. Note that all algorithms are parameter-free in the sense that they only need the desired error probability δ and the number of arms n, but no other adjustable (hyper-)parameters.

The two leftmost plots in Figure 1 show the average sample complexity (and their standard deviation in brackets) of the four algorithms for the two problem classes with $n = 20$ arms and different choices for the error probabilities δ. All algorithms have an empirical error probability of zero for each δ. This is due to the Bonferroni correction used by each algorithm to ensure that each pairwise comparison is correctly decided, making the overall decision quite conservative. As can be seen from the plots, the average sample complexities of our algorithms are by a magnitude smaller than for the existing methods and the same holds for their standard deviations. TRA-POCOWISTA has even a clear improvement over POCOWISTA, although the considered problem instances do not necessarily satisfy the transitivity property in Def. 4.1.

Condorcet Winner Identification. Next, we consider the setting of identifying a Condorcet winner (CW) from a given strict preference probability matrix

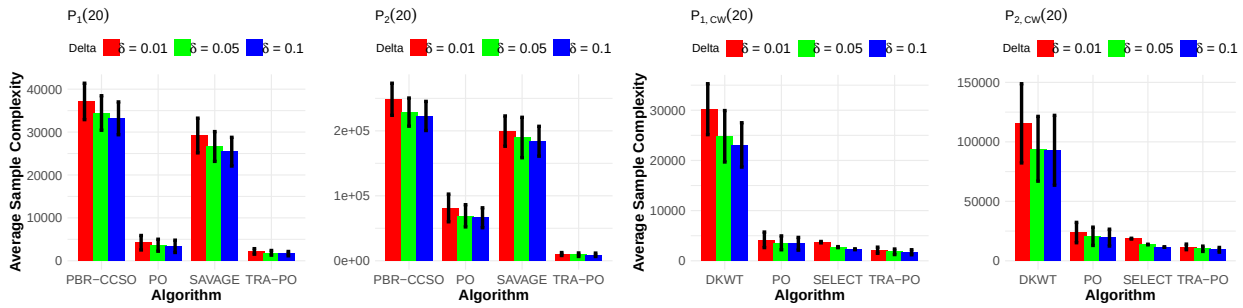

Figure 1: Average sample complexities (and standard deviations) for the considered dueling bandit classes of the considered learning algorithms.

$P^{\succ} \in \mathcal{P}_{\mathrm{CW}}(n)$ for n arms with an existing CW, where $\mathcal{P}_{\mathrm{CW}}(n) := \left\{P^{\succ} \in \mathcal{P}(n) \,\middle|\, \exists \rho : P^{\succ}_{\rho,j} > \frac{1}{2}, \forall j \neq \rho\right\}$. Again, we distinguish between two classes of a (conventional) dueling bandits problem with different difficulties for this task similarly as above by defining $\mathcal{P}_{i,\mathrm{CW}}(n) := \mathcal{P}_i(n) \cap \mathcal{P}_{\mathrm{CW}}(n)$ for $i = 1, 2$. The experimental set-up (i.e., sampling a problem instance and repetition number) is similar as above.

Note that POCOWISTA, TRA-POCOWISTA and DKWT are parameter-free, while SELECT needs as a parameter the number of duels carried out per arm pair. In order to ensure that SELECT fulfills the sought error probability δ this parameter needs to be chosen as a function of $\min_{i<j} |P^{\succ}_{i,j} - 1/2|$ i.e., the *unknown* minimal gap (in the conventional dueling bandit setting). In the following, we use the optimal choice for SELECT's parameter, although this gives it a clear advantage.

The results for $n = 20$ arms and different choices for the error probabilities δ for this experiment setting are reported in the rightmost plots of Figure 1. Again, all algorithms have an empirical error probability of zero due to the Bonferroni correction. The results show that DKWT is inferior to all other three algorithms, while POCOWISTA requires on average a slightly higher sample complexity than SELECT with the optimal choice of its parameter, and TRA-POCOWISTA's sample complexity is the lowest. Nevertheless, the standard deviation of SELECT's sample complexity is lower compared to our algorithms, which is due to the fact that the parameter of SELECT determines exactly how many pairwise comparisons are performed per pair of arms. The variance then arises from the different numbers of arm pairs used in total to arrive at the decision, which varies accordingly due to the random selection of the problem instance in each repetition. It is worth noting that the sampled problem instances do not necessarily satisfy the transitivity property in Def. 4.1, so that the results are again in favor of TRA-POCOWISTA.

6 CONCLUSION

In this paper, we considered an extension of the dueling bandits problem, where feedback in the form of an indifference can be observed in addition to the binary strict preference feedback. We have studied the pure exploration problem of finding a Copeland winner within a fixed confidence setting, for which we provided instance-dependent lower bounds on the sample complexity. Furthermore, we proposed POCOWISTA, which can solve this task almost optimally for worst-case scenarios, and extended it to TRA-POCOWISTA for the case where the preference probabilities satisfy a certain type of stochastic transitivity that lead to improved sample complexity bounds.

For future work, it would be interesting to investigate the considered extension of the dueling bandit problem in a regret minimization setting, or to combine it with other variants or extensions of the dueling bandits problem such as the multi-dueling setting (Saha and Gopalan, 2020) or the non-stationary preference variants (Saha and Gupta, 2022; Kolpaczki et al., 2022; Buening and Saha, 2023; Suk and Agarwal, 2023) or contextualized variants (Saha, 2021; Bengs et al., 2022; Saha and Krishnamurthy, 2022).

Since our motivation for extending the dueling bandits problem stemmed from real-world examples, a more in-depth experimental study in such practical application areas would certainly be a desirable avenue for future work, e.g., in algorithm configuration or learning-to-rank problems for which preference-based bandit algorithms have been used before (Brost et al., 2016; Schuth et al., 2016; Oosterhuis et al., 2016; Zhao and King, 2016; El Mesaoudi-Paul et al., 2020; Brandt et al., 2023).

References

Ailon, N., Karnin, Z., and Joachims, T. (2014). Reducing dueling bandits to cardinal bandits. In *Proceedings of the International Conference on Machine*

Learning (ICML), pages 856–864.

Alfaro, J. C., Aledo, J. A., and Gámez, J. A. (2022). Integrating bayesian network classifiers to deal with the partial label ranking problem. In *International Conference on Probabilistic Graphical Models (PGM)*, volume 186 of *Proceedings of Machine Learning Research*, pages 337–348. PMLR.

Alfaro, J. C., Aledo, J. A., and Gámez, J. A. (2023). Pairwise learning for the partial label ranking problem. *Pattern Recognition*, 140:109590.

Asudeh, A., Zhang, G., Hassan, N., Li, C., and Záruba, G. V. (2015). Crowdsourcing pareto-optimal object finding by pairwise comparisons. In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management, (CIKM)*, pages 753–762. ACM.

Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. E. (2002). The nonstochastic multiarmed bandit problem. *SIAM journal on computing*, 32(1):48–77.

Bengs, V., Busa-Fekete, R., El Mesaoudi-Paul, A., and Hüllermeier, E. (2021). Preference-based online learning with dueling bandits: A survey. *The Journal of Machine Learning Research*, 22(7):1–108.

Bengs, V., Saha, A., and Hüllermeier, E. (2022). Stochastic contextual dueling bandits under linear stochastic transitivity models. In *Proceedings of the International Conference on Machine Learning (ICML)*, volume 162, pages 1764–1786. PMLR.

Brady, H. E. and Ansolabehere, S. (1989). The nature of utility functions in mass publics. *American Political Science Review*, 83(1):143–163.

Brandt, F., Brill, M., and Harrenstein, P. (2016). Tournament solutions. In *Handbook of Computational Social Choice*, pages 57–84. Cambridge University Press.

Brandt, F., Dau, A., and Seedig, H. G. (2015). Bounds on the disparity and separation of tournament solutions. *Discrete Applied Mathematics*, 187:41–49.

Brandt, J., Bengs, V., Haddenhorst, B., and Hüllermeier, E. (2022). Finding optimal arms in non-stochastic combinatorial bandits with semi-bandit feedback and finite budget. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pages 20621–20634.

Brandt, J., Schede, E., Haddenhorst, B., Bengs, V., Hüllermeier, E., and Tierney, K. (2023). AC-Band: A combinatorial bandit-based approach to algorithm configuration. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 12355–12363. AAAI Press.

Brost, B., Seldin, Y., Cox, I. J., and Lioma, C. (2016). 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.

Buening, T. K. and Saha, A. (2023). ANACONDA: An improved dynamic regret algorithm for adaptive non-stationary dueling bandits. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 3854–3878. PMLR.

Busa-Fekete, R., Szörényi, B., and Hüllermeier, E. (2014). PAC rank elicitation through adaptive sampling of stochastic pairwise preferences. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 1701–1707.

Busa-Fekete, R., Szörényi, B., Weng, P., Cheng, W., and Hüllermeier, E. (2013). Top-k selection based on adaptive sampling of noisy preferences. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 1094–1102.

Carterette, B., Bennett, P. N., Chickering, D. M., and Dumais, S. T. (2008). Here or there. In *Advances in Information Retrieval: 30th European Conference on IR Research, ECIR*, volume 4956 of *Lecture Notes in Computer Science*, pages 16–27. Springer.

Clarke, C. L., Vtyurina, A., and Smucker, M. D. (2021). Assessing top-k preferences. *ACM Transactions on Information Systems (TOIS)*, 39(3):1–21.

Copeland, A. H. (1951). A reasonable social welfare function. *Seminar on Applications of Mathematics to Social Sciences. University of Michigan, Ann Arbor*.

Dewancker, I., Bauer, J., and McCourt, M. (2018). Sequential preference-based optimization. *arXiv preprint arXiv:1801.02788*.

El Mesaoudi-Paul, A., Weiß, D., Bengs, V., Hüllermeier, E., and Tierney, K. (2020). Pool-based realtime algorithm configuration: A preselection bandit approach. In *International Conference on Learning and Intelligent Optimization*, pages 216–232. Springer.

Firth, D., Kosmidis, I., and Turner, H. (2019). Davidson-Luce model for multi-item choice with ties. *arXiv preprint arXiv:1909.07123*.

Gajane, P., Urvoy, T., and Clérot, F. (2015). A relative exponential weighing algorithm for adversarial utility-based dueling bandits. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 218–227.

González, J., Dai, Z., Damianou, A., and Lawrence, N. D. (2017). Preferential Bayesian optimization. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 1282–1291. PMLR.

Haddenhorst, B., Bengs, V., and Hüllermeier, E. (2021a). Identification of the generalized Condorcet winner in multi-dueling bandits. In *Proceedings of Advances in Neural Information Processing Systems (NeurIP S)*, volume 34, pages 25904–25916.

Haddenhorst, B., Bengs, V., and Hüllermeier, E. (2021b). On testing transitivity in online preference learning. *Machine Learning*, 110:2063–2084.

Hansson, S. O. and Grüne-Yanoff, T. (2022). Preferences. *The Stanford Encyclopedia of Philosophy*.

Henderson, D. A. (2022). Modelling and analysis of rank ordered data with ties via a generalized Plackett-Luce model. *arXiv preprint arXiv:2212.08543*.

Jain, S. A., Shah, R., Gupta, S., Mehta, D., Nair, I. J., Vora, J., Khyalia, S., Das, S., Ribeiro, V. J., and Kalyanakrishnan, S. (2022). PAC mode estimation using PPR martingale confidence sequences. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 5815–5852. PMLR.

Karnin, Z. (2016). Verification based solution for structured MAB problems. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, pages 145–153.

Kaufmann, E., Cappé, O., and Garivier, A. (2016). On the complexity of best-arm identification in multi-armed bandit models. *The Journal of Machine Learning Research*, 17(1):1–42.

Kolpaczki, P., Bengs, V., and Hüllermeier, E. (2022). Non-stationary dueling bandits. *arXiv preprint arXiv:2202.00935*.

