The Distortion of Public-Spirited Participatory Budgeting

Mark Bedaywi[1], Bailey Flanigan[2], Mohamad Latifian[1], and Nisarg Shah[1]

[1]University of Toronto
[2]Harvard University

Abstract

Participatory budgeting (PB) is an increasingly popular tool for democratically allocating limited budgets to public-good projects. In PB, constituents vote on their preferred projects via *ballots*, and then an *aggregation rule* selects a set of projects whose total cost fits within the budget. Recent work studies how to design PB ballots and aggregation rules that yield *low distortion* outcomes (informally, outcomes with high social welfare). Existing distortion bounds, however, rely on strong assumptions that restrict voters' latent utilities. We prove that low distortion PB outcomes can be achieved by dropping these assumptions and instead leveraging the established idea that voters can be *public-spirited*: they may consider others' interests alongside their own when voting.

Flanigan et al. [2023] prove that in public-spirited *single-winner* voting (the special case of PB where exactly one project can be funded) with ranking ballots, deterministic aggregation rules can achieve constant distortion. Our first contribution is to extend this analysis to PB; there, we prove that the best distortion permitted by deterministic rules with ranking ballots grows *linearly* in the number of projects. We find that this impossibility—a problem in practice, where m is often large—holds for other known ballots as well. Our second contribution is the design of a new PB ballot format that breaks this linear distortion barrier. This ballot asks voters to rank a predetermined set of *entire feasible bundles* of projects. We design multiple protocols for implementing these ballots, each striking a different trade-off between the number of bundles voters must rank and the distortion: with m bundles, we get sublinear distortion; with polynomial bundles, we get logarithmic distortion; and with pseudopolynomial bundles, we get constant distortion.

1 Introduction

Governments at all scales regularly face the question: Which public-good projects — e.g., building bike paths or installing streetlamps — should they fund with their limited budget? To make such decisions democratically, governments are increasingly using *participatory budgeting* (PB), where a group of constituents convenes to vote on which projects their government should fund. In PB, the government supplies a budget B and a list of m potential projects with corresponding costs. Voters submit their preferences via *ballots*, which are then aggregated via an *aggregation rule*. The output of this rule is a set of projects to be funded whose total cost is at most B. PB is now used all over the world to allocate public funds[1] [Participedia, 2023, De Vries et al., 2022, Wampler et al., 2021].

When designing the PB process, one goal that many consider important is ensuring that the ultimate allocation of funds has high societal benefit. As have many others (e.g., Benadè et al. [2021]), we formalize the "societal benefit" of an allocation by its *utilitarian social welfare*: the total utility it gives to all voters. As such, we adopt the standard model of latent additive utilities: each voter i has *utility* $u_i(a) \in \mathbb{R}_{\geqslant 0}$ for each project a, and her total utility for a set of projects S being funded is $u_i(S) = \sum_{a \in S} u_i(a)$. Then, the *social welfare* of S is equal to $\mathsf{sw}(S) = \sum_{i \in N} u_i(S)$.

[1]See https://en.wikipedia.org/wiki/List_of_participatory_budgeting_votes for a list of use cases.

If voters' utilities were known, choosing the maximum-welfare allocation would amount to solving the knapsack problem. However, in practice voters' preferences can only be elicited more coarsely through *ballots*. One popular PB ballot format is *rankings by value*, where voters rank all individual projects. It is not hard to see that this ballot format loses far too much information about voters' utilities: suppose there are two projects a and b, and they both cost B so we can fund only one. If the utilities for (a, b) are $(1, 0)$ for half the population and $(0, X)$ for the other half (so, the welfare of b is X times that of a), the resulting ballots will be symmetric. Any deterministic aggregation rule will choose a without loss of generality, suffering unbounded welfare loss as X grows large; the best a randomized aggregation rule can do is to choose a project uniformly at random.

This example illustrates a prohibitive impossibility: in the worst case, *any* deterministic rule over ranking ballots will select an outcome with arbitrarily sub-optimal social welfare (and while randomized rules can do better, they do so trivially by ignoring voters' preferences). In fact, this impossibility holds for all widely-studied ballots in the PB literature due to the same example (see Appendix B). Formally, this sub-optimality is captured with the *distortion*: the worst-case (over latent utilities) ratio between the best possible social welfare and that of the outcome. Existing work sidesteps this impossibility by assuming that each voter's utilities sum to 1 [Benadè et al., 2021]. Although this permits bounded distortion in theory, it is unclear whether these bounds apply in practice: for example, this assumption may not hold in the likely case that all public-good projects on the ballot will more greatly benefit lower-income constituents.

Fortunately, recent work by Flanigan et al. [2023] offers a source of hope: under unrestricted utilities, they achieve low distortion in single-winner elections by leveraging the idea that voters may be *public-spirited*: when casting their ballots, voters consider others' interests in addition to their own. While it is not clear that such public-spirited voting behavior would be reliably present in the wild, as Flanigan et al. argue in-depth, research suggests that public spirit can be cultivated via democratic deliberation [Kinder and Kiewiet, 1981, Wang et al., 2020, Gastil et al., 2010] — *a practice that is already commonplace in PB elections* [Participedia, 2023, De Vries et al., 2022]. The possibility of cultivating public spirit among PB participants motivates our main research question:

> **Question:** *If voters are public-spirited, can we design PB elections that achieve low (perhaps even constant) distortion with unrestricted voter utilities?*

An affirmative answer to this question would support deliberation as a practicable approach to achieving higher-welfare PB outcomes. While this question builds on Flanigan et al. [2023], answering it will require fundamentally new methods because Flanigan *et al*'s results apply only to single-winner voting, a substantially restricted version of the PB setting in which all projects cost B.

1.1 Our Contributions

In the public-spirited voting model of Flanigan et al. [2023], each voter i has some *public spirit level* $\gamma_i \in [0, 1]$. She then evaluates each alternative (project) $a \in [m]$ according to her *public-spirited (PS) value* $v_i(a) := (1 - \gamma_i)u_i(a) + \gamma_i \mathsf{sw}(a)$, the convex combination of her own utility and the social welfare. Note that this generalizes the standard model in which γ_i is assumed to be 0 (that is, i evaluates a based on just her own utility). We extend Flanigan et al.'s model to PB by assuming additive valuations, so i's PS-value for a set of projects S is simply $v_i(S) = \sum_{a \in S} v_i(a)$. Like Flanigan et al. [2023], our distortion bounds are parameterized by $\gamma_{\min} = \min_i \gamma_i$, the minimum public spirit level of any voter. For simplicity, we summarize our main results below assuming $\gamma_{\min}$ is a constant.

We first consider the canonical *rankings* ballot format also studied by Flanigan et al. [2023], where voters rank all individual projects by their value. For our first main contribution, we show that the best distortion achievable by any deterministic PB rule using ranking ballots is $\Theta(m)$; for randomized rules, it is $\Theta(\log m)$ (Section 3). Our upper bounds are proven via general reductions from PB to single-winner elections, which may be of independent interest; proving and applying these reductions also leads to new results for the single-winner setting. Our lower bounds imply a fundamental separation between the single-winner voting and PB under public spirit: in single-winner voting with ranking ballots, there are deterministic rules that achieve *constant* distortion [Flanigan et al., 2023].

Ranking-by-value		Public-Spirit	Unit-Sum
Single-winner	Deterministic	$\Theta\left(1/\gamma_{\min} \cdot \min\{m, 1/\gamma_{\min}\}\right)$	$\Theta(m^2)$
	Randomized	$\Theta\left(\min\{m, 1/\gamma_{\min}\}\right)$	$\Theta(\sqrt{m})$
Participatory Budgeting	Deterministic	$\Omega\left(m/\gamma_{\min}\right)$, $\mathcal{O}\left(m/\gamma_{\min} \cdot \min\{m, 1/\gamma_{\min}\}\right)$	$\Theta(m^2)$
	Randomized	$\Omega\left(\log m\right)$, $\mathcal{O}\left(\min\{m, (\log m)/\gamma_{\min}\}\right)$	$\Omega(\sqrt{m})$, $\mathcal{O}\left(\sqrt{m}\log m\right)$

Table 1: Asymptotic (in $m, \gamma_{\min}$) distortion bounds for rankings-by-value, comparing results for Single-winner and Participatory Budgeting ballots. The unit-sum results are derived in Benadè et al. [2021] and are included for comparison.

Our linear lower bound of $\Omega(m)$ for deterministic rules is especially bad news: deterministic rules are typically used in practice, and in many PB elections, m can be in the hundreds or thousands (see Footnote 1). Focusing henceforth on deterministic rules, we pursue sublinear distortion by broadening our consideration to other ballot formats. Unfortunately, in Section 4 we find that none of the main PB ballot formats studied in past work (*rankings by value-for-money*, *k-approval*, *knapsack*, and *threshold approval* — see Benadè et al. [2021] for an overview) permit sublinear distortion (in m). Establishing this linear distortion barrier faced by existing PB ballot formats under public spirit constitutes our next main contribution.

Motivated by this impossibility, in Section 5 we introduce *rankings of predefined bundles*, a novel PB ballot format that asks voters to rank *entire bundles* of projects rather than individual projects. We show that with carefully-chosen bundles, this ballots format *does* permit sublinear distortion in PB. Further, with sufficient (but still fairly limited) information about voters' preferences elicited ahead of time, they can even drop the distortion to *constant*. We study three protocols for using this ballot format, each eliciting more information than the last in exchange for lower distortion:

1. *Protocol 1* permits $O(\sqrt{m})$ distortion and asks voters to rank at most m feasible bundles.
2. *Protocol 2* permits $O(\log m)$ distortion and is a two-round protocol: in round 1 voters rank individual projects, then in round 2 they rank at most $\log m$ bundles crafted based on their votes in round 1.
3. *Protocol 3* permits $O(1)$ distortion. It is similar to Protocol 2 except that in round 2, voters rank $O(m^{1+\log\log m})$ feasible bundles in the worst case.

While Protocol 3 may be impractical in the worst case, we provide empirical evidence that in realistic PB elections, even Protocol 3 would require voters to rank less than m bundles (Section 6). From a theoretical standpoint, Protocol 3 demonstrates the possibility of constant distortion with only pseudo-polynomial bits of information, raising the tantalizing open question of whether constant distortion can be achieved with only *polynomially* many bits.

1.2 Related Work

Our work builds most directly on Flanigan et al. [2023], who introduced the public-spirited model. We generalize their work from single-winner elections to the more general PB setting; and while they consider only deterministic aggregation rules, we additionally consider randomized rules. In the process, we prove new insights for the single-winner case. Our work also directly builds on the works of Benadè et al. [2021], who analyzed distortion in PB under the unit-sum utilities assumption. We contrast our bounds to those achievable in their model in Section 4.