Komiyama, J., Honda, J., Kashima, H., and Nakagawa, H. (2015). Regret lower bound and optimal algorithm in dueling bandit problem. In *Proceedings of Annual Conference on Learning Theory (COLT)*, pages 1141–1154.

Komiyama, J., Honda, J., and Nakagawa, H. (2016). Copeland dueling bandit problem: Regret lower bound, optimal algorithm, and computationally efficient algorithm. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 1235–1244.

Lattimore, T. and Szepesvári, C. (2020). *Bandit Algorithms*. Cambridge University Press.

Li, K., Tucker, M., Bıyık, E., Novoseller, E., Burdick, J. W., Sui, Y., Sadigh, D., Yue, Y., and Ames, A. D. (2021). ROIAL: Region of interest active learning for characterizing exoskeleton gait preference landscapes. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3212–3218. IEEE.

Mohajer, S., Suh, C., and Elmahdy, A. (2017). Active learning for top-k rank aggregation from noisy comparisons. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 2488–2497.

Nguyen, Q. P., Tay, S., Low, B. K. H., and Jaillet, P. (2021). Top-k ranking Bayesian optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 35, pages 9135–9143.

Oosterhuis, H., Schuth, A., and de Rijke, M. (2016). Probabilistic multileave gradient descent. In *Proceedings of European Conference on Information Retrieval (ECIR)*, pages 661–668.

Owen, G. (2013). *Game Theory*. Emerald Group Publishing.

Ramamohan, S. Y., Rajkumar, A., and Agarwal, S. (2016). Dueling bandits: Beyond Condorcet winners to general tournament solutions. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, pages 1253–1261.

Ren, W., Liu, J., and Shroff, N. (2019). 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.

Ren, W., Liu, J., and Shroff, N. (2020). The sample complexity of best-k items selection from pairwise comparisons. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 8051–8072.

Saha, A. (2021). Optimal algorithms for stochastic contextual preference bandits. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, pages 30050–30062.

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

Saha, A. and Gupta, S. (2022). Optimal and efficient dynamic regret algorithms for non-stationary dueling bandits. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 19027–19049.

Saha, A. and Krishnamurthy, A. (2022). Efficient and optimal algorithms for contextual dueling bandits under realizability. In *International Conference on Algorithmic Learning Theory (ALT)*, volume 167 of *Proceedings of Machine Learning Research*, pages 968–994. PMLR.

Schuth, A., Oosterhuis, H., Whiteson, S., and de Rijke, M. (2016). Multileave gradient descent for fast online learning to rank. In *Proceedings of ACM Inter-*

national Conference on Web Search and Data Mining (WSDM), pages 457–466.

Sui, Y., Zoghi, M., Hofmann, K., and Yue, Y. (2018). Advancements in dueling bandits. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 5502–5510.

Suk, J. and Agarwal, A. (2023). When can we track significant preference shifts in dueling bandits? *arXiv preprint arXiv:2302.06595*.

Tiwisina, J. and Külpmann, P. (2019). Probabilistic transitivity in sports. *Computers & Operations Research*, 112:104765.

Turner, H. L., van Etten, J., Firth, D., and Kosmidis, I. (2020). Modelling rankings in R: the plackettluce package. *Computational Statistics*, 35(3):1027–1057.

Urvoy, T., Clerot, F., Féraud, R., and Naamane, S. (2013). Generic exploration and k-armed voting bandits. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 91–99.

Waudby-Smith, I. and Ramdas, A. (2020). Confidence sequences for sampling without replacement. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 20204–20214.

Wu, H. and Liu, X. (2016). Double Thompson sampling for dueling bandits. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, pages 649–657.

Yan, X., Luo, C., Clarke, C. L., Craswell, N., Voorhees, E. M., and Castells, P. (2022). Human preferences as dueling bandits. *arXiv preprint arXiv:2204.10362*.

Yue, Y., Broder, J., Kleinberg, R., and Joachims, T. (2012). The k-armed dueling bandits problem. *Journal of Computer and System Sciences*, 78(5):1538–1556.

Yue, Y. and Joachims, T. (2009). Interactively optimizing information retrieval systems as a dueling bandits problem. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 1201–1208.

Zhao, T. and King, I. (2016). Constructing reliable gradient exploration for online learning to rank. In *Proceedings of ACM International Conference on Information and Knowledge Management (CIKM)*, pages 1643–1652.

Zoghi, M., Karnin, Z., Whiteson, S., and de Rijke, M. (2015a). Copeland dueling bandits. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, pages 307–315.

Zoghi, M., Whiteson, S., and de Rijke, M. (2015b). Mergerucb: A method for large-scale online ranker evaluation. In *Proceedings of ACM International Conference on Web Search and Data Mining (WSDM)*, pages 17–26.

Zoghi, M., Whiteson, S., Munos, R., and de Rijke, M. (2014). Relative upper confidence bound for the k-armed dueling bandit problem. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 10–18.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes]

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   (b) Complete proofs of all theoretical results. [Yes]
   (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [No]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Not Applicable]
   (b) The license information of the assets, if applicable. [Not Applicable]
   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

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	
$\prec, \succ$	strict preference relation for objects, i.e., $o \succ o'$ (or $o \prec o'$) iff object o is preferred over object o' (or otherwise)
$\cong$	indifference relation for objects, i.e., $o \cong o'$ iff object o is not preferred over object o' and vice versa
$1_{[\cdot]}$	indicator function
$\mathbb{N}$	set of natural numbers (without 0), i.e., $\mathbb{N} = \{1, 2, 3, \dots\}$
$\mathbb{R}$	set of real numbers
$[n]$	the set $\{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$
$p_{(1)}, p_{(2)}, p_{(3)}$	order statistics of values p_1, p_2, p_3, i.e., $p_{(1)} \geq p_{(2)} \geq p_{(3)}$ and $\{p_{(1)}, p_{(2)}, p_{(3)}\} = \{p_1, p_2, p_3\}$
$f_{\text{Beta}}(x; a, b)$	probability density function of the Beta distribution with parameters $a, b > 0$ at point $x \in \mathbb{R}$
$\text{KL}(\mathbf{p}, \mathbf{q})$	Kullback-Leibler divergence for two categorical distributions $\mathbf{p} = (p_1, \dots, p_K) \in [0,1]^K$ and $\mathbf{q} = (q_1, \dots, q_K) \in [0,1]^K$ such that $\sum_{i=1}^K p_i = \sum_{i=1}^K q_i = 1$
$\text{kl}(p, q)$	Kullback-Leibler divergence for two Bernoulli distributions with success probabilities $p, q \in [0,1]$ i.e, $\text{kl}(p,q) = \text{KL}((p, 1-p), (q, 1-q))$
Modeling related	
n	number of arms
$\mathcal{A} = [n]$	set of arms
$P^{\succ}$	strict preference probability matrix with $P^{\succ}_{i,j}$ being the probability of observing $i \succ j$ (element of $[0,1]^{n \times n}$)
$P^{\prec}$	strict preference probability matrix with $P^{\prec}_{i,j}$ being the probability of observing $i \prec j$ (element of $[0,1]^{n \times n}$)
$P^{\cong}$	indifference probability matrix with $P^{\cong}_{i,j}$ being the probability of observing $i \cong j$ (element of $[0,1]^{n \times n}$)
$P^{(1)}_{i,j}, P^{(2)}_{i,j}, P^{(3)}_{i,j}$	order statistic of $P^{\succ}_{i,j}, P^{\cong}_{i,j}$ and $P^{\prec}_{i,j}$, i.e., $P^{(1)}_{i,j} \geq P^{(2)}_{i,j} \geq P^{(3)}_{i,j}$ and $\{P^{(1)}_{i,j}, P^{(2)}_{i,j}, P^{(3)}_{i,j}\} = \{P^{\succ}_{i,j}, P^{\cong}_{i,j}, P^{\prec}_{i,j}\}$
$P_{i,j}$	ternary probability distribution with probabilities $P^{\succ}_{i,j}, P^{\cong}_{i,j}$ and $P^{\prec}_{i,j}$ for $i \succ j, i \cong j$ and $i \prec j$
$\mathbf{P}$	family of ternary distributions characterizing a dueling bandits problem instance with indifferences i.e., $((P^{\succ}_{i,j}, P^{\cong}_{i,j}, P^{\prec}_{i,j}))_{i<j}$
$\Delta(i,j)$	gap of the mode of $P_{i,j}$, i.e., $\Delta(i,j) = \lvert P^{(1)}_{i,j} - P^{(2)}_{i,j} \rvert$
$\text{CP}(j)$; $\text{CP}(\mathbf{P}, j)$	Copeland score of an arm $j \in \mathcal{A}$ (see (1)); for a given problem instance $\mathbf{P}$
$\mathcal{C}$;$\mathcal{C}(\mathbf{P})$	Copeland set (see (2)); for a given problem instance $\mathbf{P}$
d_j;$d_j(\mathbf{P})$	difference between the largest Copeland score and some arm j's Copeland score ; for a given problem instance $\mathbf{P}$
Learner related	
A	a learning for the dueling bandits problem with indifferences
(i_t, j_t)	pair of arms chosen by the learner in round t (action of the learner in round t)
o_t	preference observation in round t for the learner's action in round t either $i_t \succ j_t$, $i_t \cong j_t$, or $i_t \prec j_t$
$\tau^{\text{A}}(\mathbf{P})$	sample complexity of the learning algorithm A, when started on a dueling bandits problem with indifferences specified by $\mathbf{P}$
$\hat{i}_{\text{A}}$	Copeland Winner candidate returned by A
δ	specified probability of error (failure probability)
(TRA-)POCOWISTA related	
$\widehat{CP}(i)$	current Copeland score of i
$\overline{CP}(i)$	potential Copeland score of i
$D(i)$	set of already compared arms to i
$W(i)$	defeated arms by i
$I(i)$	indifferent arms to i
$L(i)$	superior arms to i

B DERIVATION OF LOWER BOUNDS

Before giving the proof and discussion of Theorem 2.1, we need some additional notation and auxiliary results. The *Kullback-Leibler divergence* for two categorical distributions $\mathbf{p} = (p_1, \ldots, p_K) \in [0,1]^K$ and $\mathbf{q} = (q_1, \ldots, q_K) \in [0,1]^K$ such that $\sum_{i=1}^K p_i = \sum_{i=1}^K q_i = 1$ is given by

$$\mathrm{KL}(\mathbf{p}, \mathbf{q}) = \begin{cases} \sum_{i \in [K]: p_i > 0} p_i \ln\left(\frac{p_i}{q_i}\right), & \text{if } \forall j \in [K] : q_j = 0 \Rightarrow p_j = 0, \\ \infty, & \text{otherwise.} \end{cases}$$

If $K = 2$, we will simply write $\mathrm{kl}(p, q) \coloneqq \mathrm{KL}((p, 1-p), (q, 1-q))$ for any $p, q \in [0,1]$.

Lemma B.1. *(i) For any two categorical distributions* $\mathbf{p} = (p_1, \ldots, p_K) \in [0,1]^K$ *and* $\mathbf{q} = (q_1, \ldots, q_K) \in [0,1]^K$*, it holds that*

$$\mathrm{KL}(\mathbf{p}, \mathbf{q}) \leq \sum\nolimits_{i=1}^K \frac{(p_i - q_i)^2}{q_i}.$$

(ii) For any $\delta \in (0,1)$ *it holds that* $\mathrm{kl}(\delta, 1-\delta) \geq \ln((2.4\delta)^{-1})$.

(iii) For any $p, q \in [0,1]$ *it holds that* $2(p-q)^2 \leq \mathrm{kl}(p,q) \leq \frac{(p-q)^2}{q(1-q)}$.