Procaccia and Rosenschein [2006] introduce the distortion framework in single-winner elections under the unit-sum assumption. We now know that the best distortion achievable by deterministic and randomized rules for this special case are $\Theta(m^2)$ [Caragiannis and Procaccia, 2011, Caragiannis et al., 2017] and $\Theta(\sqrt{m})$ [Boutilier et al., 2015, Ebadian et al., 2022], respectively. Optimal distortion bounds have also been identified for k-committee selection [Caragiannis et al., 2017, Borodin et al., 2022], which is still a special case

	Public-Spirit	Unit-Sum
k-approvals $(k > 1)$	∞	∞
1-*approval*	$\Theta\left({m^2}/{\gamma_{\min}}\right)$	$\Theta(m^2)$
Knapsack	$\Omega\left({m}/{\gamma_{\min}}\right)$, $\mathcal{O}\left({m^3}/{\gamma_{\min}^2}\right)$	$\Omega({2^m}/{\sqrt{m}})$, $\mathcal{O}\left(m2^m\right)$
Ranking of Predefined Bundles (One Round)	$\mathcal{O}\left({\sqrt{m}}/{\gamma_{\min}^2}\right)$	$\Omega(m^2)$
Ranking of Predefined Bundles (Two Round)	$\mathcal{O}\left({(\log m)}/{\gamma_{\min}^4}\right)$	$\Omega(m^2)$

Table 2: Asymptotic (in $m, \gamma_{\min}$) deterministic distortion bounds across ballot formats other than ranking-by-value. The bottom rows are the new ballots introduced in this paper. The unit-sum results are derived in Benadè et al. [2021] and are included for comparison.

of PB. Some have studied unit-range utilities or metric costs in place of unit-sum utilities [Filos-Ratsikas and Miltersen, 2014, Anshelevich et al., 2018], but all these models directly restrict voters' cardinal preferences. For further details, see the survey of Anshelevich et al. [2021].

Multiple approaches other than distortion have been studied for PB. The axiomatic approach has been used to identify aggregation rules satisfying desirable axioms such as various monotonicity properties Talmon and Faliszewski [2019], Baumeister et al. [2020], Rey et al. [2020]. Another important consideration in PB is whether the allocation of funds is fair with respect to (groups of) voters [Fain et al., 2018, Peters et al., 2021, Brill et al., 2023]. For further details, we suggest the survey of Rey and Maly [2023] and the book chapter of Aziz and Shah [2021].

2 Model and Preliminaries

We let $[k] = \{1, \ldots, k\}$ for any $k \in \mathbb{N}$, and for a finite set S, let $\Delta(S)$ denote the set of probability distributions over S. We introduce the general framework of participatory budgeting (PB) first, and later introduce single-winner and multiwinner voting as its special cases.

Alternatives A, budget B, and costs c. In a PB instance, there is a set of n *voters* $N = [n]$ and a set of m *alternatives* (projects) A. We denote voters by i, j and alternatives by a, b. There is a total budget of B, which is normalized to 1 without loss of generality, and a cost function $c : A \to [0, 1]$, where $c(a)$ is the *cost* of a. Slightly abusing notation, denote by $c(S) = \sum_{a \in S} c_a$ the total cost of alternatives in S.

Utilities U. Each voter $i \in N$ has a *utility* for each alternative $a \in A$ denoted by $u_i(a) \in \mathbb{R}_{\geqslant 0}$. Together, these utilities form a *utility matrix* $U \in \mathbb{R}_{\geqslant 0}^{n \times m}$. The *social welfare* of $a \in A$ w.r.t. utility matrix U is $\mathsf{sw}(a, U) = \sum_{i \in N} u_i(a)$; for any *set* of alternatives $S \subseteq A$, $\mathsf{sw}(S, U) = \sum_{a \in S} \mathsf{sw}(a, U)$. We use $\mathsf{sw}(a)$ or $\mathsf{sw}(S)$ when U is clear from context.

PS-levels $\vec{\gamma}$ and PS-values V. Following Flanigan et al. [2023], we assume that each voter $i \in N$ has a *public spirit (PS) level* $\gamma_i \in [0, 1]$, and, together, these PS-levels form the PS-vector $\vec{\gamma} \in [0, 1]^n$. For a given $\vec{\gamma}$, we let $\gamma_{\min} := \min_{i \in N} \gamma_i$ be the minimum level of public spirit among voters. Each voter i evaluates each alternative a by her *PS-value* $v_i(a)$, a convex combination of her personal utility $u_i(a)$ and $\mathsf{sw}(a)/n$, the average voter's utility for a:

$$v_i(a) = (1 - \gamma_i) \cdot u_i(a) + \gamma_i \cdot \mathsf{sw}(a)/n.$$

Note that this model does not restrict voters' utilities; rather, it assumes something about how they translate their utilities into votes. These PS-values form the *PS-value matrix* $V_{\vec{\gamma}, U} \in \mathbb{R}_{\geqslant 0}^{n \times m}$. For each $S \subseteq A$, let $v_i(S) := \sum_{a \in S} v_i(a)$.

Instances and special cases. An *instance* of the PB problem is composed of the elements defined so far: $I = (A, B, c, U, \vec{\gamma})$. Let $\mathcal{I}$ be the set of all PB instances. Let $\mathcal{F}(I) = \{S \subseteq A\colon c(S) \leqslant 1\}$ be the set of *budget-feasible* subsets of A in instance I. $\mathcal{F}$ will be a generic such set.

We will sometimes build our results using ideas from *k-committee selection* and *single-winner voting*—two restrictions of the PB setting. Formally, all instances of *k-committee selection* are captured when $\mathcal{I}$ is exclusively restricted to instances I with $c(\cdot) = 1/k$ (i.e., all alternatives have cost $1/k$), so $\mathcal{F}(I)$ consists of all subsets of alternatives of size k. Single-winner voting is the further restriction in which $c(\cdot) = 1$. We will let $\mathcal{I}^{\text{single-win}} := \{I | c(\cdot) = 1\}$ denote the set of all single-winner voting instances.

Ballot formats. Since it is cognitively burdensome for voters to report cardinal preferences, preferences are often elicited using discrete ballots. We denote a generic ballot format as X, and let $\rho_i(I, \mathsf{X})$ be the ballot submitted by voter i in instance I. Correspondingly, let $\vec{\rho}(I, \mathsf{X}) = (\rho_1(I, \mathsf{X}), \ldots, \rho_n(I, \mathsf{X}))$ be the *vote profile*. When I, X are clear, we will drop these from the notation. In Section 5, we will design multi-round elicitation protocols; when there are multiple rounds, a "vote profile" will refer to the profile of votes collected in the final round of elicitation.

We primarily consider ordinal ballot formats, of which we study two types. First, we consider canonical *ranking* ballots ($\mathsf{X} = \mathsf{rank}$), which ask voters to rank alternatives. Then, $\rho_i(I, \mathsf{rank})$ is the permutation of A implied by the ordering of i's PS-values, so $v_i(a) > v_i(b) \Rightarrow a \succ_{\rho_i(I, \mathsf{rank})} b$ for all $a, b \in A$ (ties are broken arbitrarily, and $a \succ_\rho b$ denotes that a is ranked ahead of b in ballot ρ). We then introduce a novel ordinal ballot format, *ranking of predefined bundles* ($\mathsf{X} = \mathsf{rank\text{-}b}$), which asks voters to rank *entire bundles* of alternatives. Formally, the $\mathsf{rank\text{-}b}$ ballot format accepts an argument of a collection of predefined bundles $\mathcal{P} \subseteq \mathcal{F}(I)$; then, $\rho_i(I, \mathsf{rank\text{-}b}(\mathcal{P}))$ is a permutation of the elements of $\mathcal{P}$ such that $v_i(S) > v_i(S') \Rightarrow S \succ_{\rho_i(I, \mathsf{rank\text{-}b}(\mathcal{P}))} S'$ for all $S, S' \in \mathcal{P}$. To paint a more complete picture, we also consider other non-ordinal ballot formats in Section 4.

Aggregation rules. A (randomized) *aggregation rule* f takes as input the vote profile $\vec{\rho}$ (from the final round of elicitation, if there are multiple rounds) and returns a distribution over feasible bundles (an element of $\Delta(\mathcal{F})$). We say that f is deterministic if its output always has singleton support. We will sometimes talk about *single-winner rules* versus *PB rules*. Formally, a single-winner rule must output an element of $\Delta([m])$ while a PB rule can output any element of $\Delta(\mathcal{F})$.

We will frequently use the rule *Copeland*, so we define it here. *Copeland* is traditionally defined for the single-winner case with rank ballots. We make the natural extension here to define *Copeland* also for $\mathsf{rank\text{-}b}(\mathcal{P})$ ballot formats. All other rules we consider are defined as needed.

Definition 1 (*Copeland*)**.** Each alternative has a *score*, equal to the number of alternatives it defeats in pairwise elections. The *Copeland* winner is the one with the highest score.

When we want to choose multiple winners W, we often use the rule *Iterative Copeland*: *Copeland* is used, the winner is added to W and removed from the election, and then *Copeland* is run again on the remaining instance, and so on.

Distortion. The *distortion* measures the efficiency of a combination of a ballot format and an aggregation rule (if there are multiple rounds, the rule applies to the final round). Formally, it is the *worst-case* over all instances of the ratio between the best achievable social welfare and the output of the aggregation rule. Our bounds will depend explicitly on m and $\gamma_{\min}$, so we denote the subset of $\mathcal{I}$ with m, $\gamma_{\min}$ as

$$\mathcal{I}_{m,\gamma_{\min}} := \{I \in \mathcal{I} \,:\, |A| = m \,\wedge\, \min_{i \in N} \gamma_i = \gamma_{\min}\}.$$

Then, the distortion is defined as

$$\text{dist}_{\mathsf{X}}(f) = \sup_{n \geqslant 1} \sup_{I \in \mathcal{I}_{m,\gamma_{\min}}} \frac{\max_{S \in \mathcal{F}(I)} \mathsf{sw}(S, U)}{\mathbb{E}_{S' \sim f(\vec{\rho}(I, \mathsf{X}))} \mathsf{sw}(S', U)}.$$

We sometimes study a rule's distortion in the *single-winner* case, where the set of instances is restricted to $\mathcal{I}^{\text{single-win}}$. The *single-winner distortion* $\text{dist}_{\mathsf{X}}^{\text{single-win}}(f)$ is therefore defined identically to $\text{dist}_{\mathsf{X}}(f)$ except the second supremum is taken over $\mathcal{I}_{m,\gamma_{\min}}^{\text{single-win}}$ (analogous to $\mathcal{I}_{m,\gamma_{\min}}$).

Our distortion bounds will assume $m \geqslant 2$ and $\gamma_{\min} \in (0, 1]$. We are interested in the lowest distortion possible by *any* aggregation rule using a given ballot format; this is a measure of the usefulness of the information contained in the ballot format for social welfare maximization.

Preliminaries. For comparison, in Appendix B we prove that with no public spirit and unrestricted utilities, for all ballot formats we consider, all deterministic rules have unbounded distortion and the randomized rules have at least m distortion. In Section 4 we also contrast many of our bounds to those achievable under unit-sum utilities.

Our upper bounds will often use the following lemma, which is a simple generalization of Lemma 3.1 of Flanigan et al. [2023].

Lemma 1. *Let $A_1, A_2 \subseteq A$ be any two subsets of alternatives. Fix any $\alpha \geqslant 0$ and define $N_{A_1 \succ A_2} = \{i \in N : \alpha v_i(A_1) \geqslant v_i(A_2)\}$. Then:*

$$\frac{\mathsf{sw}(A_2)}{\mathsf{sw}(A_1)} \leqslant \alpha \cdot \left(\frac{1-\gamma_{\min}}{\gamma_{\min}} \frac{n}{|N_{A_1 \succ A_2}|} + 1\right).$$

Proof. The proof is the same as that of Lemma 3.1 by Flanigan et al. [2023]. Indeed, for each voter $i \in N_{A_1 \succ A_2}$, we know that $\alpha v_i(A_1) \geqslant v_i(A_2)$, and so,

$$\alpha \left((1-\gamma_i)u_i(A_1) + \gamma_i \frac{\mathsf{sw}(A_1)}{n}\right) \geqslant (1-\gamma_i)u_i(A_2) + \gamma_i \frac{\mathsf{sw}(A_2)}{n} \geqslant \gamma_i \frac{\mathsf{sw}(A_2)}{n}.$$

Dividing by γ_i and using the fact that $\frac{1-\gamma_i}{\gamma_i}$ is decreasing in γ_i we have,

$$\alpha \left(\frac{1-\gamma_{\min}}{\gamma_{\min}} u_i(A) + \frac{\mathsf{sw}(A_1)}{n}\right) \geqslant \frac{\mathsf{sw}(A_2)}{n}.$$

Summing over all voters in $N_{A_1 \succ A_2}$,

$$\alpha\left(\frac{1-\gamma_{\min}}{\gamma_{\min}} \sum_{i \in N_{A_1 \succ A_2}} u_i(A_1) + \frac{\mathsf{sw}(A_1)\,|N_{A_1 \succ A_2}|}{n}\right) \geqslant \frac{\mathsf{sw}(A_2)\,|N_{A_1 \succ A_2}|}{n}.$$

Using the fact that $\sum_{i \in N_{A_1 \succ A_2}} u_i(A_1) \leqslant \sum_{i \in N} u_i(A_1) = \mathsf{sw}(A_1)$,

$$\alpha \left(\frac{1-\gamma_{\min}}{\gamma_{\min}} \mathsf{sw}(A_1) + \frac{\mathsf{sw}(A_1)\,|N_{A_1 \succ A_2}|}{n}\right) \geqslant \frac{\mathsf{sw}(A_2)\,|N_{A_1 \succ A_2}|}{n}.$$

So, after some simplification, we finally get the desired upper bound:

$$\frac{\mathsf{sw}(A_2)}{\mathsf{sw}(A_1)} \leqslant \alpha \left(\frac{1-\gamma_{\min}}{\gamma_{\min}} \frac{n}{|N_{A_1 \succ A_2}|} + 1\right).$$ □

In Appendix D, we also prove a *robust* version of Lemma 1 showing that its guarantee degrades smoothly as an increasing number of voters have $\gamma_i = 0$. We further show there that we can replace Lemma 1 with its robust version in all our upper bound proofs, meaning that our upper bounds degrade smoothly as well.

3 PB with Rankings over Projects

We begin by studying *ranking* ballot format rank, the canonical ballot format in single-winner election and the one studied by Flanigan et al. [2023] (for only deterministic aggregation rules). Here we extend their results to PB for deterministic *and* randomized rules.

3.1 Deterministic Rules

We begin by upper-bounding the distortion of *Copeland* in PB, due to its strong performance in the single-winner case.

Theorem 1. $\text{dist}_{\mathsf{rank}}(\textit{Copeland}) \in \mathcal{O}\left(m/\gamma_{\min}^2\right)$.

Proof. First, we prove a general reduction that converts any deterministic single-winner rule to a deterministic PB rule.

Lemma 2 (PB → Single-Winner). *For any $d \geqslant 1$, any deterministic rule f with distortion d in single-winner voting has distortion $\text{dist}_{\mathsf{rank}}(f) \leqslant m \cdot d$ in participatory budgeting.*

Proof. The proof of this lemma is straightforward: fix any instance and let f return the singleton set $\{a\}$. Let A^* be an optimal budget-feasible set. Then,

$$\frac{\mathsf{sw}(A^*)}{\mathsf{sw}(a)} = \sum_{a^* \in A^*} \frac{\mathsf{sw}(a^*)}{\mathsf{sw}(a)} \leqslant m \cdot \max_{a^* \in A^*} \frac{\mathsf{sw}(a^*)}{\mathsf{sw}(a)} \leqslant m \cdot d. \qquad \square$$

Intuitively, a factor of m should be incurred from single-winner to PB: unlike in the single-winner case, in PB even when $\gamma_{\min} = 1$ (so all voters vote unanimously for the highest-welfare alternative), it can remain unclear how to make cost trade-offs without cardinal information, and deterministic rules still incur $\Omega(m)$ distortion. To conclude the proof, we apply this reduction via the following bound on *Copeland*'s distortion in the single-winner case, proven in Thm. 3.3 of Flanigan et al. [2023]:

$$\text{dist}_{\mathsf{rank}}^{\text{single-win}}(\textit{Copeland}) \in O(1/\gamma_{\min}^2). \tag{1}$$

□

Now, we prove a lower bound on the distortion achievable by *any* deterministic aggregation rule in the PB setting.

Theorem 2. *Every deterministic rule f has distortion*

$$\text{dist}_{\mathsf{rank}}(f) \in \Omega\left(m/\gamma_{\min}\right).$$

Proof sketch. The full proof is in Appendix A.1. The lower bound instance has only two (maximal) feasible sets, one containing a single alternative a and the other containing the remaining alternatives. A few voters rank a above the other alternatives, while all other voters do the opposite. Through detailed calculations, we show that there exist utilities for which either choice can be sub-optimal by $\Omega(m/\gamma_{\min})$. □

Together, Theorems 1 and 2 imply that *Copeland* achieves optimal dependency on m and is within a $1/\gamma_{\min}$ factor of optimal overall. There are two possible sources of this remaining gap: our use of (or analysis of) a single-winner rule via the reduction in Lemma 2, and our choice to apply the reduction specifically *Copeland*. To shed light on the role of each of these, we now prove a universal lower bound showing that at least for large m, *Copeland* is an optimal single-winner rule. This is a novel finding of independent interest for the single-winner case, given that Flanigan et al. [2023] do not give any universal lower bounds.

Theorem 3. *For all deterministic single-winner rules f,*

$$\text{dist}_{\mathsf{rank}}^{\text{single-win}}(f) \in \Omega(\min\{m/\gamma_{\min},\, 1/\gamma_{\min}^2\}).$$

The proof, found in Appendix A.2, uses a cyclic profile, where an equal number of voters submit each of m cyclically shifted permutations. The contribution is in the intricate derivation of the piecewise bound.

Comparing Equation (1) and Theorem 3 when $m \in \Omega(1/\gamma_{\min})$, *Copeland*'s distortion matches this lower bound. When $m \in o(1/\gamma_{\min})$, *Plurality*, which selects the most common first-choice alternative, provides a matching distortion upper bound of $O(m/\gamma_{\min})$ [Flanigan et al., 2023, Proposition 3.5]. Hence, this lower bound "resolves" the deterministic single-winner case in that for every regime of m and $\gamma_{\min}$, there is some voting rule that asymptotically matches it. Whether a single, $\gamma_{\min}$-oblivious rule can do so remains open for future work.

3.2 Randomized Rules

We take a parallel approach to analyze what distortion is achievable with rank ballots and *randomized* rules. Because Flanigan et al. [2023] did not study randomized single-winner rules, we must first identify and analyze a low-distortion single-winner rule anew—a result that is of independent interest for the single-winner case.

An Upper Bound With Maximal Lotteries

We select the *Maximal Lottery* rule, a single-winner rule originally proposed by Kreweras [1965].[2]

Definition 2 (*Maximal Lottery*)**.** The (directed) domination graph G consists of a vertex corresponding to each alternative $a \in A$, and an edge from a to b whenever a defeats b in a pairwise election (ties can be broken arbitrarily). The maximal lottery rule returns a distribution p over the vertices such that for any vertex $b \in A$, the probability of picking b or a vertex a with an edge to b is at least $1/2$. The existence of such a distribution can be inferred from, e.g., Farkas' lemma (see Thm. 2.4 of Harutyunyan et al. [2017]).

We now upper-bound *Maximal Lottery*'s distortion in PB:

Theorem 4. $\mathrm{dist}_{\mathsf{rank}}(\textit{Maximal Lottery}) \in O(\log(m)/\gamma_{\min})$.

Proof. We again begin by proving a general-purpose reduction to convert single-winner rules to PB rules. The reduction in the randomized case is more involved, and we do it in two steps: we first reduce PB to committee selection (Lemma 3), and then reduce that to single-winner voting (Lemma 4). The first reduction incurs an $O(\log m)$ overhead; the latter incurs none (asymptotically).

Lemma 3 (**PB → Committee**)**.** *Fix any $d \geqslant 1$. If there exists a randomized k-committee selection rule $f_{m',k}$ with distortion at most d for each $m' \leqslant m$ and $k \in [m']$, then there exists a randomized participatory budgeting rule f with distortion at most* $\mathrm{dist}_{\mathsf{rank}}(f) \leqslant 2d \cdot (\lceil \log_2(m) \rceil + 1)$.

Proof. Fix any PB instance. Split the alternatives into buckets $A_0, A_1, \ldots, A_{\lceil \log_2(m) \rceil}$, where $A_0 = \{a \in A : c_a \leqslant 1/m\}$ and for $i \neq 0$, $A_i = \left\{a \in A : 2^{i-1}/m < c_a \leqslant 2^i/m\right\}$.

The randomized PB rule f is as follows:

1. Sample $j \in \{0, 1, \ldots, \lceil \log_2(m) \rceil\}$ uniformly.
2. Consider the restricted instance with only the alternatives in A_j. That is, with $m' = |A_j|$ and $k = \min(m', \lfloor \frac{m}{2^j} \rfloor)$, use the k-committee selection rule $f_{m',k}$ to pick a set of k alternatives and return it.

Let A^* be the optimal budget-feasible subset of the alternatives, L_j^* be the optimal $\lfloor \frac{m}{2^j} \rfloor$-committee of A_j, and L_j be the one selected by the k-committee rule. For $j \neq 0$, $A^* \cap A_j$ is of size at most $\frac{m}{2^{j-1}}$. That means $\mathsf{sw}(A^* \cap A_j) \leqslant 2\mathsf{sw}(L_j^*)$ for any $j \neq 0$.

In addition, for $j = 0$, $L_0^* = A_0$ which implies $\mathsf{sw}(A^* \cap A_j) \leqslant \mathsf{sw}(L_j^*)$. Since the k-committee selection rule has distortion of d for any j, we have $\mathsf{sw}(L_j^*) \leqslant d\mathsf{sw}(L_j)$, implying that $\mathsf{sw}(A^* \cap A_j) \leqslant 2d\mathsf{sw}(L_j)$. Letting δ be the distribution of the mechanism output, we deduce the desired bound:

$$\begin{aligned}
\mathbb{E}_{L \sim \delta}[\mathsf{sw}(L)] &= \frac{1}{\lceil \log_2(m) \rceil + 1} \sum_{j=0}^{\lceil \log_2(m) \rceil} \mathsf{sw}(L_j) \\
&\geqslant \frac{1}{\lceil \log_2(m) \rceil + 1} \sum_{j=0}^{\lceil \log_2(m) \rceil} \frac{\mathsf{sw}(A^* \cap A_j)}{2d} \\
&\geqslant \frac{\mathsf{sw}(A^*)}{2d(\lceil \log_2(m) \rceil + 1)}.
\end{aligned}$$

□

[2]This rule has been rediscovered numerous times [Laffond et al., 1993, Fishburn, 1984, Fisher and Ryan, 1995, Rivest and Shen, 2010]. To the best of our knowledge, this is the first analysis of this rule's utilitarian distortion.