For any learning algorithm A for the dueling bandits problem with indifferences let $\tau^{\mathrm{A}}(\mathbf{P})$ denote its number of samples, when started on a dueling bandits problem with indifferences specified by $\mathbf{P} = ((P_{i,j}^{\succ}, P_{i,j}^{\cong}, P_{i,j}^{\prec}))_{i<j}$ Further, let us write d_t^{A} for the duel (element of $\mathcal{A} \times \mathcal{A}$) made at time step t. Define $\tau_{i,j}^{\mathrm{A}}(\mathbf{P})$ to be the number of times A compares (i,j) or equivalently (j,i) before termination, i.e.,

$$\tau_{i,j}^{\mathrm{A}}(\mathbf{P}) = \sum_{t=1}^{\tau^{\mathrm{A}}(\mathbf{P})} 1_{[\![d_t^{\mathrm{A}} = \{i,j\}]\!]},$$

so that $\tau^{\mathrm{A}}(\mathbf{P}) = \sum_{i<j} \tau_{i,j}^{\mathrm{A}}(\mathbf{P})$. In the following, we will simply write τ^{A} for $\tau^{\mathrm{A}}(\mathbf{P})$ and $\tau_{i,j}^{\mathrm{A}}$ for $\tau_{i,j}^{\mathrm{A}}(\mathbf{P})$.

Let o_t^{A} be the feedback observed by A at time step t, after conducting the duel d_t^{A}. Write $\mathcal{F}_t^{\mathrm{A}} = \sigma(d_1^{\mathrm{A}}, o_1^{\mathrm{A}}, \ldots, d_t^{\mathrm{A}}, o_t^{\mathrm{A}})$ for the sigma algebra generated by the choices and the corresponding observed feedback of A until time t, and as usual $\mathcal{F}_{\tau^{\mathrm{A}}} = \{B \in \sigma(\bigcup_t \mathcal{F}_t^{\mathrm{A}}) : B \cap \{\tau^{\mathrm{A}} \leq t\} \in \mathcal{F}_t^{\mathrm{A}} \forall t \in \mathbb{N}\}$. Note that A can be interpreted as a learning algorithm for the classical multi-armed bandit with $\binom{n}{2}$ many arms (one for each possible pair) and "rewards" $r(o_t^{\mathrm{A}}) \in \{-1, 0, 1\}$, where for $o_t^{\mathrm{A}} \in \{i_t \succ j_t, i_t \cong j_t, i_t \prec j_t\}$ we set

$$r(o_t^{\mathrm{A}}) = \begin{cases} -1, & o_t^{\mathrm{A}} = i_t \prec j_t, \\ 0, & o_t^{\mathrm{A}} = i_t \cong j_t, \\ 1, & o_t^{\mathrm{A}} = i_t \succ j_t. \end{cases}$$

For sake of convenience, let us write $P_{i,j} = (P_{i,j}^{\succ}, P_{i,j}^{\cong}, P_{i,j}^{\prec})$, so that $\mathbf{P} = (P_{i,j})_{i<j}$. With this, we may transfer Lemma 1 from Kaufmann et al. (2016) to our setting as follows:

Lemma B.2. *Let* $\mathbf{P} = (P_{i,j}^{\succ}, P_{i,j}^{\cong}, P_{i,j}^{\prec})_{i<j}$, $\widetilde{\mathbf{P}} = (\widetilde{P}_{i,j}^{\succ}, \widetilde{P}_{i,j}^{\cong}, \widetilde{P}_{i,j}^{\prec})_{i<j}$ *be two problem instances of the dueling bandits problem with indifferences such that* $\min_{i<j} P_{i,j} > 0$ *and* $\min_{i<j} \widetilde{P}_{i,j} > 0$*. For any learning algorithm* A *for the dueling bandits problem with indifferences, which fulfills* $\mathbb{E}_{\mathbf{P}}[\tau^{\mathrm{A}}(\mathbf{P})], \mathbb{E}_{\widetilde{\mathbf{P}}}[\tau^{\mathrm{A}}(\widetilde{\mathbf{P}})] < \infty$*, it holds that*

$$\sum\nolimits_{i<j} \mathbb{E}_{\mathbf{P}}\left[\tau_{i,j}^{\mathrm{A}}(\mathbf{P})\right] \mathrm{KL}(P_{i,j}, \widetilde{P}_{i,j}) \geq \sup_{\mathcal{E} \in \mathcal{F}_{\tau^{\mathrm{A}}}} \mathrm{kl}\left(\mathbb{P}_{\mathbf{P}}(\mathcal{E}), \mathbb{P}_{\mathbf{P}'}(\mathcal{E})\right).$$

In case $\mathbf{P}$ *and* $\widetilde{\mathbf{P}}$ *do not have indifferences (i.e., if* $\max_{i,j} P_{i,j}^{\cong} = 0 = \max_{i,j} \widetilde{P}_{i,j}^{\cong}$*), the same inequality holds with* $\mathrm{KL}(P_{i,j}, \widetilde{P}_{i,j})$ *replaced by* $\mathrm{kl}(P_{i,j}^{\succ}, \widetilde{P}_{i,j}^{\succ})$*, respectively.*

We are now ready to prove Theorem 2.1.

Recall $L(j)$ and $I(j)$, let $\mathrm{CP}^* = \max_i \mathrm{CP}(i)$ and $d_j = \mathrm{CP}^* - \mathrm{CP}(j)$ be as above. In order to capture the dependence on the underlying instance $\mathbf{P}$ of interest, we may simply write $L(\mathbf{P}, j)$ and $I(\mathbf{P}, j)$ as well as $d_j(\mathbf{P})$,

$\mathrm{CP}^*(\mathbf{P})$ and $\mathrm{CP}(\mathbf{P}, j)$ for these corresponding terms. For any set $X \subseteq [n]$ and $k \in \mathbb{N}$, let us denote by $X^{[k]}$ the set of k-sized subsets of X. If $\mathbf{P}$ has no indifferences, let us simply write

$$\kappa_{x,y}(\mathbf{P}) \coloneqq \mathrm{kl}\left(P^{\succ}_{x,y}, 1 - P^{\succ}_{x,y}\right)$$

for any $x, y \in \mathcal{A}$.

Theorem B.3. *If* A *is an algorithm that correctly identifies the Copeland winner with confidence* $1-\delta$ *for any* $\mathbf{P}$ *without indifferences, then we have for all such* $\mathbf{P}$ *with* $\min_{i<j} \min\{P^{\succ}_{i,j}, 1 - P^{\succ}_{i,j}\} > 0$ *and* $\mathcal{C}(\mathbf{P}) = \{i^*\}$ *(for some* i^**, i.e., there is a unique Copeland winner), then*

$$\mathbb{E}_{\mathbf{P}}[\tau^{\mathrm{A}}] \geq \ln \frac{1}{2.4\delta} \sum_{j \in \mathcal{A} \setminus \{i^*\}} \max\left\{ \frac{|L(j)|}{d_j + 1} 1_{\llbracket |L(j)| \geq d_j + 1 \rrbracket}, \frac{|L(j)| - 1}{|L(j)| + d_j - 2} 1_{\llbracket i^* \in L(j) \rrbracket} \right\} \min_{z \in L(j)} \frac{1}{\kappa_{j,z}(\mathbf{P})}. \tag{6}$$

Note that the right-hand side of (6) depends not only via $\kappa_{j,z}(\mathbf{P})$ but also via $d_j = d_j(\mathbf{P})$, $L(j) = L(\mathbf{P}, j)$ on the underlying instance $\mathbf{P}$. Before stating its proof, the following lemma assures us that the lower bound from Theorem B.3 is in any case non-trivial.

Lemma B.4. *Suppose* $\mathbf{P}$ *has no indifferences and* $\mathcal{C}(\mathbf{P}) = \{i^*\}$ *holds and let* $j \in \mathcal{A} \setminus \{i^*\}$ *be arbitrary. If* $i^* \notin L(j)$*, then* $|L(j)| \geq d_j + 1$*.*

Proof of Lemma B.4. As $\mathbf{P}$ has no indifferences, the Copeland scores are given as

$$\mathrm{CP}(\mathbf{P}, l) = (n-1) - |L(l)|$$

for any $l \in \mathcal{A}$. If $i^* \notin L(j)$, then $j \in L(i^*)$, which implies $\mathrm{CP}(\mathbf{P}, i^*) \leq (n-1) - 1 = n - 2$ and thus

$$d_j = \mathrm{CP}(\mathbf{P}, i^*) - \mathrm{CP}(\mathbf{P}, j) \leq (n-2) - ((n-1) - |L(j)|) = |L(j)| - 1$$

follows directly. □

According to the previous lemma, for any instance $\mathbf{P}$, at least one of the indicators appearing in (6) is 1, whence the lower bound is larger than 0. In case $i^* \in L(j)$ and $L(j)$ is large, it is possible that both indicators are 1. We proceed with the proof of the theorem.

Proof of Theorem B.3. Suppose A and $\mathbf{P}$ with $\mathcal{C}(\mathbf{P}) = \{i^*\}$ without indifferences are fixed.

Claim 1: The following holds:

(i) If $\mathcal{L} \subseteq L(j)$ fulfills $|\mathcal{L}| \geq d_j + 1$, then

$$\ln \frac{1}{2.4\delta} \leq \sum\nolimits_{z \in \mathcal{L}} \mathbb{E}_{\mathbf{P}}[\tau^{\mathrm{A}}_{jz}] \kappa_{j,z}(\mathbf{P}). \tag{7}$$

(ii) If $i^* \in L(j)$ and $\mathcal{L} \subseteq L(j) \setminus \{i^*\}$ fulfills $|\mathcal{L}| \geq d_j - 1$, then (7) holds as well.

Proof of Claim 1: To prove (i), suppose $\mathcal{L} \subseteq L(j)$ with $|\mathcal{L}| \geq d_j + 1$ to be arbitrary but fixed for the moment. Define the instance $\widetilde{\mathbf{P}}$ via

$$\widetilde{P}^{\succ}_{x,y} \coloneqq \begin{cases} 1 - P^{\succ}_{x,y}, & \text{if } (x,y) \in \{(j,z),(z,j)\} \text{ for } z \in \mathcal{L}, \\ P^{\succ}_{x,y}, & \text{otherwise.} \end{cases}$$

By construction we have $\mathrm{CP}(\widetilde{\mathbf{P}}, i^*) \leq \mathrm{CP}(\mathbf{P}, i^*)$ and obtain

$$\begin{aligned} \mathrm{CP}(\widetilde{\mathbf{P}}, j) &= \mathrm{CP}(\mathbf{P}, j) + |\mathcal{L}| \geq \mathrm{CP}(\mathbf{P}, j) + d_j + 1 \\ &= \mathrm{CP}(\mathbf{P}, j) + (\mathrm{CP}(\mathbf{P}, i^*) - \mathrm{CP}(\mathbf{P}, j)) + 1 = \mathrm{CP}(\mathbf{P}, i^*) + 1 \geq \mathrm{CP}(\widetilde{\mathbf{P}}, i^*) + 1. \end{aligned}$$

This shows $i^* \notin \mathcal{C}(\widetilde{\mathbf{P}})$, and by assumption on A the event $\mathcal{E} \coloneqq \{i^* \in \mathcal{C}(\mathbf{P})\} \in \mathcal{F}_{\tau^{\mathrm{A}}}$ has the properties

$$\mathbb{P}_{\mathbf{P}}(\mathcal{E}) \geq 1 - \delta \quad \text{and} \quad \mathbb{P}_{\widetilde{\mathbf{P}}}(\mathcal{E}) \leq \delta.$$

Therefore, Lemma B.2 and part (ii) of Lemma B.1 assure

$$\sum\nolimits_{x<y} \mathbb{E}_{\mathbf{P}}\left[\tau^{\mathrm{A}}_{x,y}\right] \mathrm{KL}\left(P_{x,y}, \widetilde{P}_{x,y}\right) \geq \mathrm{kl}\left(\mathbb{P}_{\mathbf{P}}(\mathcal{E}), \mathbb{P}_{\widetilde{\mathbf{P}}}(\mathcal{E})\right) \geq \mathrm{kl}(1-\delta, \delta) \geq \ln \frac{1}{2.4\delta}. \tag{8}$$

In case $(x,y) \notin \{(j,z),(z,j)\}$ for any $z \in \mathcal{L}$, it holds that $\widetilde{P}_{x,y} = P_{x,y}$ so that $\mathrm{KL}\left(P_{x,y}, \widetilde{P}_{x,y}\right) = 0$. In case $z \in \mathcal{L} \subseteq L(j)$ we have

$$\mathrm{KL}\left(P_{j,z}, \widetilde{P}_{j,z}\right) = \mathrm{KL}\left((P^{\succ}_{j,z}, P^{\prec}_{j,z}), (P^{\prec}_{j,z}, P^{\succ}_{j,z})\right) = \mathrm{kl}\left(P^{\succ}_{j,z}, 1 - P^{\succ}_{j,z}\right) = \kappa_{j,z}(\mathbf{P}).$$

Combining these estimates with (8) proves (7).