Lemma 4 (**Committee → Single-Winner**). *Fix any $k \in [m]$ and $d \geqslant 1$. If there exists a single-winner rule $f_{m'}$ with distortion at most d for each $m' \leqslant m$, then there exists a k-committee selection rule f with distortion at most d. If $f_{m'}$ is deterministic then so is f.*

Proof. Let $A^* = \{a_1^*, \ldots, a_k^*\}$ be the optimal budget-feasible set, sorted from highest social welfare to the lowest so that $i < j \implies \mathsf{sw}(a_i^*) \geqslant \mathsf{sw}(a_j^*)$. Let S denote the set of alternatives that our algorithm picks.

Consider the ith iteration of the procedure. Let ${a^+}_i$ be the alternative with the highest social welfare among the remaining alternatives, and a_i be the random alternative picked by the single-winner voting rule in this round. We know that $\mathsf{sw}(a_i^+) \geqslant \mathsf{sw}(a_i^*)$ and since the single-winner rule has expected distortion of d, we have $\mathbb{E}[\mathsf{sw}(a_i)] \geqslant \frac{\mathsf{sw}(a_i^+)}{d}$ which implies $\mathbb{E}[\mathsf{sw}(a_i)] \geqslant \frac{\mathsf{sw}(a_i^*)}{d}$. Summing this over all iterations and using linearity of expectation, we get that

$$\sum_{i=0}^{k} \mathbb{E}[\mathsf{sw}(a_i)] \geqslant \sum_{i=0}^{k} \mathsf{sw}(a_i^*) \,/\, d \implies \mathsf{sw}(A^*) \,/\, \mathbb{E}[\mathsf{sw}(S)] \leqslant d. \qquad \square$$

The $\log(m)$ overhead in Lemma 3 comes from partitioning the alternatives into $O(\log m)$ buckets and then applying a k-committee selection rule to a random bucket (similar approaches appear in other work, e.g. Benadè et al. [2021]). The proof of Lemma 4 generalizes ideas from an analogous reduction for deterministic rules by Goel et al. [2018].

Next, to bound the distortion of *Maximal Lottery* via Lemmas 3 and 4, we must first upper-bound its distortion in the public-spirited *single-winner* setting. The approach is to apply Lemma 1 using the insight that *Maximal Lottery* picks either the optimal alternative or an alternative that pairwise-defeats it with probability at least 1/2.

Theorem 5. $\mathrm{dist}_{\mathsf{rank}}^{\text{single-win}}(\textit{Maximal Lottery}) \in \mathcal{O}\left(1/\gamma_{\min}\right)$.

Proof. Let a^* be the optimal alternative. If we pick a^* or an alternative b that beats a^* in a pairwise election, by Lemma 1 we get distortion:

$$\frac{\mathsf{sw}(a^*)}{\mathsf{sw}(b)} \leqslant 2\frac{1-\gamma_{\min}}{\gamma_{\min}} + 1.$$

Let the set of such alternatives be $A' = \{b \in A : |\{i \in N : b \succ_i a^*\}| \geqslant n/2\}$. Then, the distortion of our rule is:

$$\begin{aligned} \frac{\mathsf{sw}(a^*)}{\sum_{a\in A} p(a)\mathsf{sw}(a)} &\leqslant \frac{\mathsf{sw}(a^*)}{\sum_{a\in A'} p(a)\mathsf{sw}(a)} \\ &\leqslant \frac{\mathsf{sw}(a^*)}{(\min_{a\in A'} \mathsf{sw}(a)) \sum_{a\in A'} p(a)} \\ &\leqslant 2\frac{\mathsf{sw}(a^*)}{\min_{a\in A'} \mathsf{sw}(a)} \leqslant 4\frac{1-\gamma_{\min}}{\gamma_{\min}} + 2 = \frac{4}{\gamma_{\min}} - 2. \qquad \square \end{aligned}$$

Finally, applying Theorem 5 along with our reductions, we conclude that in the PB setting, *Maximal Lottery* has distortion at most $O(\log(m)/\gamma_{\min})$, as needed. □

Lower Bounds

Now, we lower bound the distortion achievable by *any* randomized aggregation rule in the PB setting.

Theorem 6. *For all randomized rules f,*

$$\mathrm{dist}_{\mathsf{rank}}(f) \in \Omega(\log(m)).$$

Proof. Define $k = \lceil \sqrt{m} \rceil - 1$ and partition the alternatives into $k+1$ buckets $A_1, \ldots, A_k, B$ such that for $\ell \in [k]$, A_ℓ consists of ℓ alternatives with cost $1/\ell$ each, and B includes the rest of the alternatives with cost 1 each. Note that each A_ℓ is a feasible subset.

Suppose that all the voters have the same ranking where they rank every alternative in A_ℓ higher than every alternative in $A_{\ell'}$ for all $\ell < \ell'$ (and breaks ties within each A_ℓ arbitrarily), and rank members of B at the end of their ranking.

Consider any aggregation rule. For each $a \in A$, let p_a denote the marginal probability of alternative a being included in the distribution returned by the rule on this profile. For each $\ell \in [k]$, define $\bar{p}_\ell = \frac{1}{\ell} \sum_{a \in A_\ell} p_a$ as the average of the marginal probabilities of alternatives in A_ℓ being chosen. Since the rule returns a distribution over budget-feasible subsets of alternatives (with total cost at most 1), the expected cost under this distribution is also at most 1. Due to additivity of cost and linearity of expectation, the expected cost can be written as

$$\sum_{a \in A} p_a \cdot c_a \geqslant \sum_{\ell \in [k]} \left(\frac{1}{\ell} \sum_{a \in A_\ell} p_a \right) = \sum_{\ell \in [k]} \bar{p}_\ell \leqslant 1. \tag{2}$$

Next, fix an arbitrary $t \in [k]$. Consider the following consistent utility function of the agent (which, in this case, is also her PS-value function): $v(a) = u(a) = 1$ if $a \in \cup_{\ell \in [t]} A_\ell$ and $v(a) = u(a) = 0$ otherwise. It is evident that the budget-feasible subset with the highest social welfare (i.e., one which contains the highest number of alternatives of value 1 to the agent) is A_t, and $\mathsf{sw}(A_t) = t$. In contrast, using the additivity of the utility function over the alternatives and linearity of expectation, we can write the expected social welfare under the rule as $\sum_{a \in \cup_{\ell \in [t]} A_\ell} p_a \cdot 1 = \sum_{\ell \in [t]} \ell \cdot \bar{p}_\ell$, which means the distortion is at least

$$D_t = \frac{t}{\sum_{\ell \in [t]} \ell \cdot \bar{p}_\ell}.$$

Because $t \in [k]$ was fixed arbitrarily, we get that the distortion is at least $D = \max_{t \in [k]} D_t$. Our goal is to show that $D = \Omega(\log m)$.

Note that for each $t \in [k]$, we have

$$\frac{t}{\sum_{\ell \in [t]} \ell \cdot \bar{p}_\ell} \leqslant D \Rightarrow \sum_{\ell \in [t]} \ell \cdot \bar{p}_\ell \geqslant \frac{t}{D}.$$

Dividing both sides by $t(t+1)$, we have that

$$\sum_{\ell \in [t]} \frac{\ell}{t(t+1)} \cdot \bar{p}_\ell \geqslant \frac{1}{D \cdot (t+1)}, \forall t \in [k].$$

Summing over $t \in [k]$, the right hand side sums to $(H_{k+1} - 1)/D$. On the left hand side, the coefficient of each $\bar{p}_\ell$ is

$$\ell \sum_{t=\ell}^{k} \frac{1}{t(t+1)} = \ell \left(\sum_{t=\ell}^{k} \frac{1}{t} - \frac{1}{t+1} \right) = \ell \left(\frac{1}{\ell} - \frac{1}{k+1} \right) \leqslant 1.$$

Hence, the left hand side sums to at most $\sum_{\ell \in [k]} \bar{p}_\ell \leqslant 1$. Since the left hand side is at least the right hand side, we have that

$$1 \geqslant \frac{H_{k+1} - 1}{D} \Rightarrow D \geqslant H_{k+1} - 1 = H_{\lceil \sqrt{m} \rceil} - 1,$$

which completes the proof after observing that

$$H_{\lceil \sqrt{m} \rceil} \geqslant \ln(\lceil \sqrt{m} \rceil) \geqslant \ln(\sqrt{m}) = \ln(m)/2. \qquad \square$$

Together, Theorems 4 and 6 imply that *Maximal Lottery* achieves optimal dependency on m and is within a $1/\gamma_{\min}$ factor of optimal overall. As before, we explore the source of this $1/\gamma_{\min}$ gap by showing that at least for large m, *Maximal Lottery* is the optimal randomized single-winner rule:

Theorem 7. *For all randomized single-winner rules f,*

$$\text{dist}^{\text{single-win}}_{\mathsf{rank}}(f) \in \Omega\left(\min\left\{m, 1/\gamma_{\min}\right\}\right).$$

This lower bound is proven in Appendix A.3 by the same construction as in Theorem 3, the analogous lower bound for the deterministic case. As with *Copeland* in the deterministic case, this bound shows that *Maximal Lottery* is optimal when $m \geqslant \Omega(1/\gamma_{\min})$. When $m \in o(1/\gamma_{\min})$, this bound is matched by the rule that chooses a uniformly random alternative. As before, whether a *single* $\gamma_{\min}$-oblivious randomized rule can match this lower bound remains open.

4 PB with Other Known Ballot Formats

In Section 3, we found something that may initially seem strange: in both the randomized and deterministic cases, voting rules designed for the single-winner setting—which output just a single alternative, even in PB—achieved optimal dependence on m. This is due to the weakness of the rank ballot format: designed for single-winner voting, rank fails to capture fundamental aspects of the PB setting.

Others have tried to address this problem in the unit-sum utilities model by designing better ballot formats. We now pursue the same approach in the public spirit model: focusing henceforth on deterministic rules for their practicality, we aim to identify a ballot format that achieves sublinear distortion in m, thereby surpassing our lower bound of $\Omega(m)$ in Theorem 2. To this end, we examine four other known PB ballot formats (see Benadè et al. [2021]): *rankings by value for money*, where each voter i ranks alternatives by $v_i(a)/c(a)$; *k-approvals*, where each voter i submits a set of k alternatives with the highest $v_i(a)$; *knapsack*, where each voter i submits the budget-feasible set with the highest value $v_i(S)$; and *threshold approvals*, where the ballot format specifies a threshold t and each voter i submits the set of alternatives with $v_i(a) \geqslant t$. Unfortunately, the answer is resoundingly negative for all these ballot formats.

The main goal of this section is to formally define the ballot formats mentioned in Section 4. In each subsection we focus on one of the ballot formats and prove bounds on the optimal achievable distortion with that ballot format using either deterministic or randomized rules. Furthermore, in Table 2 we show an overview of all the bounds that we prove in this paper, and compare them to the results of Benadè et al. [2021] under the unit-sum assumption.

Following the model of Benadè et al. [2021], a *ballot format* $\mathsf{X} : \mathbb{R}^m_{\geqslant 0} \times [0,1]^m \to \mathcal{L}_{\mathsf{X}}$ turns every PS-value function into a "vote", which takes values from a (usually finite) set $\mathcal{L}_{\mathsf{X}}$, sometimes using the cost function over the alternatives. Under this ballot format, each voter i submits the vote $\rho_i = \mathsf{X}(v_i)$; together, these votes form the *input profile* $\vec{\rho} = \{\rho_1, \ldots, \rho_n\}$. We use $V_{\vec{\gamma},U} \rhd_{\mathsf{X}} \vec{\rho}$ to indicate that PS-value matrix $V_{\vec{\gamma},U}$ induces input profile $\vec{\rho}$ under ballot format X. Alternatively, we say that $\vec{\rho}$ is consistent with $V_{\vec{\gamma},U}$. We omit X when it is clear from the context.

4.1 Rankings by Value for Money

In the ballot format *rankings by value for money* (vfm), $\mathcal{L}_{\mathsf{vfm}}$ is still the set of all rankings over alternatives, but now each voter i submits a ranking ρ_i of the alternatives by their PS-value divided by cost, i.e., such that for every $a, b \in A$, $v_i(a)/c(a) > v_i(b)/c(b)$ implies $a \succ_{\rho_i} b$; the voter can break ties arbitrarily.

Deterministic Rules

Benadè et al. [2021] show that no deterministic rule for rankings by value for money can achieve bounded distortion, even under the unit-sum assumption. Moreover, in their construction, all voters submit the same ranking. Adding any amount of public spirit would therefore leave the rankings and their analysis unchanged, implying that the distortion remains unbounded even with public spirit. We formalize this in Theorem 8.

Theorem 8 (lower bound)**.** *For rankings by value for money, every deterministic rule f has unbounded distortion:* $\text{dist}_{\mathsf{vfm}}(f) = \infty$.

Proof. We use the exact same construction used by Benadè et al. [2021]. Fix $a, b \in A$, and let $c_a = \epsilon > 0$ and $c_x = 1$ for all $x \in A \setminus \{a\}$. Construct an input profile $\vec{\rho}$ where each voter has alternatives a and b in positions 1 and 2, and let f be some deterministic aggregation rule.

If $f(\vec{\rho}, c) \neq a$, then construct a utility profile where $u_i(a) = 1$ and $u_i(x) = 0$ for all $x \in A \setminus \{a\}$. Then the distortion is infinite.

If $f(\vec{\rho}, c) = a$, then construct a utility profile where $u_i(a) = \epsilon$, $u_i(b) = 1$ and $u_i(x) = 0$ for $x \in A \setminus \{a, b\}$. Then,

$$\frac{v_i(a)}{c_a} = \frac{(1-\gamma_i)\epsilon + \gamma_i \frac{(n\epsilon)}{n}}{\epsilon} = \frac{(1-\gamma_i) + \gamma_i}{1} = \frac{v_i(b)}{c_b},$$

and so the ranking of each voter is consistent with this utility profile. But, the distortion is:

$$\frac{n}{n\epsilon} = \frac{1}{\epsilon},$$

which, as $\epsilon \to 0$, tends to infinity. □

Randomized Rules

For randomized rules, we show the same upper bound (up to a constant) for rankings by value for money as for rankings by value. The result uses a similar construction, too: First, we bucket alternatives as in Lemma 3, so that the alternatives in each bucket differ in cost by a factor of at most 2. Due to these similar costs, a ranking by value for money of the alternatives within any is a good approximation of their ranking by value, allowing us to apply our reductions from PB to committee selection to single-winner selection, except we lose an additional factor of 2.

Theorem 9 (upper bound)**.** *For rankings by value for money, there exists a randomized rule f with distortion*

$$\text{dist}_{\mathsf{vfm}}(f) \leqslant 8\left(\lceil \log_2(m) \rceil + 1\right)\left(2\gamma_{\min}^{-1} - 1\right).$$

Lemma 5. *For rankings by value for money, there exists a k-committee-selection voting rule f such that on all sets of alternatives with costs in $[2^{-\ell}, 2^{1-\ell}]$ for some $\ell \geqslant 0$, f has distortion $4(2\gamma_{\min}^{-1} - 1)$.*

Proof. Notice that if a beats b, then $v_i(a)/c_a \geqslant v_i(b)/c_b$ at least $n/2$ times. Since the costs differ by at most a factor of 2, $2v_i(a) \geqslant v_i(b)$.

We can use the exact same rule as in Theorem 5. Indeed, everything is the same, except that when b beats a^* in a pairwise election (i.e. at least $n/2$ times), we get the following distortion by Lemma 1:

$$\frac{\mathsf{sw}(a^*)}{\mathsf{sw}(b)} \leqslant 2\left(2\frac{1-\gamma_{\min}}{\gamma_{\min}} + 1\right).$$

Then, the distortion of our rule is, by the same analysis in Theorem 5:

$$8\frac{1-\gamma_{\min}}{\gamma_{\min}} + 4.$$

From here, we can convert this single-winner rule into a committee selection rule with the same distortion by using Lemma 4. □

Having proved this lemma, we utilise an argument similar to Lemma 3.

Proof of Theorem 9. Let g be the rule in Lemma 5, and let the distortion it achieves, $\left(4\frac{1-\gamma_{\min}}{\gamma_{\min}}+2\right)$, be d. By the same mechanism in Lemma 3, we will convert g to a ranking by value per cost rule.

Indeed, divide the alternatives into buckets $A_0, A_1, \ldots, A_{\lceil \log_2(m) \rceil}$, where for $i \neq 0$:

$$A_i = \left\{ a \in A : \frac{2^{i-1}}{m} < c_a \leqslant \frac{2^i}{m} \right\},$$

and

$$A_0 = \{a \in A : c_a \leqslant 1/m\}.$$

Recall the mechanism used:

1. Pick the bucket A_j uniformly at random.
2. Consider the restricted election with only the alternatives in A_j.
3. Use g to pick the top $\lfloor \frac{m}{2^j} \rfloor$ alternatives in the restricted election.

Consider any PB instance. Split the alternatives into buckets $A_0, A_1, \ldots, A_{\lceil \log_2(m) \rceil}$, where for $i \neq 0$:

$$A_i = \left\{ a \in A : 2^{i-1}/m < c_a \leqslant 2^i/m \right\},$$

and

$$A_0 = \{a \in A : c_a \leqslant 1/m\}.$$

The randomized PB rule f is as follows:

1. Pick $j \in \{0, 1, \ldots, \lceil \log_2(m) \rceil\}$ uniformly at random.
2. Consider the restricted instance with only the alternatives in A_j.
3. With $m' = |A_j|$ and $k = \min(m', \lfloor \frac{m}{2^j} \rfloor)$, use the k-committee selection rule $f_{m',k}$ on this restricted instance to pick a set of k alternatives and return it.

Let A^* be the optimal budget-feasible subset of the alternatives, L_j^* be the optimal $\lfloor \frac{m}{2^j} \rfloor$-committee of A_j, and L_j be the one selected by the k-committee rule. For $j \neq 0$, $A^* \cap A_j$ is of size at most $\frac{m}{2^{j-1}}$. That means $\mathsf{sw}(A^* \cap A_j) \leqslant 2\mathsf{sw}(L_j^*)$ for any $j \neq 0$.

In addition for $j = 0$, $L_0^* = A_0$ which implies $\mathsf{sw}(A^* \cap A_j) \leqslant \mathsf{sw}(L_j^*)$. Since the k-committee selection rule has distortion of d for any j we have $\mathsf{sw}(L_j^*) \leqslant d\mathsf{sw}(L_j)$ which gives us $\mathsf{sw}(A^* \cap A_j) \leqslant 2d\mathsf{sw}(L_j)$. Let δ be the distribution of the output of the mechanism, we have:

$$\begin{aligned}
\mathbb{E}_{L\sim\delta}[\mathsf{sw}(L)] &= \frac{1}{\lceil \log_2(m) \rceil + 1} \sum_{j=0}^{\lceil \log_2(m) \rceil} \mathsf{sw}(L_j) \\
&\geqslant \frac{1}{\lceil \log_2(m) \rceil + 1} \sum_{j=0}^{\lceil \log_2(m) \rceil} \frac{\mathsf{sw}(A^* \cap A_j)}{2d} \\
&\geqslant \frac{\mathsf{sw}(A^*)}{2d(\lceil \log_2(m) \rceil + 1)},
\end{aligned}$$

which gives us the desired distortion bound. □

Whether this is (asymptotically) the best distortion that randomized rules for rankings by value for money can achieve remains an open question.

4.2 k-approval ballots

Approval-Based Ballots Another popular type of ballot—especially in participatory budgeting—is to ask voters to simply *approve* their favorite items, rather than rank items relative to one another. The most common type of approval-based ballots in practice is the *k-approval ballot*, in which voters "vote" by identifying their k favorite alternatives. However, this ballot format has an important limitation in the PB context: as we show, it allows voters to approve items or sets of items that are *not budget-feasible*. In the worst case, this can leave the voting rule with little or no information about which *budget-feasible* allocations are desirable, in which case it can do nothing better than making an arbitrary choice.

A natural potential fix for this is allowing voters to approve *only sets of items that are budget-feasible*. This is can be achieved by either restricting our use to 1-approval ballots (and removing all items which individually exceed the budget), or using *Knapsack ballots*, an approval-based ballot format in which voters can approve any set of projects whose total cost does not exceed the budget. We explore both these directions in this and the next subsections.

For the ballot format k-approval (k-app), the set of possible ballots $\mathcal{L}_{\text{k-app}}$ is the set of all subsets of size k of A. That means each voter submits the set of her top k alternatives (breaking the ties arbitrarily). We start by showing that asking voters to approve more than one alternative leads to an unbounded distortion.