To prove (ii) suppose $i^* \in L(j)$ holds and $\mathcal{L} \subseteq L(j) \setminus \{i^*\}$ fulfills $|\mathcal{L}| \geq d_j - 1$. Define the instance $\widetilde{\mathbf{P}}$ via

$$\widetilde{P}^{\succ}_{x,y} \coloneqq \begin{cases} 1 - P^{\succ}_{x,y}, & \text{if } (x,y) \in \{(j,z),(z,j)\} \text{ for } z \in \mathcal{L} \cup \{i^*\} \\ P^{\succ}_{x,y}, & \text{otherwise.} \end{cases}$$

Regarding that $\widetilde{P}^{\succ}_{j,i^*} = 1 - P^{\succ}_{j,i^*}$ and $\widetilde{P}^{\succ}_{z,i^*} = P^{\succ}_{z,i^*}$ for $z \neq j$, we have $\mathrm{CP}(\widetilde{\mathbf{P}}, i^*) = \mathrm{CP}(\mathbf{P}, i^*) - 1$, and similarly we see $\mathrm{CP}(\widetilde{\mathbf{P}}, j) = \mathrm{CP}(\mathbf{P}, j) + |\mathcal{L}| + 1$. Together with $|\mathcal{L}| \geq d_j - 1$, we obtain with the same argumentation as before that $\mathrm{CP}(\widetilde{\mathbf{P}}, j) \geq \mathrm{CP}(\widetilde{\mathbf{P}}, i^*) + 1$ and thus $i^* \notin \mathcal{C}(\widetilde{\mathbf{P}})$. Therefore, following the lines from above, we conclude that (7) also holds in this case. ■

To prove the theorem, abbreviate for convenience $\kappa_{x,y} = \kappa_{x,y}(\mathbf{P})$ in the following, and let us at first suppose that $|L(j)| \geq d_j + 1$ holds. When summing the inequality (7) over all $\binom{|L(j)|}{d_j+1}$ many $\mathcal{L} \subseteq L(j)$ of size $|\mathcal{L}| = d_j + 1$, any of the summands $\mathbb{E}_{\mathbf{P}}[\tau^{\mathrm{A}}_{jz}]\kappa_{j,z}$, with $z \in L(j)$, appears exactly $\binom{|L(j)|}{d_j+1}$ times, i.e., we have

$$\begin{aligned} \binom{|L(j)|}{d_j+1} \ln \frac{1}{2.4\delta} &\leq \sum\nolimits_{z \in \mathcal{L}} \binom{|L(j)|-1}{d_j} \mathbb{E}_{\mathbf{P}}[\tau^{\mathrm{A}}_{jz}]\kappa_{j,z} \\ &\leq \binom{|L(j)|-1}{d_j} \left(\max\nolimits_{z \in L(j)} \kappa_{j,z}\right) \sum\nolimits_{z \in L(j)} \mathbb{E}_{\mathbf{P}}[\tau^{\mathrm{A}}_{jz}]. \end{aligned}$$

Using that $\binom{a}{b}/\binom{a-1}{b-1} = \frac{a}{b}$ holds for any $a, b \in \mathbb{N}$ with $a \leq b$, we infer

$$\mathbb{E}_{\mathbf{P}}\left[\tau^{\mathrm{A}}_j\right] \geq \sum\nolimits_{z \in L(j)} \mathbb{E}_{\mathbf{P}}\left[\tau^{\mathrm{A}}_{jz}\right] \geq \left(\ln \frac{1}{2.4\delta}\right) \frac{|L(j)|}{d_j + 1} \min\nolimits_{z \in L(j)} \frac{1}{\kappa_{j,z}}.$$

Now, suppose $i^* \in L(j)$. When summing (7) over all $\binom{|L(j)|-1}{d_j-1}$ many $\mathcal{L} \subseteq L(j) \setminus \{i^*\}$ with $|\mathcal{L}| = d_j - 1$, we observe $\mathbb{E}_{\mathbf{P}}[\tau^{\mathrm{A}}_{ji^*}]\kappa_{j,i^*}$ exactly $\binom{|L(j)|-1}{d_j-1}$ times as a summand and each of the terms $\mathbb{E}_{\mathbf{P}}[\tau^{\mathrm{A}}_{jz}]\kappa_{j,z}$, with $z \in L(j)$, exactly[2] $\binom{|L(j)|-2}{d_j-2} 1_{[\![d_j \geq 2]\!]}$ times as a summand. Therefore, we get

$$\binom{|L(j)|-1}{d_j-1} \ln \frac{1}{2.4\delta} \leq \binom{|L(j)|-1}{d_j-1} \mathbb{E}_{\mathbf{P}}[\tau^{\mathrm{A}}_{ji^*}]\kappa_{j,i^*} + \sum\nolimits_{z \in \mathcal{L}} \binom{|L(j)|-2}{d_j-2} \mathbb{E}_{\mathbf{P}}[\tau^{\mathrm{A}}_{jz}]\kappa_{j,z} 1_{[\![d_j \geq 2]\!]}. \tag{9}$$

If $d_j \geq 2$, again using that $\binom{a}{b}/\binom{a-1}{b-1} = \frac{a}{b}$ holds for any $a, b \in \mathbb{N}$ with $a \leq b$, we obtain from

$$\binom{|L(j)|-1}{d_j-1} + \binom{|L(j)|-2}{d_j-2} = \frac{|L(j)| + d_j - 2}{|L(j)| - 1} \binom{|L(j)|-2}{d_j-2}$$

that

$$\mathbb{E}_{\mathbf{P}}[\tau^{\mathrm{A}}_j] \geq \mathbb{E}_{\mathbf{P}}[\tau^{\mathrm{A}}_{jz}] \geq \ln \frac{1}{2.4\delta} \frac{|L(j)| - 1}{|L(j)| + d_j - 2} \left(\min\nolimits_{z \in L(j)} \frac{1}{\kappa_{j,z}}\right)$$

and (6) follows. In the other case $d_j = 1$, we have $\frac{|L(j)|+d_j-2}{|L(j)|-1} = 1$ and thus (6) can be inferred from (9). This completes the proof. □

[2]Note here that $\mathcal{L} = \emptyset$ if $d_j = 1$.

Next, we want to prove an analogon of the above lower bound for the more sophisticated scenario of dueling bandits with indifferences. To prepare this, define for any instance $\mathbf{P}$ with indifferences and $x, y \in \mathcal{A}$ the terms

$$D_{x,y}(\mathbf{P}) \coloneqq \max\Big\{\mathrm{KL}\left(\left(P^{\succ}_{x,y}, P^{\cong}_{x,y}, P^{\prec}_{x,y}\right), \left(P^{\cong}_{x,y}, P^{\succ}_{x,y}, P^{\prec}_{x,y}\right)\right), \\ \mathrm{KL}\left(\left(P^{\succ}_{x,y}, P^{\cong}_{x,y}, P^{\prec}_{x,y}\right), \left(P^{\prec}_{x,y}, P^{\cong}_{x,y}, P^{\succ}_{x,y}\right)\right)\Big\}.$$

If $\mathbf{P}$ is fixed or clear by the context, we may simply write $D_{x,y}$ instead of $D_{x,y}(\mathbf{P})$.

Theorem B.5. *If* A *is an algorithm that correctly identifies the Copeland winner with confidence* $1-\delta$ *for any* $\mathbf{P}$*, then we have for all* $\mathbf{P}$ *with* $\min_{i<j} \min\{P^{\succ}_{i,j}, P^{\cong}_{i,j}, P^{\prec}_{i,j}\} > 0$ *and* $\mathcal{C}(\mathbf{P}) = \{i^*\}$ *(for some* i^**, i.e., there is a unique Copeland winner) the bound*

$$\mathbb{E}_{\mathbf{P}}[\tau^{\mathrm{A}}] \geq \ln \frac{1}{2.4\delta} \sum\nolimits_{j \in \mathcal{A}\setminus\{i^*\}} \max\left\{C_j, C'_j 1_{[\![i^* \in I(j)]\!]}, C''_j 1_{[\![i^* \in L(j)]\!]}\right\} \min_{z \in L(j)\cup I(j)} \frac{1}{D_{j,z}(\mathbf{P})}$$

where

$$C_j \coloneqq \max_{(i,l)\in\Psi(j)} \frac{\binom{|I(j)|}{i}\binom{|L(j)|}{l}}{\binom{|I(j)|-1}{i-1}\binom{|L(j)|}{l}1_{[\![i\geq 1]\!]} + \binom{|I(j)|}{i}\binom{|L(j)|-1}{l-1}1_{[\![l\geq 1]\!]}},$$

$$C'_j \coloneqq \max_{(i,l)\in\Psi'(j)} \frac{\binom{|I(j)|-1}{i}\binom{|L(j)|}{l}}{\binom{|I(j)|-1}{i}\binom{|L(j)|}{l} + \binom{|I(j)|-2}{i-1}\binom{|L(j)|}{l}1_{[\![i\geq 1]\!]} + \binom{|I(j)|-1}{i}\binom{|L(j)|-1}{l-1}1_{[\![l\geq 1]\!]}},$$

$$C''_j \coloneqq \max_{(i,l)\in\Psi''(j)} \frac{\binom{|I(j)|}{i}\binom{|L(j)|-1}{l}}{\binom{|I(j)|}{i}\binom{|L(j)|-1}{l} + \binom{|I(j)|-1}{i-1}\binom{|L(j)|-1}{l}1_{[\![i\geq 1]\!]} + \binom{|I(j)|}{i}\binom{|L(j)|-2}{l-1}1_{[\![l\geq 1]\!]}}$$

with

$$\Psi(j) \coloneqq \{(i,l) \,:\, i \in \{0,\dots,|I(j)|\}, l \in \{0,\dots,|L(j)|\} \text{ and } i+2l \geq 2d_j+1\},$$
$$\Psi'(j) \coloneqq \{(i,l) \,:\, i \in \{0,\dots,|I(j)|-1\}, l \in \{0,\dots,|L(j)|\} \text{ and } i+2l \geq 2d_j-1\},$$
$$\Psi''(j) \coloneqq \{(i,l) \,:\, i \in \{0,\dots,|I(j)|\}, l \in \{0,\dots,|L(j)|-1\} \text{ and } i+2l \geq 2d_j-3\}.$$

Before providing its proof, let us briefly discuss this lower bound. At first, note that in fact all of the binomial coefficients appearing in the definitions of C_j, C'_j and C''_j are well-defined and of the form $\binom{n}{k}$ for $0 \leq k \leq n$. In the definition of C_j this is assured by means of the indicator functions $1_{[\![i\geq 1]\!]}$ and $1_{[\![l\geq 1]\!]}$, and in the definition of C'_j resp. C''_j this follows from the definitions of $\Psi'(j)$ resp. $\Psi''(j)$. For example, if $(i,l) \in \Psi'(j)$, then $i \geq 0$ assures $|I(j)| \geq i+1 \geq 1$, and in case $i \geq 1$ we have $|I(j)| \geq 2$, whence $\binom{|I(j)|-2}{i-1}\binom{|L(j)|}{l}$ is well-defined.