Theorem 10 (LB - Deterministic)**.** *For k-approval ballot format with $k \geqslant 2$, any deterministic PB rule has unbounded distortion.*

Proof. Suppose we are using k-approval ballots. Let A be the alternatives, and suppose that each $a \in A$ has cost $\frac{1}{k-1}$. Suppose all agents have the same utilities, where $\epsilon > 0$ is arbitrarily small, giving 1 utility to a_1, ϵ utility for all of $a_2 \ldots a_k$, and 0 for all $A \setminus \{a_1, \ldots, a_k\}$. Then, everyone's public-spirited values are identical to their utilities. All agents approve $a_1, \ldots, a_k$, and the deterministic rule must pick $k-1$ of these arbitrarily. Let the deterministic rule pick $a_2 \ldots a_k$. The best possible welfare is n, achieved by any $k-1$-subset including a_1; the winner has welfare ϵn, making the distortion $\frac{1}{\epsilon}$ (unbounded). □

These lower bounds were for $k \geqslant 2$; one can also realize the same bounds with $k = 1$, where all voters approve items whose costs exceed 1, giving the voting rule no information about which budget-feasible set to choose. However, an obvious fix for this is to remove all items ahead of time that exceed the budget. If we assume every *individual* item has cost at most 1, then 1-approval ballots ensure that voters can only approve budget-feasible sets, escaping the problem described above. Then, 1-approval-based ballots are akin to plurality voting, and they permit the following positive result:

Proposition 1 (UB, 1-app, Deterministic)**.** *If all alternatives have cost at most* 1*, then for* 1*-approval ballot format, there exists a deterministic voting rule f with distortion*

$$\text{dist}_{\textit{1-app}}(f) \in \mathcal{O}\left(\frac{m^2}{\gamma_{\min}}\right).$$

Proof. Pick the most approved alternative a. This is in fact the plurality winner and by Theorem 1, the plurality rule achieves the claimed distortion. □

The following proposition shows that this is the best we can hope for.

Proposition 2 (LB, 1-app, Deterministic)**.** *For* 1*-approval ballot format, every deterministic rule f has distortion*

$$\text{dist}_{\textit{1-app}}(f) \in \Omega\left(\frac{m^2}{\gamma_{\min}}\right).$$

Proof. We take m to be sufficiently large. Consider an instance with $\frac{m}{2}$ alternatives $a_1, \ldots, a_{m/2}$ of cost 1 and $\frac{m}{2}$ alternatives $b_1, \ldots, b_{m/2}$ of cost $\frac{2}{m}$, and all the voters have the same PS-value of $\gamma = \gamma_{\min}$. Suppose $\frac{2n}{m}$ voters vote for each a_i.

If a PB rule picks the bundle $b_1, \dots, b_{m/2}$, then consider the instance where every voter assigns a value of 1 to each a_i and a value of 0 to each b_i. This is consistent with the input, and results in infinite distortion.

Instead, suppose the PB rule, without the loss of generality, picks $a_{m/2}$. Then, suppose that every voter who votes for $a_{m/2}$ gives it a value of $\gamma \frac{m-2}{m-2\gamma}$, and everything else a value of 0, and suppose that all other voters give their top choice a value of 1, the b_i a value of $\frac{m-\gamma(m-2)}{m-2\gamma}$, and everything else a value of zero.

Then, $\mathsf{sw}(b_i) = \frac{m-\gamma(m-2)}{m-2\gamma} \cdot \frac{m-2}{m} \cdot n$ for all i from 1 to $\frac{m}{2}$, and $\mathsf{sw}(a_i) = \frac{2n}{m}$ for $i \neq \frac{m}{2}$ with $\mathsf{sw}(a_{m/2}) = \frac{2n}{m} \cdot \gamma \frac{m-2}{m-2\gamma}$.

Then, the utilities for voters i who vote for $a_{m/2}$ are consistent as

$$\begin{aligned}
v_i(a_{m/2}) &= (1-\gamma)\frac{\gamma(m-2)}{m-2\gamma} + \gamma \frac{m-2}{m-2\gamma}\frac{2}{m} \\
&= \frac{\gamma(m-2)}{m-2\gamma}\left(1-\gamma+\frac{2}{m}\right) \\
&= \frac{\gamma(m-2)}{m-2\gamma}\frac{m-\gamma m+2}{m} \\
&\geqslant \gamma \frac{m-\gamma(m-2)}{m-2\gamma}\frac{m-2}{m} = v_i(b_j)
\end{aligned}$$

for all b_j, where the last inequality holds because $2 \geqslant 2\gamma$. Similarly,

$$\begin{aligned}
v_i(a_{m/2}) &= (1-\gamma)\frac{m-2}{m-2\gamma} + \gamma \frac{m-2}{m-2\gamma}\frac{2}{m} \\
&= \frac{m-2}{m-2\gamma}\frac{m-\gamma(m-2)}{m} \\
&\geqslant \gamma \frac{2}{m} = v_i(a_j)
\end{aligned}$$

for all $a_j \neq a_{m/2}$, where the last inequality holds for sufficiently large m, so $a_{m/2}$ is indeed the alternative of highest value.

The utilities of voters i who vote for $a_j \neq a_{m/2}$ is consistent, as

$$\begin{aligned}
v_i(b_i) &= (1-\gamma)\frac{m-\gamma(m-2)}{m-2\gamma} + \gamma \frac{m-\gamma(m-2)}{m-2\gamma} \cdot \frac{m-2}{m} \\
&= \frac{m-\gamma(m-2)}{m-2\gamma}\left(1-\gamma+\gamma\frac{m-2}{m}\right) \\
&= \frac{m-\gamma(m-2)}{m} \\
&= (1-\gamma) + \gamma \cdot \frac{2}{m} = v_i(a_j)
\end{aligned}$$

for all b_i. And $v_i(a_j) \geqslant v_i(a_k)$ for all $k \neq j$ as $\mathsf{sw}(a_k) \leqslant \mathsf{sw}(a_j)$ and voter i gives a_k zero utility. So, a_j is indeed the highest ranking alternative.

But, the distortion we get is

$$\begin{aligned}
\frac{\sum_i \mathsf{sw}(b_i)}{\mathsf{sw}(a_{m/2})} &= \frac{m}{2} \cdot \frac{m-\gamma(m-2)}{m-2\gamma} \cdot n \cdot \left(\frac{2n}{m} \cdot \gamma \frac{m-2}{m-2\gamma}\right)^{-1} \\
&= \frac{m^2}{4} \cdot \frac{m-\gamma(m-2)}{\gamma(m-2)} \\
&= \frac{m^2}{4} \cdot \left(\frac{1}{\gamma} \cdot \frac{m}{m-2} - 1\right) \\
&\geqslant \frac{m^2}{4} \cdot \frac{1-\gamma}{\gamma},
\end{aligned}$$

as claimed. □

Remark 1. While not explicitly studied in Benadè et al. [2021], a deterministic distortion of $\Theta(m^2)$ in the 1-approval ballot format follows from their analysis of the ranking by value ballot format immediately, as it simply uses a plurality rule to aggregate voter preferences.

While 1-approval ballot sounds practical, it does not yield a good distortion since the basic potential of PB (which is selecting multiple alternatives if the budget allows) is not used. However, this is really the best we can hope for with k-approval ballots. This motivates the consideration of knapsack ballots, which elicits the top budget-feasible subset from each voter's perspective.

4.3 Knapsack ballots

For the ballot format *knapsack* (knap), the set of possible ballots $\mathcal{L}_{\mathsf{knap}} = \mathcal{F}$ is the set of all budget-feasible subsets of A. Each voter i submits the subset she values most: $\rho_i \in \arg\max_{S \in \mathcal{F}} v_i(S)$. This amounts to asking each voter to solve her own personal knapsack problem.

Unfortunately, similar to what happens with 1-app ballots, an instance similar to the one in Proposition 2 also applies to knapsack ballots, since voters are only permitted to approve budget-feasible allocations, which all consist of one single item.

Corollary 1 (LB, knap, Deterministic)**.** *For knapsack ballot format, every deterministic rule f has distortion*

$$\mathrm{dist}_{knap}(f) \geqslant m\gamma_{\min}^{-1} - m + 1 \in \Omega\left(\frac{m}{\gamma_{\min}}\right).$$

For randomized rules, we prove a slightly weaker lower bound that is $\gamma_{\min}$ times our lower bound for deterministic rules. As $\gamma_{\min}$ goes from 0 to 1, the lower bound for deterministic rules goes from unbounded to 1 while that for randomized rules goes from m to 1. It is easy to observe that both lower bounds are tight at both extremes, but there may be room for improvement for intermediate values of $\gamma_{\min}$. The proof is in Theorem 11.

Theorem 11 (LB, knap, Randomized)**.** *For knapsack ballot format, every randomized rules f has distortion*

$$\mathrm{dist}_{knap}(f) \geqslant m(1-\gamma_{\min}) + \gamma_{\min}.$$

Proof. Formally, consider a case where n is divisible by m, all the voters have the same PS-value of $\gamma = \gamma_{\min}$, and every alternative $a \in A$ has a cost of $c_a = 1$. In this case, each vote consists of exactly one alternative. For any alternative $a \in A$, let N_a be the set of voters who vote for alternative a. Choose the input profile $\vec{\rho}$ so that n/m voters vote for each alternative so that $|N_a| = \frac{n}{m}$ for all $a \in A$. Our randomized voting rule f must pick some alternative a^* with probability at most $1/m$.

Suppose that all voters in N_{a^*} have a utility of $\frac{m(1-\gamma)+\gamma}{\gamma}$ for a^* and utility zero for everything else. Moreover, voters in N_a with $a \neq a^*$ have utility 1 for a and zero utility for the rest of the alternatives. We can see that the social welfare of a^* is $\frac{m(1-\gamma)+\gamma}{\gamma} \cdot \frac{n}{m}$, and the social welfare of any other alternative is $\frac{n}{m}$.

First of all, we have to make sure that this utility matrix and PS-vector yield a value matrix consistent with the input profile. For any $a \neq a^*$ and $i \in N_a$ we have:

$$\begin{aligned} v_i(a^*) &= \gamma \frac{m(1-\gamma)+\gamma}{\gamma} \cdot \frac{1}{m} \\ &= \frac{m(1-\gamma)+\gamma}{m} = (1-\gamma) + \frac{\gamma}{m} \\ &= v_i(a). \end{aligned}$$

Furthermore, for voter $i \in N_{a^*}$ and any $a \neq a*$ as:

$$\begin{aligned} v_i(a^*) &= (1-\gamma)\frac{m(1-\gamma)+\gamma}{\gamma} + \gamma\frac{m(1-\gamma)+\gamma}{\gamma}\cdot\frac{1}{m} \\ &= \left(1-\gamma\frac{m-1}{m}\right)\frac{m(1-\gamma)+\gamma}{\gamma} \\ &= \frac{m-\gamma(m-1)}{m}\cdot\frac{m(m-\gamma)+\gamma}{\gamma} \\ &= \frac{\gamma}{m}\cdot\frac{(1-\gamma)m+\gamma}{\gamma}\cdot\frac{m(m-\gamma)+\gamma}{\gamma} \\ &\geqslant \frac{\gamma}{m} = v_i(a), \end{aligned}$$

where the last inequality follows from the fact that $\gamma \leqslant 1$. That means the value matrix is consistent with the input profile for all the voters.

After that, we can see the distortion that the rule incurs. We could have gotten a utility of $\frac{n}{m}\cdot\frac{m(1-\gamma)+\gamma}{\gamma}$ by choosing a^*, but instead, we got the expected utility of the following

$$\begin{aligned} \mathbb{E}_{a\sim f(\vec{\rho},c)}[\mathsf{sw}(a)] &\leqslant \frac{1}{m}\mathsf{sw}(a^*) + \frac{m-1}{m}\cdot\frac{n}{m} \\ &= \frac{1}{m}\cdot\frac{n}{m}\cdot\frac{m(1-\gamma)+\gamma}{\gamma} + \frac{m-1}{m}\cdot\frac{n}{m} \\ &= n\left(\frac{m(1-\gamma)+\gamma+(m-1)\gamma}{m^2\gamma}\right) \\ &= \frac{n}{\gamma m}, \end{aligned}$$

and so the distortion is at least:

$$\begin{aligned} \mathrm{dist}_{\mathsf{knap}}(f,\vec{\rho},c) &= \frac{\mathsf{sw}(a^*)}{\mathbb{E}_{a\sim f(\vec{\rho},c)}[\mathsf{sw}(a)]} \\ &\geqslant \frac{\frac{n}{m}\cdot\frac{m(1-\gamma_{\min})+\gamma_{\min}}{\gamma_{\min}}}{\frac{n}{\gamma_{\min}m}} \\ &= m(1-\gamma_{\min})+\gamma_{\min}. \end{aligned}$$

□

This lower bound is trivially tight in m. We show this by having m alternatives of cost 1 each, and $\frac{n}{m}$ voters approving each one.

Remark 2 (UB, knap, Randomized)**.** The voting rule f which ignores all the ballots and simply picks a single alternative uniformly at random trivially yields an upper bound of $\mathrm{dist}_{\mathsf{knap}}(f) \leqslant m$.

Finally, we present upper bounds for knapsack due to its importance in the literature. In the unit-sum model, Benadè et al. [2021] give exponential lower bounds for the knapsack ballot format. We are able to prove that in the public-spirit model, it is possible to break this exponential barrier, showing that the worst-case instances for knapsack in the unit-sum model rely on potentially infeasible voter preferences. In doing so, we rely on new techniques for aggregating knapsack votes. This illustrates how public spirit can be much more powerful than that pervasive assumption (which is hard to justify) in mitigating distortion, especially when the number of alternatives is at all large.

Theorem 12 (UB, knap, Deterministic)**.** *For knapsack votes, there exists a deterministic rule f with distortion*

$$\mathrm{dist}_{\mathsf{knap}}(f) \leqslant 4m^3(\gamma_{\min}^{-2}-\gamma_{\min}^{-1})+3m \in O\left(\frac{m^3}{\gamma_{\min}^2}\right).$$

Proof. For any subset of alternatives $S \subseteq A$, let $n_S := \sum_{i \in N} \mathbb{I}(S \subseteq \rho_i)$ be the number of voters whose knapsack set contains S. We use shorthand $n_a := n_{\{a\}}$ and $n_{a,b} := n_{\{a,b\}}$ for all $a, b \in A$. Then, informally, $n_{a,b}$ is the number of voters who vote for both a and b.

For an arbitrary input, define $A_0 := \{a \in A \colon n_a \geqslant \frac{n}{2m}\}$ and initialize $A^- = A_0$ and $A^+ = \emptyset$. We will return A^+ after running the following until A^- is empty:

1. Remove the alternative b with the highest cost in A^- and add it to A^+.
2. Remove from A^- all alternatives a such that

$$\frac{n_{a,b}}{n_b} \leqslant \frac{m-1}{m}.$$

First, we will prove that this algorithm always returns a budget-feasible subset. Suppose for the sake of contradiction that at some point, the max-cost item in A^-, call it a^{m}, is no longer within budget: i.e., $c_{a^{\mathsf{m}}} + \sum_{b \in A^+} c_b > 1$. We will show that there exists some $b \in A^+$ such that $\frac{n_{b,a^{\mathsf{m}}}}{n_b} \leqslant \frac{m-1}{m}$.

Let $b^{\mathsf{m}} \in A^+$ be the first alternative added to A^+, so that it has maximum cost. Then, for all $b \in A^+ \setminus \{b^{\mathsf{m}}\}$, because b wasn't pruned in step 2 directly after adding b^{m}, it must be that $\frac{n_{b,b^{\mathsf{m}}}}{n_{b^{\mathsf{m}}}} > \frac{m-1}{m}$. By the same reasoning, the same must be true for a^{m} —that is, $\frac{n_{a^{\mathsf{m}},b^{\mathsf{m}}}}{n_{b^{\mathsf{m}}}} > \frac{m-1}{m}$. Summing over these inequalities, we get that:

$$n_{a^{\mathsf{m}},b^{\mathsf{m}}} + \sum_{b \in A^+ \setminus \{b^{\mathsf{m}}\}} n_{b^{\mathsf{m}},b} > n_{b^{\mathsf{m}}} \left[\frac{m-1}{m} + \frac{m-1}{m}\left(\left|A^+\right| - 1\right)\right] = n_{b^{\mathsf{m}}} \frac{m-1}{m} \left|A^+\right|.$$

Notice that the left hand side is at most the number of voters who voted for b^{m}, multiplied by the number of other alternatives in $\{a^{\mathsf{m}}\} \cup |A^+|$ they could have voted for. Since $\{a^{\mathsf{m}}\} \cup A^+$ is an infeasible set, no voter could have voted for all of them. Thus, each voter can only vote for $|A^+|$ alternatives in $\{a^{\mathsf{m}}\} \cup |A^+|$, and so only $|A^+| - 1$ alternatives other than b^{m}. The left hand side is then at most $(|A^+| - 1)n_{b^{\mathsf{m}}}$, and therefore

$$(\left|A^+\right| - 1)n_{b^{\mathsf{m}}} > n_{b^{\mathsf{m}}} \frac{m-1}{m} \left|A^+\right|.$$

Simplifying, we can see that this is impossible, as this is equivalent to the inequality:

$$\left|A^+\right| - 1 > \left|A^+\right| - \left|A^+\right|/m.$$

We have encountered a contradiction, so our premise — that we added an a to A^+ that exceeded the budget — must have been false.

Now, we will show that if an $a \in A^-$ is pruned in Step 2, then $\frac{\mathsf{sw}(a)}{\mathsf{sw}(A^+)} \leqslant 2m^2 \frac{1-\gamma_{\min}}{\gamma_{\min}} + 1$. Indeed, because we prune it, there exists some $b \in A^+$ such that:

$$\frac{n_{a,b}}{n_b} \leqslant \frac{m-1}{m}.$$

Since $b \in A_0$, we have $n_b \geqslant n/2m$ and so $n_b - n_{a,b}$, the number of voters that vote for b but not a, is at least $n/(2m^2)$:

$$n_b - n_{a,b} \geqslant n_b - \frac{m-1}{m} n_b \geqslant \frac{n}{2m^2}.$$

Notice that because we pick the highest cost alternative b in each iteration, any alternative pruned later by the algorithm must have a cost lower than c_b. Therefore, any time a voter votes for b but not a, they could have replaced b with a and have gotten another feasible set. The fact that they did not means that they prefer b to a. We have at least $n/(2m^2)$ of such voters (that prefer b to a), by Lemma 1 we can conclude that $\frac{\mathsf{sw}(a)}{\mathsf{sw}(A^+)} \leqslant 2m^2 \frac{1-\gamma_{\min}}{\gamma_{\min}} + 1$, as needed.

Extending this result, define $m_0 := |A_0|$, we get that

$$\frac{\mathsf{sw}(A_0)}{\mathsf{sw}(A^+)} \leqslant m_0 \left(2m^2 \frac{1-\gamma_{\min}}{\gamma_{\min}} + 1\right).$$

On the other hand, for alternatives outside of A_0, the distortion must be small. Let A^* be the optimal budget-feasible set of alternatives. Then:

$$\frac{\mathsf{sw}(A^* \setminus A_0)}{\mathsf{sw}(A^+)} = \frac{\mathsf{sw}(A^* \setminus A_0)}{\mathsf{sw}(A_0)} \cdot \frac{\mathsf{sw}(A_0)}{\mathsf{sw}(A^+)}.$$

It remains to bound $\frac{\mathsf{sw}(A^* \setminus A_0)}{\mathsf{sw}(A_0)}$. Because at most $n/(2m)$ voters include each alternative in $A \setminus A_0$ in their knapsack set, and there are at most $m - m_0$ such alternatives, we know that at most $n(m - m_0)/2m$ voters vote for alternatives in $A \setminus A_0$, that is at least $n(m+m_0)/2m$ voters only vote for alternatives in A_0. Observing that $A^* \setminus A_0 \in \mathcal{F}$ (since $A^* \in \mathcal{F}$), it must be that for all $n(m+m_0)/2m$ voters i who vote for only alternatives in A_0, $v_i(A_0) \geqslant v_i(\rho_i) \geqslant v_i(A^* \setminus A_0)$ for each $a \in A \setminus A_0$. Therefore, by Lemma 1,

$$\frac{\mathsf{sw}(A^* \setminus A_0)}{\mathsf{sw}(A_0)} \leqslant \frac{2m}{m + m_0} \cdot \frac{1 - \gamma_{\min}}{\gamma_{\min}} + 1.$$

Thus,

$$\begin{aligned}
\frac{\mathsf{sw}(A^*)}{\mathsf{sw}(A^+)} &\leqslant \frac{\mathsf{sw}(A_0)}{\mathsf{sw}(A^+)} + \frac{\mathsf{sw}(A^* \setminus A_0)}{\mathsf{sw}(A^+)} = \frac{\mathsf{sw}(A_0)}{\mathsf{sw}(A^+)} + \frac{\mathsf{sw}(A^* \setminus A_0)}{\mathsf{sw}(A_0)} \cdot \frac{\mathsf{sw}(A_0)}{\mathsf{sw}(A^+)} \\
&\leqslant \frac{\mathsf{sw}(A_0)}{\mathsf{sw}(A^+)} \left(1 + \frac{m}{m_0} \cdot \frac{1 - \gamma_{\min}}{\gamma_{\min}} + 1\right) \\
&\leqslant m_0 \left(2m^2 \frac{1 - \gamma_{\min}}{\gamma_{\min}} + 1\right) \left(\frac{m}{m_0} \cdot \frac{1 - \gamma_{\min}}{\gamma_{\min}} + 2\right) \\
&\leqslant 2m^3 \left(\frac{1 - \gamma_{\min}}{\gamma_{\min}}\right)^2 + 4m^3 \frac{1 - \gamma_{\min}}{\gamma_{\min}} + m \frac{1 - \gamma_{\min}}{\gamma_{\min}} + 2m \\
&\leqslant 4m^3 \left(\gamma_{\min}^{-2} - \gamma_{\min}^{-1}\right) + 3m.
\end{aligned}$$

□

It's possible that for general Knapsack voting, this cannot be improved to match the lower bound that is achieved in the case that reduces to plurality voting. This is because in the general case where people can approve more than 1 alternative, although we have *budget-feasible information*, we don't know what people's *favorite* element is in their approval set if it is greater than size 1.

Knapsack for Committee Selection

We can improve the analysis of the knapsack voting when all alternatives have the same cost.

Theorem 13. *We can get a distortion of* $1 + \frac{m}{2} + \frac{1-\gamma_{\min}}{\gamma_{\min}} m^2$ *in the deterministic knapsack setting for* $m/2$*-multiwinner elections (or equivalently when* $c_a = \frac{2}{m}$ *for all* $a \in A$*).*

Proof. Recall the notation used in the proof of Theorem 12. For any subset of alternatives $S \subseteq A$, let $n_S := \sum_{i \in N} \mathbb{I}(S \subseteq \rho_i)$ be the number of voters whose knapsack set contains S. We use shorthand $n_a := n_{\{a\}}$ and $n_{a,b} := n_{\{a,b\}}$ for all $a, b \in A$. Then, informally, $n_{a,b}$ is the number of voters who vote for both a and b.