Note that Thm. 2.1 is a direct consequence of the stated lower bounds. For the sake of completeness, we formulate it as follows.

Proof of Theorem 2.1. Part (i) is exactly Theorem B.3, and part (ii) follows due to $\max\{C_j, C'_j 1_{[\![i^* \in I(j)]\!]}, C''_j 1_{[\![i^* \in L(j)]\!]}\} \geq C_j$ directly from Theorem B.5. □

The maximum in the proof of Thm. B.5 actually assures that the lower bound is non-trivial on any instance $\mathbf{P}$ considered in the theorem. This is made formal in the upcoming lemma.

Lemma B.6. *Let* $\mathbf{P}$ *be an instance with indifferences such that* $\mathcal{C}(\mathbf{P}) = \{i^*\}$*. Then, for any* $j \neq i^*$ *exactly one of the following holds:*

(i) $i^* \notin L(j) \cup I(j)$ *and* $\Psi(j) \neq \emptyset$*,*

(ii) $i^* \in I(j)$ *and* $\Psi'(j) \neq \emptyset$*,*

(iii) $i^* \in L(j)$ *and* $\Psi''(j) \neq \emptyset$*.*

Proof of Lemma B.6. Suppose $\mathbf{P}$ with $\mathcal{C}(\mathbf{P}) = \{i^*\}$ and $j \neq i^*$ to be fixed. Regarding the definition of the Copeland score, we have

$$\mathrm{CP}(\mathbf{P}, l) = (n-1) - |L(l)| - \frac{1}{2}|I(l)|$$

for any $l \in \mathcal{A}$. For fixed $j \neq i^*$ we have in particular

$$\mathrm{CP}(\mathbf{P}, i^*) \leq \begin{cases} n-2, & \text{if } i^* \notin I(j) \cup L(j), \\ n-3/2, & \text{if } i^* \in I(j), \\ n-1, & \text{if } i^* \in L(j) \end{cases}$$

and thus

$$d_j = d_j(\mathbf{P}) \leq \begin{cases} |L(j)| + |I(j)|/2 - 1, & \text{if } i^* \notin L(j) \cup I(j), \\ |L(j)| + |I(j)|/2 - 1/2, & \text{if } i^* \in I(j), \\ |L(j)| + |I(j)|/2, & \text{if } i^* \in L(j). \end{cases}$$

If $i^* \notin L(j) \cup I(j)$, then $2d_j \leq 2|L(j)| + |I(j)| - 2$ and $(|I(j)|, |L(j)|) \in \Psi(j)$ follows. In case $i^* \in I(j)$, $2d_j \leq 2|L(j)| + |I(j)| - 1$ and thus $(|I(j)| - 1, |L(j)|) \in \Psi'(j)$, and similarly we see in case $i^* \in L(j)$ that $2d_j \leq 2|L(j)| + |I(j)|$ implies $(|I(j)|, |L(j)| - 1) \in \Psi''(j)$. □

In contrast to Thm. B.5, the corresponding simplified version stated in Thm. 2.1 is e.g. trivial on the instance $\mathbf{P}$ defined via

$$(P^{\succ}_{i,j})_{i,j} = \begin{pmatrix} - & 1/2 & 1/2 \\ 1/4 & - & 1/2 \\ 1/4 & 1/4 & - \end{pmatrix}.$$

To see this, note that $P^{\cong}_{x,y} = 1/4$ holds for all $x, y \in \mathcal{A}$, $\mathcal{C}(\mathbf{P}) = \{1\}$ and observe that $d_2 = 1 = |L(2)|$ resp. $d_3 = 2 = |L(3)|$ and $|I(2)| = |I(3)| = 0$ imply $\Psi(2) = \emptyset$ resp. $\Psi(3) = \emptyset$.

Now, let us proceed with the proof of Thm. B.5. The proof idea is similar to that of Thm. B.3, but as it is more sophisticated and technical, we prove it for the sake of completeness in detail.

Proof of Theorem B.5. Suppose A and $\mathbf{P}$ with $\mathcal{C}(\mathbf{P}) = \{i^*\}$ are fixed. Assume w.l.o.g. $i^* = 1$ and let $j \in [n] \setminus \{i^*\}$ be arbitrary but fixed for the moment.

Claim 1: The following holds:

(i) If $(i, l) \in \Psi(j)$, we have for each $\mathcal{I} \in I(j)^{[i]}, \mathcal{L} \in L(j)^{[l]}$ that

$$\ln \frac{1}{2.4\delta} \leq \sum\nolimits_{z \in \mathcal{I}} \mathbb{E}_{\mathbf{P}}\left[\tau^{\mathrm{A}}_{jz}\right] D_{j,z}(\mathbf{P}) + \sum\nolimits_{z \in \mathcal{L}} \mathbb{E}_{\mathbf{P}}\left[\tau^{\mathrm{A}}_{jz}\right] D_{j,z}(\mathbf{P}). \tag{10}$$

(ii) If $(i, l) \in \Psi'(j)$ and $i^* \in I(j)$, (10) holds for all $\mathcal{I} \in I(j)^{[i]}, \mathcal{L} \in L(j)^{[l]}$ with $i^* \in \mathcal{I}$.

(iii) If $(i, l) \in \Psi''(j)$ and $i^* \in L(j)$, (10) holds for all $\mathcal{I} \in I(j)^{[i]}, \mathcal{L} \in L(j)^{[l]}$ with $i^* \in \mathcal{L}$.

Proof of Claim 1: To prove (i), suppose $(i, l) \in \Psi(j)$, that is, $i \in \{0, \ldots, |I(j)|\}$ and $l \in \{0, \ldots, |L(j)|\}$ are such that $i + 2l \geq 2d_j + 1$. Let $\mathcal{I} \in I(j)^{[i]}$ and $\mathcal{L} \in L(j)^{[l]}$ be arbitrary but fixed. Define $\widetilde{P}^{\succ}$ as a modification of $P^{\succ}$ via $\widetilde{P}^{\succ}_{x,y} := P^{\succ}_{x,y}$ for any $(x, y) \in \mathcal{A} \times \mathcal{A}$ with $x \neq j \neq y$,

$$\widetilde{P}^{\succ}_{j,z} := P^{\cong}_{j,z}, \quad \widetilde{P}^{\cong}_{j,z} := P^{\succ}_{j,z} \quad \text{and} \quad \widetilde{P}^{\prec}_{j,z} := P^{\prec}_{j,z} \tag{11}$$

for all $z \in \mathcal{I}$, and

$$\widetilde{P}^{\succ}_{j,z} := P^{\prec}_{j,z}, \quad \widetilde{P}^{\cong}_{j,z} := P^{\cong}_{j,z} \quad \text{and} \quad \widetilde{P}^{\prec}_{j,z} := P^{\succ}_{j,z} \tag{12}$$

for all $z \in \mathcal{L}$. By construction of $\widetilde{P}^{\succ}$, we have[3] $\mathrm{CP}(\widetilde{P}^{\succ}, i^*) \leq \mathrm{CP}(P^{\succ}, i^*)$. Moreover, as the modes of the (j,z)-components of $P^{\succ}$ have been *flipped to the "category"* $j \succ z$ for $z \in \mathcal{I} \cup \mathcal{L}$ and remained unchanged otherwise, we have

$$\mathrm{CP}(\widetilde{P}^{\succ}, j) = \mathrm{CP}(P^{\succ}, j) + \frac{1}{2}|\mathcal{I}| + |\mathcal{L}| = \mathrm{CP}(P^{\succ}, j) + \frac{i}{2} + l.$$

As $i + 2l \geq 2d_j + 1$ holds by assumption, we thus get

$$\begin{aligned}\mathrm{CP}(\widetilde{P}^{\succ}, j) &\geq \mathrm{CP}(P^{\succ}, j) + d_j + 1/2 \\ &= \mathrm{CP}(P^{\succ}, j) + (\mathrm{CP}(P^{\succ}, i^*) - \mathrm{CP}(P^{\succ}, j)) + 1/2 \\ &= \mathrm{CP}(P^{\succ}, i^*) + 1/2 \geq \mathrm{CP}(\widetilde{P}^{\succ}, i^*) + 1/2.\end{aligned}$$

This shows[4] $i^* \notin \mathcal{C}(P^{\succ})$, and by assumption on A, the event $\mathcal{E} \coloneqq \{i^* \in \mathcal{C}(\mathbf{P})\} \in \mathcal{F}_{\tau^{\mathrm{A}}}$ has the properties

$$\mathbb{P}_{P^{\succ}}(\mathcal{E}) \geq 1 - \delta, \quad \mathbb{P}_{\widetilde{P}^{\succ}}(\mathcal{E}) \leq \delta.$$

Consequently, Lemma B.2 and part (ii) of Lemma B.1 assure

$$\sum\nolimits_{x<y} \mathbb{E}_{\mathbf{P}}\left[\tau^{\mathrm{A}}_{x,y}\right] \mathrm{KL}\left(P_{x,y}, \widetilde{P}_{x,y}\right) \geq \mathrm{kl}\left(\mathbb{P}_{\mathbf{P}}(\mathcal{E}), \mathbb{P}_{\widetilde{\mathbf{P}}}(\mathcal{E})\right) \geq \mathrm{kl}(1-\delta, \delta) \geq \ln \frac{1}{2.4\delta}. \tag{13}$$

In case $x \neq j \neq y$, $\widetilde{P}_{x,y} = P_{x,y}$ assures $\mathrm{KL}\left(P_{x,y}, \widetilde{P}_{x,y}\right) = 0$. In case $z \in \mathcal{I} \subseteq I(j)$, we have

$$\mathrm{KL}\left(P_{j,z}, \widetilde{P}_{j,z}\right) = \mathrm{KL}\left(\left(P^{\succ}_{j,z}, P^{\cong}_{j,z}, P^{\prec}_{j,z}\right), \left(P^{\cong}_{j,z}, P^{\succ}_{j,z}, P^{\prec}_{j,z}\right)\right) \leq D_{j,z}(\mathbf{P}),$$

and in case $z \in \mathcal{L} \subseteq L(j)$ we similarly see

$$\mathrm{KL}\left(P_{j,z}, \widetilde{P}_{j,z}\right) = \mathrm{KL}\left((P^{\succ}_{j,z}, P^{\cong}_{j,z}, P^{\prec}_{j,z}), (P^{\prec}_{j,z}, P^{\cong}_{j,z}, P^{\succ}_{j,z})\right) \leq D_{j,z}(\mathbf{P}).$$

Combining these estimates with (13) proves (10).

To prove (ii), suppose now $(i,l) \in \Psi'(j)$ and $\mathcal{I} \in I(j)^{[i]}$, $\mathcal{L} \in L(j)^{[l]}$ with $i^* \in \mathcal{I}$ are given. Similarly as above, one may construct an instance $\widetilde{\mathbf{P}}$, which differs from $\mathbf{P}$ only on positions (j,z), $z \in \mathcal{I} \cup \mathcal{L} \cup \{i^*\}$, such that (11) for all $z \in \mathcal{I}$ and (11) for all $z \in \mathcal{L}$ and $\widetilde{P}^{\succ}_{j,i^*} = P^{\cong}_{j,i^*}$, $\widetilde{P}^{\cong}_{j,i^*} = P^{\succ}_{j,i^*}$ and $\widetilde{P}^{\prec}_{j,i^*} = P^{\prec}_{j,i^*}$. Then, $\mathrm{CP}(\widetilde{P}^{\succ}, i^*) = \mathrm{CP}(P^{\succ}, i^*)$ holds and again $\mathrm{CP}(\widetilde{P}^{\succ}, j) = \mathrm{CP}(P^{\succ}, j) + \frac{i}{2} + l$. Due to $i + 2l \geq 2d_j - 1$ we obtain $\mathrm{CP}(\widetilde{P}^{\succ}, j) \geq \mathrm{CP}(\widetilde{P}^{\succ}, j) + 1/2$ and thus $i^* \notin \mathcal{C}(\widetilde{P}^{\succ})$. Thus, the same argumentation as above shows that (10) also holds in this case.

For proving (iii), construct $\widetilde{P}^{\succ}$ such that it differs from P only on positions (j,z), $z \in \mathcal{I} \cup \mathcal{L} \cup \{i^*\}$, fulfills (11) for all $z \in \mathcal{I}$ and (11) for all $z \in \mathcal{L}$ and further $\widetilde{P}^{\succ}_{j,i^*} = P^{\prec}_{j,i^*}$, $\widetilde{P}^{\cong}_{j,i^*} = P^{\cong}_{j,i^*}$ and $\widetilde{P}^{\prec}_{j,i^*} = P^{\succ}_{j,i^*}$. Then, $\mathrm{CP}(\widetilde{P}^{\succ}, i^*) = \mathrm{CP}(P^{\succ}, i^*) - 1$ holds, and the assumptions stated in (iii) suffice to show $i^* \notin \mathcal{C}(\widetilde{P}^{\succ})$. Therefore, repeating the arguments from above shows (10).■

As $\mathbf{P}$ is fixed, we may simply write $D_{x,y}$ for $D_{x,y}(\mathbf{P})$ throughout the rest of the proof. First, let $(i,l) \in \Psi(j)$ be arbitrary but fixed, i.e., $i \in \{0, \ldots, |I(j)|\}$ and $l \in \{0, \ldots, |L(j)|\}$ and $i + 2l \geq 2d_j + 1$ hold. According to Part (i) of Claim 1, (10) holds for any of the $\binom{|I(j)|}{i}\binom{|L(j)|}{l}$ many $(\mathcal{I}, \mathcal{L}) \in I(j)^{[i]} \times L(j)^{[l]}$. When summing (10) over all such $(\mathcal{I}, \mathcal{L})$, the summand $\mathbb{E}_{\mathbf{P}}\left[\tau^{\mathrm{A}}_{jz}\right] D_{j,z}$ appears exactly $\binom{|I(j)|-1}{i-1}\binom{|L(j)|}{l} 1_{[\![i \geq 1]\!]}$ many times if $z \in I(j)$,

[3] In fact, the difference $\mathrm{CP}(P^{\succ}, i^*) - \mathrm{CP}(\widetilde{P}^{\succ}, i^*)$ is 1 resp. 1/2 resp. 0 if $i^* \in \mathcal{L}$ resp. $i^* \in \mathcal{I}$ resp. $i^* \notin \mathcal{I} \cup \mathcal{L}$.

[4] In fact, by construction we even have $\mathcal{C}(\widetilde{P}^{\succ}) = \{j\}$.

and it appears $\binom{|I(j)|}{i}\binom{|L(j)|-1}{l-1}1_{[\![l\geq 1]\!]}$ many times if $z \in L(j)$. Consequently, we have

$$\begin{aligned}
&\binom{|I(j)|}{i}\binom{|L(j)|}{l}\ln\frac{1}{2.4\delta} \\
&\quad\leq \binom{|I(j)|-1}{i-1}\binom{|L(j)|}{l}1_{[\![i\geq 1]\!]}\sum\nolimits_{z\in\mathcal{I}}\mathbb{E}_{\mathbf{P}}\left[\tau_{jz}^{\mathrm{A}}\right]D_{j,z} \\
&\qquad+\binom{|I(j)|}{i}\binom{|L(j)|-1}{l-1}1_{[\![l\geq 1]\!]}\sum\nolimits_{z\in\mathcal{L}}\mathbb{E}_{\mathbf{P}}\left[\tau_{jz}^{\mathrm{A}}\right]D_{j,z} \\
&\quad\leq \max\nolimits_{z\in I(j)\cup L(j)}D_{j,z}\cdot\sum\nolimits_{z\in I(j)\cup L(j)}\mathbb{E}_{\mathbf{P}}\left[\tau_{jz}^{\mathrm{A}}\right] \\
&\qquad\cdot\left[\binom{|I(j)|-1}{i-1}\binom{|L(j)|}{l}1_{[\![i\geq 1]\!]}+\binom{|I(j)|}{i}\binom{|L(j)|-1}{l-1}1_{[\![l\geq 1]\!]}\right].
\end{aligned}$$

Thus, we obtain for $\tau_j^{\mathrm{A}}=\sum_{z\neq j}\tau_{jz}^{\mathrm{A}}$ the estimate

$$\begin{aligned}
\mathbb{E}_{\mathbf{P}}\left[\tau_j^{\mathrm{A}}\right] &\geq \sum\nolimits_{z\in I(j)\cup L(j)}\mathbb{E}_{\mathbf{P}}\left[\tau_{jz}^{\mathrm{A}}\right] \\
&\geq \frac{\binom{|I(j)|}{i}\binom{|L(j)|}{l}}{\binom{|I(j)|-1}{i-1}\binom{|L(j)|}{l}1_{[\![i\geq 1]\!]}+\binom{|I(j)|}{i}\binom{|L(j)|-1}{l-1}1_{[\![l\geq 1]\!]}}\left(\ln\frac{1}{2.4\delta}\right)\min\nolimits_{z\in I(j)\cup L(j)}\frac{1}{D_{j,z}}. \qquad (14)
\end{aligned}$$

Next, suppose $i^*\in I(j)$ and $(i,l)\in\Psi'(j)$ be fixed for the moment. For any of the $\binom{|I(j)|-1}{i}\binom{|L(j)|}{l}$ many $(\mathcal{I},\mathcal{L})\in I(j)^{[i]}\times\mathcal{L}(j)^{[l]}$ with $i^*\in\mathcal{I}$, Part (ii) of Claim 1 yields that (10) holds. Summing this over all such $(\mathcal{I},\mathcal{L})$, we observe:

- The summand $\mathbb{E}_{\mathbf{P}}\left[\tau_{ji^*}^{\mathrm{A}}\right]D_{j,i^*}$ appears $\binom{|I(j)|-1}{i}\binom{|L(j)|}{l}$ many times.
- For $z\in I(j)\setminus\{i^*\}$, the summand $\mathbb{E}_{\mathbf{P}}\left[\tau_{jz}^{\mathrm{A}}\right]D_{j,z}$ appears $\binom{|I(j)|-2}{i-1}\binom{|L(j)|}{l}$ many times if $|I(j)|>i\geq 1$, and it does not appear at all if $i=0$. Thus, this summand appears $\binom{|I(j)|-2}{i-1}\binom{|L(j)|}{l}1_{[\![|I(j)|>i\geq 1]\!]}$ many times.
- For $z\in L(j)$, the summand $\mathbb{E}_{\mathbf{P}}\left[\tau_{jz}^{\mathrm{A}}\right]D_{j,z}$ appears $\binom{|I(j)|-1}{i}\binom{|L(j)|-1}{l-1}1_{[\![l\geq 1]\!]}$ many times.

Thus, we obtain

$$\begin{aligned}
&\binom{|I(j)|-1}{i}\binom{|L(j)|}{l}\ln\frac{1}{2.4\delta} \\
&\leq \binom{|I(j)|-1}{i}\binom{|L(j)|}{l}\mathbb{E}_{\mathbf{P}}\left[\tau_{ji^*}^{\mathrm{A}}\right]D_{j,i^*} \\
&\quad+\binom{|I(j)|-2}{i-1}\binom{|L(j)|}{l}1_{[\![i\geq 1]\!]}\sum\nolimits_{z\in\mathcal{I}}\mathbb{E}_{\mathbf{P}}\left[\tau_{jz}^{\mathrm{A}}\right]D_{j,z} \\
&\quad+\binom{|I(j)|-1}{i}\binom{|L(j)|-1}{l-1}1_{[\![l\geq 1]\!]}\sum\nolimits_{z\in\mathcal{L}}\mathbb{E}_{\mathbf{P}}\left[\tau_{jz}^{\mathrm{A}}\right]D_{j,z} \\
&\leq \max\nolimits_{z\in I(j)\cup L(j)}D_{j,z}\cdot\sum\nolimits_{z\in I(j)\cup L(j)}\mathbb{E}_{\mathbf{P}}\left[\tau_{jz}^{\mathrm{A}}\right] \\
&\quad\cdot\left[\binom{|I(j)|-1}{i}\binom{|L(j)|}{l}+\binom{|I(j)|-2}{i-1}\binom{|L(j)|}{l}1_{[\![i\geq 1]\!]}+\binom{|I(j)|-1}{i}\binom{|L(j)|-1}{l-1}1_{[\![l\geq 1]\!]}\right],
\end{aligned}$$

which shows similarly as above that

$$\begin{aligned}
\mathbb{E}_{\mathbf{P}}\left[\tau_j^{\mathrm{A}}\right] \geq &\frac{\binom{|I(j)|-1}{i}\binom{|L(j)|}{l}}{\binom{|I(j)|-1}{i}\binom{|L(j)|}{l}+\binom{|I(j)|-2}{i-1}\binom{|L(j)|}{l}1_{[\![i\geq 1]\!]}+\binom{|I(j)|-1}{i}\binom{|L(j)|-1}{l-1}1_{[\![l\geq 1]\!]}}\left(\ln\frac{1}{2.4\delta}\right) \\
&\cdot\min\nolimits_{z\in I(j)\cup L(j)}\frac{1}{D_{j,z}}. \qquad (15)
\end{aligned}$$

Finally, suppose $i^*\in L(j)$ and let $(i,l)\in\Psi''(j)$ be fixed for the moment. For any of the $\binom{|I(j)|}{i}\binom{|L(j)|-1}{l}$ many $(\mathcal{I},\mathcal{L})\in I(j)^{[i]}\times L(j)^{[l]}$ with $i^*\in L(j)$, Part (iii) of Claim 1 yields that (10) holds. When summing over all these $(\mathcal{I},\mathcal{L})$, we see:

- The summand $\mathbb{E}_{\mathbf{P}}\left[\tau^{\mathrm{A}}_{ji^*}\right] D_{j,i^*}$ appears $\binom{|I(j)|}{i}\binom{|L(j)|-1}{l}$ many times.
- For $z \in I(j)$, the summand $\mathbb{E}_{\mathbf{P}}\left[\tau^{\mathrm{A}}_{jz}\right] D_{j,z}$ appears $\binom{|I(j)|-1}{i-1}\binom{|L(j)|-1}{l}1_{[\![i\geq 1]\!]}$ many times.
- For $z \in L(j)\setminus\{i^*\}$, the summand $\mathbb{E}_{\mathbf{P}}\left[\tau^{\mathrm{A}}_{jz}\right] D_{j,z}$ appears $\binom{|I(j)|}{i}\binom{|L(j)|-2}{l-1}1_{[\![l\geq 1]\!]}$ many times.

Analogously as above, we obtain

$$\mathbb{E}_{\mathbf{P}}\left[\tau^{\mathrm{A}}_j\right] \geq \frac{\binom{|I(j)|}{i}\binom{|L(j)|-1}{l}}{\binom{|I(j)|}{i}\binom{|L(j)|-1}{l} + \binom{|I(j)|-1}{i-1}\binom{|L(j)|-1}{l}1_{[\![i\geq 1]\!]} + \binom{|I(j)|}{i}\binom{|L(j)|-2}{l-1}1_{[\![l\geq 1]\!]}} \left(\ln\frac{1}{2.4\delta}\right) \cdot \min_{z\in I(j)\cup L(j)} \frac{1}{D_{j,z}}. \tag{16}$$

As (14) holds for all $(i,l)\in\Psi(j)$, (15) for all $(i,l)\in\Psi'(j)$, if $i^*\in I(j)$, and (16) holds for all $(i,l)\in\Psi''(j)$ if $i^*\in L(j)$, combining these estimates concludes the proof. □

For the sake of comparison, let us state the consequences from Theorem B.5 for the particular case when the indifferences are non-dominant in the sense that $I(z)=\emptyset$ for all z. Note that the maximum appearing therein is exactly the same term as that in the lower bound for Copeland winner identification without indifferences (Thm. B.3).

Corollary B.7. *If* A *is an algorithm that correctly identifies the Copeland winner with confidence* $1-\delta$ *for any* $\mathbf{P}$ *(possibly with indifferences), then we have for all such* $\mathbf{P}$ *with* $\min_{i<j}\min\{P^{\succ}_{i,j}, P^{\cong}_{i,j}, P^{\prec}_{i,j}\} > 0$, $\max_{z\in\mathcal{A}}|I(z)| = 0$ *and* $\mathcal{C}(\mathbf{P}) = \{i^*\}$ *that*

$$\mathbb{E}_{\mathbf{P}}\left[\tau^{\mathrm{A}}\right] \geq \ln\frac{1}{2.4\delta}\sum_{j\neq i^*}\max\left\{\frac{|L(j)|}{d_j+1}1_{[\![|L(j)|\geq d_j+1]\!]}, \frac{|L(j)|-1}{|L(j)|+d_j-2}1_{[\![i^*\in L(j)]\!]}\right\} \min_{z\in L(j)\cup I(j)}\frac{1}{D_{j,z}(\mathbf{P})}.$$

Proof of Corollary B.7. Suppose $\mathbf{P}$ is such that $\max_{z\in\mathcal{A}}|I(z)|=0$ and recall the definitions of C_j, C''_j, $\Psi(j)$ and $\Psi''(j)$ from Thm. B.5. Then, $\mathrm{CP}(j)$ and d_j are integers for any $j\in\mathcal{A}$. Moreover, for any $j\in\mathcal{A}\setminus\{i^*\}$, $|I(j)|=0$ directly implies $\Psi(j)=\{(0,d_j+1),\dots,(0,|L(j)|)\}$ and $\Psi''(j)=\{(0,d_j-1),\dots,(0,|L(j)|-1)\}$. Note that $\Psi(j)\neq\emptyset$ iff $|L(j)|\geq d_j+1$, whereas $|L(j)|\geq d_j$ shows that $\Psi''(j)\neq\emptyset$ in any case. Using that $\binom{a}{b}/\binom{a-1}{b-1}=\frac{a}{b}$ holds for any $a,b\in\mathbb{N}$ with $a\leq b$, we thus obtain

$$\begin{aligned} C_j &= \max_{(i,l)\in\Psi(j)}\frac{\binom{|I(j)|}{i}\binom{|L(j)|}{l}}{\binom{|I(j)|}{i}\binom{|L(j)|-1}{l-1}} \\ &= \max_{l\in\{d_j+1,\dots,|L(j)|\}}\frac{|L(j)|}{l} = \frac{|L(j)|}{d_j+1}1_{[\![|L(j)|\geq d_j+1]\!]} \end{aligned}$$

and similarly

$$\begin{aligned} C''_j &= \max_{(i,l)\in\Psi''(j)}\frac{\binom{|I(j)|}{i}\binom{|L(j)|-1}{l}}{\binom{|I(j)|}{i}\binom{|L(j)|-1}{l} + \binom{|I(j)|}{i}\binom{|L(j)|-2}{l-1}} \\ &= \max_{l\in\{d_j-1,\dots,|L(j)|-1\}}\frac{\binom{|L(j)|-1}{l}}{\binom{|L(j)|-1}{l} + \frac{l}{|L(j)|-1}\binom{|L(j)|-1}{l}} \\ &= \max_{l\in\{d_j-1,\dots,|L(j)|-1\}}\frac{|L(j)|-1}{|L(j)|-1+l} \\ &= \frac{|L(j)|-1}{|L(j)|+d_j-2}. \end{aligned}$$

Thus, the statement follows from Thm. B.5. □

To conclude this section, we state in the following corollary worst-case consequences of Thm. B.5 and Thm. B.3. They show in particular that – in both learning scenarios with and without indifferences – identifying the Copeland winner of $\mathbf{P}$ requires $\Omega(n^2)$ samples in the worst case.

Corollary B.8. *Let $f : \mathbb{N} \to \mathbb{N}$ with $1 \leq f(n) \leq \frac{n}{2} - 1$ for all $n \in \mathbb{N}$ and $f(n) \in o(n)$ as $n \to \infty$ and let $\Delta \in (0, 1/6)$ be arbitrary.*

(i) There exists a sequence $(\mathbf{P}^n)_{n\in\mathbb{N}}$ of instances without indifferences with $(P^n)^{\succ}_{i,j} \in \{1/2 \pm \Delta\}$ for all $i, j \in \mathcal{A}$ and $\mathrm{CP}^*(\mathbf{P}^n) \geq \lceil \frac{n}{2} + f(n) \rceil$ such that

$$\mathbb{E}_{\mathbf{P}^n}\left[\tau^{\mathrm{A}}\right] \in \Omega\left(\frac{n^2}{f(n)\Delta^2} \ln \frac{1}{\delta}\right)$$

for any algorithm A *that correctly identifies the Copeland winner of any $\mathbf{P}$ without indifferences with confidence $1 - \delta$.*

(ii) There exists a sequence $(\mathbf{P}^n)_{n\in\mathbb{N}}$ of instances with $(P^n)^{\succ}_{i,j}, (P^n)^{\cong}_{i,j}, (P^n)^{\prec}_{i,j} \in \{1/3 - \Delta, 1/3 + 2\Delta\}$ for all $i, j \in \mathcal{A}$ and $\mathrm{CP}^*(\mathbf{P}^n) \geq \lceil \frac{n}{2} + f(n) \rceil$ such that

$$\mathbb{E}_{\mathbf{P}^n}\left[\tau^{\mathrm{A}}\right] \in \Omega\left(\frac{n^2}{f(n)\Delta^2} \ln \frac{1}{\delta}\right)$$

for any algorithm A *that correctly identifies the Copeland winner of any $\mathbf{P}$ with indifferences with confidence $1 - \delta$.*

Proof. Let f and Δ be as stated above. We start with the proof of (i). By assumption on f there exists $n_0 \in \mathbb{N}$ with $n - 1 - \lceil \frac{n}{2} + f(n) \rceil \geq 1$ and $\lfloor \frac{n-1}{2} \rfloor \geq f(n) + 2$ for all $n \geq n_0$. For $n < n_0$, let $\mathbf{P}^n$ be an arbitrary allowed instance. For arbitrary but fixed $n \geq n_0$, fix a set $L_n \subseteq \mathcal{A}$ of size $|L_n| = n - 1 - \lceil \frac{n}{2} + f(n) \rceil$ and define $\mathbf{P}^n = ((P^n)^{\succ}_{x,y})_{x,y}$ via

$$(P^n)^{\succ}_{1,y} \coloneqq \begin{cases} 1/2 + \Delta, & \text{if } y \in \mathcal{A} \setminus L_n, \\ 1/2 - \Delta, & \text{if } y \in L_n \end{cases}$$

for $2 \leq y \leq n$ and

$$(P^n)^{\succ}_{x,y} \coloneqq \begin{cases} 1/2 + \Delta, & \text{if } x + y \text{ is even}, \\ 1/2 - \Delta, & \text{if } x + y \text{ is odd} \end{cases}$$

for $2 \leq x < y \leq n$. Then, $\mathbf{P}^n$ has no indifferences. Moreover, $\mathrm{CP}(\mathbf{P}^n, 1) = n - 1 - |L_n| = \lceil \frac{n}{2} + f(n) \rceil$ and $\mathrm{CP}(\mathbf{P}^n, j) \geq \lfloor \frac{n-1}{2} \rfloor - 1$ for $j \in \mathcal{A} \setminus \{1\}$ hold, which shows $\mathcal{C}(\mathbf{P}^n) = \{1\}$ and $d_j(\mathbf{P}^n) = \mathrm{CP}(\mathbf{P}^n, 1) - \mathrm{CP}(\mathbf{P}^n, j) \leq f(n) + 1$. By construction, $\lfloor \frac{n-1}{2} \rfloor \leq |L(\mathbf{P}^n, j)| \leq \lceil \frac{n-1}{2} \rceil + 1$ is fulfilled, and thus by choice of n_0 also $|L(\mathbf{P}^n, j)| \geq \lfloor \frac{n-1}{2} \rfloor \geq f(n) + 2 = d_j(\mathbf{P}^n) + 1$ holds. Using that Lemma B.4 and $\Delta < 1/6$ imply

$$\kappa_{x,y}(\mathbf{P}) = \mathrm{kl}(P^{\succ}_{x,y}, P^{\prec}_{x,y}) \leq \frac{(P^{\succ}_{x,y} - P^{\prec}_{x,y})^2}{P^{\prec}_{x,y}(1 - P^{\prec}_{x,y})} = \frac{4\Delta^2}{(1/2 - \Delta)(1/2 + \Delta)} \leq 16\Delta^2,$$

Thm. B.3 yields

$$\begin{aligned}
\mathbb{E}_{\mathbf{P}^n}\left[\tau^{\mathrm{A}}\right] &\geq \ln \frac{1}{2.4\delta} \sum\nolimits_{j \neq i^*} \frac{|L(\mathbf{P}^n, j)|}{d_j(\mathbf{P}^n)} \min_{z \in L(j)} \frac{1}{\kappa_{j,z}(\mathbf{P}^n)} \\
&\geq \frac{1}{16\Delta^2} \left(\ln \frac{1}{2.4\delta}\right) \sum_{j \neq i^*} \frac{\lfloor (n-1)/2 \rfloor}{f(n) + 1} \\
&\geq \frac{1}{32\Delta^2} \left(\ln \frac{1}{2.4\delta}\right) \sum_{j \neq i^*} \frac{n - 1}{f(n) + 1} \\
&\geq \frac{1}{32\Delta^2} \left(\ln \frac{1}{2.4\delta}\right) \frac{(n-1)^2}{f(n) + 1},
\end{aligned}$$

which concludes the proof of (i).

To prove (ii), define n_0 as before and fix allowed arbitrary $\mathbf{P}^n$ for $n < n_0$. For arbitrary but fixed $n \geq n_0$, fix

again $L_n \subseteq \mathcal{A}$ of size $n - 1 - \lceil \frac{n}{2} + f(n) \rceil$ and define the instances $\mathbf{P}^n = ((P^n)^{\succ}_{x,y})_{x,y}$ via

$$(P^n)^{\succ}_{1,y} \coloneqq \begin{cases} 1/3 + 2\Delta, & \text{if } y \in \mathcal{A} \setminus L_n, \\ 1/3 - \Delta, & \text{if } y \in L_n, \end{cases}$$
$$(P^n)^{\succ}_{y,1} \coloneqq \begin{cases} 1/3 - \Delta, & \text{if } y \in \mathcal{A} \setminus L_n, \\ 1/3 + 2\Delta, & \text{if } y \in L_n, \end{cases}$$

for $2 \leq y \leq n$ and

$$(P^n)^{\succ}_{x,y} \coloneqq \begin{cases} 1/3 + 2\Delta, & \text{if } (x+y \text{ is even and } x < y) \text{ or } (x+y \text{ is odd and } x > y) \\ 1/3 - \Delta, & \text{if } (x+y \text{ is odd and } x < y) \text{ or } (x+y \text{ is even and } x > y) \end{cases}$$

for distinct $x, y \in \{2, \dots, n\}$. Then, $\mathbf{P}^n$ has indifferences and fulfills $(P^n)^{\cong}_{x,y} = 1/3 - \Delta$ for any distinct $x, y \in \mathcal{A}$, which directly implies $I(j) = \emptyset$ for any $j \in \mathcal{A}$. By construction, we see similarly as above $\mathrm{CP}(\mathbf{P}^n, 1) = \lceil \frac{n}{2} + f(n) \rceil$, $\mathcal{C}(\mathbf{P}^n) = \{1\}$ and $d_j(\mathbf{P}^n) \leq f(n) + 1$ as well as $\lfloor \frac{n-1}{2} \rfloor \leq |L(\mathbf{P}^n, j)| \leq \lceil \frac{n-1}{2} \rceil + 1$ and $|L(\mathbf{P}^n, j)| \geq d_j(\mathbf{P}^n) + 1$ for $j \in \mathcal{A} \setminus \{1\}$. Regarding the construction of $\mathbf{P}^n$, Lemma B.4 implies due to $0 < \Delta < 1/6$ the estimate

$$\begin{aligned} D_{x,y}(\mathbf{P}^n) &= \mathrm{KL}\left(\left(\frac{1}{3} + 2\Delta, \frac{1}{3} - \Delta, \frac{1}{3} - \Delta\right), \left(\frac{1}{3} - \Delta, \frac{1}{3} + 2\Delta, \frac{1}{3} - \Delta\right)\right) \\ &\leq \frac{(3\Delta)^2}{1/3 - \Delta} + \frac{(3\Delta)^2}{1/3 + 2\Delta} \leq 3^2(6+3)\Delta^2 = 81\Delta^2. \end{aligned}$$

With the use of Cor. B.7 instead of Thm. 6, following the same argumentation as in the proof of (i) thus lets us conclude

$$\mathbb{E}_{\mathbf{P}^n}\left[\tau^{\mathrm{A}}\right] \geq \frac{1}{162\Delta^2} \left(\ln \frac{1}{2.4\delta}\right) \frac{(n-1)^2}{f(n)+1}.$$

□

C POCOWISTA ANALYSIS

Theorem 3.1. *Let* A := POCOWISTA. *For any dueling bandits problem with indifferences characterized by* $\mathbf{P} = ((P^{\succ}_{i,j}, P^{\cong}_{i,j}, P^{\prec}_{i,j}))_{i<j}$, *such that there exists no pair* $i, j \in \mathcal{A}$ *with* $i \neq j$ *and* $P^{\succ}_{i,j} = P^{\succ}_{j,i} = 1/3$, *it holds that*

$$\mathbb{P}\big(\hat{i}_{\mathrm{A}} \in \mathcal{C}(\mathbf{P}) \text{ and } \tau^{\mathrm{A}}(\mathbf{P}) \leq t(\mathbf{P}, \delta)\big) \geq 1 - \delta,$$

where

$$t(\mathbf{P}, \delta) = \sum_{i<j} t_0\big((P^{\succ}_{i,j}, P^{\cong}_{i,j}, P^{\prec}_{i,j}), \delta/\tbinom{n}{2}\big),$$

$$t_0\big((p_1, p_2, p_3), \delta\big) = \frac{c_1 p_{(1)} \ln\left(\sqrt{\frac{2\,c_2}{\delta}} \frac{p_{(1)}}{(p_{(1)} - p_{(2)})}\right)}{(p_{(1)} - p_{(2)})^2},$$

$p_{(1)} \geq p_{(2)} \geq p_{(3)}$ *is the order statistic of* $p_1, p_2, p_3 \in [0, 1]$ *with* $\sum_{i=1}^{3} p_i = 1$ *and* $c_1 = 194.07$ *and* $c_2 = 79.86$.

Proof. Let P be a categorical distribution with categories c_1, c_2 and c_3 having probabilities p_1, p_2 and p_3, i.e., $P(c_i) = p_i$ for $i = 1, 2, 3$, such that $p_{(1)} > p_{(2)} \geq p_{(3)}$. Theorem 9 in Jain et al. (2022) states that running PPR-1v1 with $\tilde{\delta} \in [0, 1]$ as the desired error probability for identifying the mode of P, leads to a sample complexity of at most

$$t_0\big((p_1, p_2, p_3), \tilde{\delta}\big) = \frac{c_1 p_{(1)} \ln\left(\sqrt{\frac{2\,c_2}{\tilde{\delta}}} \frac{p_{(1)}}{(p_{(1)} - p_{(2)})}\right)}{(p_{(1)} - p_{(2)})^2},$$

for identifying the mode with probability at least $1 - \tilde{\delta}$.

In the worst case, POCOWISTA has to use PPR-1v1 with an error probability of $\tilde{\delta} = \delta/\binom{n}{2}$ for each ternary distribution $P_{i,j}$, where $i < j$. Recall that $P_{i,j}$ is a categorical distribution with three categories $c_1 :=$ "$i \succ j$", $c_2 :=$ "$i \cong j$" and $c_3 :=$ "$i \prec j$" having probabilities $P^{\succ}_{i,j}, P^{\cong}_{i,j}$ and $P^{\prec}_{i,j}$. Moreover, by assumption each $P_{i,j}$ has a unique mode, so that we can use Theorem 9 in Jain et al. (2022). As for each epoch the probability of making an incorrect decision or exceeding $t_0\big((P^{\succ}_{i,j}, P^{\cong}_{i,j}, P^{\prec}_{i,j}), \delta/\binom{n}{2}\big)$ many samples is bounded by $\delta/\binom{n}{2}$, the probability that the overall sample complexity of POCOWISTA exceeds

$$\sum_{i<j} t_0\big((P^{\succ}_{i,j}, P^{\cong}_{i,j}, P^{\prec}_{i,j}), \delta/\tbinom{n}{2}\big)$$

is bounded by $\sum_{i<j} \delta/\binom{n}{2} = \delta$ by means of the union bound.

Next, as the modes of the ternary distributions are all correctly identified (with probability $\delta/\binom{n}{2}$), the score updates are all correct in the sense that

$$\widehat{CP}(i) \leq CP(i) \leq \overline{CP}(i) \quad \forall i \in \mathcal{A}$$

holds with probability at least $1 - \delta$. Thus, the termination criterion of POCOWISTA implies that

$$CP(\hat{i}_{\mathrm{A}}) \geq \widehat{CP}(\hat{i}_{\mathrm{A}}) \geq \overline{CP}(j) \geq CP(j) \,\forall j \in \mathcal{A} \setminus \{\hat{i}_{\mathrm{A}}\}$$

holds with probability at least $1 - \delta$, so that $\hat{i}_{\mathrm{A}}$ is an element of the Copeland set $\mathcal{C}(\mathbf{P})$. □

D TRA-POCOWISTA ANALYSIS

Theorem 4.2. *Let* A := TRA-POCOWISTA. *For any dueling bandits problem with indifferences as in Theorem 3.1 which in addition is transitive according to Def. 4.1, it holds that*

$$\mathbb{P}\big(\hat{i}_{\mathrm{A}} \in \mathcal{C}(\mathbf{P}) \text{ and } \tau^{\mathrm{A}}(\mathbf{P}) \leq \tilde{t}(\mathbf{P}, \delta)\big) \geq 1 - \delta,$$

where

$$\tilde{t}(\mathbf{P}, \delta) = \textstyle\sum_{e=1}^{E} t_0\big((P^{\succ}_{i_e,j_e}, P^{\cong}_{i_e,j_e}, P^{\prec}_{i_e,j_e}), \delta/n\big),$$

t_0 *is as in* (5) *and* $E \leq n$.

Proof. Following the lines of the proof of Theorem 3.1, we only need to verify that TRA-POCOWISTA runs for at most n many epochs. For this purpose, we only need show that the termination criterion of TRA-POCOWISTA (see line 3 in Algo. 4) is fulfilled after at most n many epochs. If the modes of the ternary distributions are all correctly identified (with probability δ/n) and transitivity as in Def. 4.1 holds, then the score updates are all correct in the sense that

$$\widehat{CP}(i) \leq CP(i) \leq \overline{CP}(i) \quad \forall i \in \mathcal{A}$$

holds with probability at least $1 - \delta$.

For sake of convenience, define $\widehat{CP}_e(i)$ as the estimated Copeland score for arm $i \in \mathcal{A}$ *before* the update in epoch e is made (i.e., the value before line 7 in Algo. 4) and likewise $\overline{CP}_e(i)$. The score updates (Algo. 5) as well as the choice of j_e (line 5 in Algo. 4) imply that

$$\widehat{CP}_e(j_e) \geq \widehat{CP}_{e-1}(j_{e-1}) + \widehat{CP}_{e-1}(i_{e-1}) + 1/2 \tag{17}$$

for any epoch e. Indeed, if one arm dominates the other in epoch $e-1$, then it's estimated Copeland score is updated to $\widehat{CP}_{e-1}(j_{e-1}) + \widehat{CP}_{e-1}(i_{e-1}) + 1$, while in case of an indifference the updated value corresponds to the right-hand side of (17). As j_e is (one of) the arm(s) with largest estimated Copeland score, it has consequently an estimated Copeland score in epoch e of at least the right-hand side of (17). Further, it holds that $\widehat{CP}_2(j_2) \geq 1/2$. If TRA-POCOWISTA has not terminated after epoch $n-1$ it must hold that there exists some epoch $s \in \{1, \ldots, n-1\}$ such that $\widehat{CP}_{s+1}(i_s) \geq 1/2$, as otherwise the "second arm" j_e has dominated in each epoch the "first arm" i_e and has stayed the same for all epochs, in which case TRA-POCOWISTA terminates and returns j_e. Combining this with (17) it holds that $\widehat{CP}_{n+1}(j_n) \geq n/2$. We distinguish now the three different cases for the outcome between the compared arms i_n and j_n in epoch n (i.e., line 6 in Algo. 4) and show that in each case TRA-POCOWISTA terminates, so that the overall number of epochs E is bounded by n.

Case 1: $k = 1$, i.e., i_n dominated j_n.

This implies that $\widehat{CP}_{n+1}(i_n) \geq \widehat{CP}_{n+1}(j_n)$ and in particular $\widehat{CP}_{n+1}(i_n) \geq \widehat{CP}_n(i_n)$. By choice of i_n it holds that for any $i \in \mathcal{A}$

$$\overline{CP}_{n+1}(i) \leq \overline{CP}_n(i) \leq \overline{CP}_n(i_n) = n - \widehat{CP}_n(i_n) \leq n - \widehat{CP}_{n+1}(i_n) \leq n/2 \leq \widehat{CP}_{n+1}(i_n).$$

Thus, the termination criterion of TRA-POCOWISTA (see line 3 in Algo. 4) is fulfilled after epoch n, as i_n fulfills the criterion.

Case 2: $k = 2$, i.e., and indifference between i_n and j_n.

This implies that $\widehat{CP}_{n+1}(i_n) = \widehat{CP}_{n+1}(j_n)$ and in particular $\widehat{CP}_{n+1}(i_n) \geq \widehat{CP}_n(i_n)$. By choice of i_n it holds that for any $i \in \mathcal{A}$

$$\begin{aligned}\overline{CP}_{n+1}(i) \leq \overline{CP}_n(i) \leq \overline{CP}_n(i_n) = n - \widehat{CP}_n(i_n) &\leq n - \widehat{CP}_{n+1}(i_n) \\ &= n - \widehat{CP}_{n+1}(j_n) \leq n/2 \leq \widehat{CP}_{n+1}(j_n).\end{aligned}$$

Thus, the termination criterion of TRA-POCOWISTA (see line 3 in Algo. 4) is fulfilled after epoch n, as j_n fulfills the criterion.

Case 3: $k = 3$, i.e., j_n dominated i_n.

In this case, the potential Copeland score of i_n in epoch $n+1$ is at most $n - \widehat{CP}_{n+1}(j_n)$. By choice of i_n it holds that for any $i \in \mathcal{A}$

$$\overline{CP}_{n+1}(i) \leq \overline{CP}_n(i) \leq \overline{CP}_n(i_n) \leq n - \widehat{CP}_{n+1}(j_n) \leq n/2 \leq \widehat{CP}_{n+1}(j_n).$$

Thus, the termination criterion of TRA-POCOWISTA (see line 3 in Algo. 4) is fulfilled after epoch n, as j_n fulfills the criterion.

□