The voting rule we will use is as follows: assign a plurality score to each alternative, where the score is simply the number of times each alternative appears.

Pick the $m/2$ alternatives with the largest plurality score, A. Indeed, every alternative can appear at most n times, as every voter can vote for them only once. Therefore, in the worst case, if the top $m/2 - 1$ alternatives appear n times there must remain $nm/2 - n(m/2 - 1) = n$ appearances of other alternatives. By the pigeonhole principle from here, the remaining plurality winner must be chosen $n/(m/2 + 1) > n/m$ times. Thus, the minimum number of times a plurality winner can appear is n/m.

Moreover, because $n_a > n_b$ for all $a \in A$ and $b \notin A$, and $\sum_{a\in A} n_a + \sum_{b\notin A} n_b = mn/2$, we get that $2\sum_{a\in A} n_a \geqslant mn/2$ and so $\sum_{a\in A} n_a \geqslant mn/4$.

Then, let A^* be the optimal set of alternatives. Note then that:

$$\begin{aligned}
\frac{\mathsf{sw}(A^*, U)}{\mathsf{sw}(A, U)} &= \frac{\sum_{a^*\in A^*} \mathsf{sw}(a^*, U)}{\sum_{a\in A} \mathsf{sw}(a, U)} \\
&= \frac{\sum_{a^*\in A^*\cap A} \mathsf{sw}(a^*, U)}{\sum_{a\in A} \mathsf{sw}(a, U)} + \frac{\sum_{a^*\in A^*\setminus A} \mathsf{sw}(a^*, U)}{\sum_{a\in A} \mathsf{sw}(a, U)} \\
&\leqslant 1 + \sum_{a^*\in A^*\setminus A} \frac{\mathsf{sw}(a^*, U)}{\sum_{a\in A} \mathsf{sw}(a, U)}. \qquad (3)
\end{aligned}$$

We will show that for all $a^* \in A^* \setminus A$, there exists some $a \in A$ such that:

$$\frac{\mathsf{sw}(a^*)}{\mathsf{sw}(a)} \leqslant 2\frac{1-\gamma_{\min}}{\gamma_{\min}} m + 1,$$

by considering two cases:

1. Suppose that for all $a^* \in A^* \setminus A$, there exists some $a \in A$ such that $n_{a,a^*}/n_a \leqslant 1/2$. Then, $n_a - n_{a,a^*} \geqslant n_a/2 \geqslant n/2m$. Therefore, by Lemma 1:

$$\frac{\mathsf{sw}(a^*)}{\mathsf{sw}(a)} \leqslant 2\frac{1-\gamma_{\min}}{\gamma_{\min}} m + 1.$$

2. Suppose that for some $a^* \in A^* \setminus A$, and for all $a \in A$, $n_{a,a^*}/n_a > 1/2$. Let $a_{\max} = \arg\max_{a\in A} n_a$ and $a_{\min} = \arg\min_{a\in A} n_a$. Then, in particular,

$$n_{a_{\max}} < 2n_{a_{\max},a^*} \leqslant 2n_{a^*} \leqslant 2n_{a_{\min}},$$

where the last equality holds because $a_{\min}$ is a plurality winner, and a^* isn't

Since $(m/2)n_{a_{\max}} \geqslant \sum_{a\in A} n_a \geqslant nm/4$, $n_{a_{\max}} \geqslant n/2$ and so $n_{a_{\min}} \geqslant n/4$. Therefore, we can improve the lower bound for plurality winners: for all $a \in A$, $n_a \geqslant n/4$.

By Lemma 6 below, we know that for all $a^* \in A^* \setminus A$, there exists some $a \in A$ such that $n_{a,a^*}/n_a \leqslant (m-2)/m$. Therefore, $n_a - n_{a,a^*} \geqslant 2n_a/m \geqslant n/2m$. Thus, by Lemma 1 in [Flanigan et al., 2023]:

$$\frac{\mathsf{sw}(a^*)}{\mathsf{sw}(a)} \leqslant 2\frac{1-\gamma_{\min}}{\gamma_{\min}} m + 1.$$

From here we can prove an m^2 bound easily by taking $a^*_{\max} = \text{argmax}_{a^*\in A^*} \mathsf{sw}(a^*, U)$. Then, continuing off of (3), and using the fact that there exists some $\hat{a} \in A$ such that $\frac{\mathsf{sw}(a^*_{\max}, U)}{\mathsf{sw}(\hat{a}, U)} \leqslant 2\frac{1-\gamma_{\min}}{\gamma_{\min}} m + 1$:

$$\begin{aligned}
\frac{\mathsf{sw}(A^*, U)}{\mathsf{sw}(A, U)} &\leqslant 1 + \frac{m}{2} \cdot \frac{\mathsf{sw}(a^*_{\max}, U)}{\sum_{a\in A} \mathsf{sw}(a, U)} \\
&\leqslant 1 + \frac{m}{2} \cdot \frac{\mathsf{sw}(a^*_{\max}, U)}{\mathsf{sw}(\hat{a}, U)} \\
&\leqslant 1 + \frac{1-\gamma_{\min}}{\gamma_{\min}} m^2 + \frac{m}{2},
\end{aligned}$$

as claimed. □

Lemma 6. *When A^* is the optimal subset and A is the subset chosen by the repeated plurality rule, for all $a^* \in A^* \setminus A$, there exists some $a \in A$ such that:*

$$\frac{N(a, a^*)}{N(a)} \leqslant (m-2)/m.$$

Proof. Note that $\sum_{a \in A} N(a, a^*)$ is the number of times a voter votes for some alternative and a^*. Each voter can vote for at most $m/2$ alternatives. Since there are then at most $m/2 - 1$ alternatives in A that any voter who votes for a^* could have voted for:

$$\sum_{a \in A} N(a, a^*) \leqslant N(a^*)(m/2 - 1) \leqslant N(a^*) \cdot \frac{m-2}{2}.$$

From here, let $a_{\min} = \text{argmin}_{a \in A} N(a, a^*)$. Then, substituting this into the inequality above, and using that $|A| = \frac{m}{2}$:

$$\frac{m}{2} N(a_{\min}, a^*) \leqslant N(a^*) \cdot \frac{m-2}{2}.$$

Since $N(a^*) \leqslant N(a_{\min})$ as a^* is not in A and therefore must occur at most as many times as any plurality winner,

$$\frac{m}{2} N(a_{\min}, a^*) \leqslant N(a_{\min}) \cdot \frac{m-2}{2},$$

and so finally

$$\frac{N(a_{\min}, a^*)}{N(a_{\min})} \leqslant \frac{m-2}{m},$$

as desired. ☐

4.4 Threshold Approval Votes

Finally, we investigate the distortion under the ballot format of *threshold approval votes*. Under this ballot format with threshold $\tau > 0$ (τ-th), each voter i reports the subset of alternatives for which her PS-value is at least a τ fraction of her total PS-value for all alternatives in A, i.e., $\rho_i = \{a \in A : v_i(a) \geqslant \tau \cdot \sum_{b \in A} v_i(b)\}$. Thus, $\mathcal{L}_{\tau\text{-th}} = 2^A$, as with knapsack votes. Benadè et al. [2021] introduce this ballot format for unit-sum utilities and our definition extends it to arbitrary utilities.[3]

It is easy to see that without a unit sum assumption, the distortion of any deterministic rule is unbounded, even with public-spirited voters.

Proposition 3. *The distortion associated with deterministic fixed thresholds (using the same definition as in [Benadè et al., 2021]) is unbounded for any choice of threshold.*

Proof. Suppose we use a threshold of t. Then, consider an input profile where no voter approves any alternative. Suppose that f picks $a^* \in A$. Then, consider a preference profile where $u_i(a^*) = 0$ and $u_i(b) = t/2$ for all $i \in N$ and all $b \neq a^*$.

Then, $v_i(a^*) = (1 - \gamma_i) \cdot 0 + \gamma_i \cdot \frac{0}{n} = 0 < t$ and $v_i(b) = (1 - \gamma_i) \cdot t/2 + \gamma_i \cdot \frac{nt/2}{n} = t/2 < t$, meaning the utility profile is consistent with the input, but the distortion is infinite. ☐

Deterministic Rules

By setting $\tau = 1/m$, we can achieve the following distortion upper bound.

[3] One could also conceive of using an *absolute* threshold (i.e., voter i asked to approve all a with $v_i(a) \geqslant \tau$), instead of making it relative to the total value. But in Proposition 3, we show that this leads to the worst possible distortion: unbounded for deterministic rules and m for randomized rules.

Theorem 14 (upper bound)**.** *For threshold approval votes with threshold $\tau = 1/m$, there exists a deterministic rule f with distortion*

$$\mathrm{dist}_{(1/m)\text{-}\mathsf{th}}(f) \leqslant m\left(m\gamma_{\min}^{-1} - m + 1\right).$$

Proof. We can use the voting rule that simply picks the plurality winner: the alternative with most approvals. Let a be the plurality winner.

Let S^* be the optimal feasible subset of alternatives. Then, if voter i approves alternative a:

$$\frac{v_i(a)}{\sum_{b \in A} v_i(b)} \geqslant 1/m,$$

and so:

$$mv_i(a) \geqslant v_i(A).$$

Notice that every voter must approve at least one alternative, as at least one alternative must have value at least the average: $\frac{\sum_{a \in A} v_i(a)}{m}$. Therefore, by the pigeonhole principle, the plurality winner must appear at least n/m times, and so $mv_i(a) \geqslant v_i(A)$ for at least n/m voters i.

By Lemma 1,

$$\frac{\mathsf{sw}(A)}{\mathsf{sw}(a)} \leqslant m\left(\frac{1-\gamma_{\min}}{\gamma_{\min}}m + 1\right).$$

as claimed. ☐

As with rankings by value, it turns out that linear distortion is unavoidable, even when voters exhibit perfect public spirit and submit the same vote.

Theorem 15 (lower bound)**.** *For all deterministic f and all threshold values $\tau > 0$,*

$$\mathrm{dist}_{\tau\text{-}\mathsf{th}}(f) \geqslant m - 1.$$

Proof. Let $t > 0$ be the threshold.

Consider the case where alternative a costs 1, and alternatives $b_1, \ldots, b_{m-1}$ cost $\frac{1}{m-1}$.

Suppose all voters approve only a. Then, we have two cases. If the voting rule f doesn't pick alternative a, then we incur infinite distortion when the utility of a is 1, and the utility of $b_1, \ldots, b_{m-1}$ is 0 for all voters.

If f does pick a, then it cannot pick anything else as the budget is exhausted. Let the utility of a be $t + \epsilon$ and the utility of b_j be $t - \epsilon$ for all voters, and any small $\epsilon > 0$.

Then, we could have gotten a utility of $(m-1)(t-\epsilon)$, but instead get $t + \epsilon$. As $\epsilon \to 0$, the distortion goes to $m - 1$. ☐

Randomized Rules

Turning to randomized rules for threshold approval votes with threshold τ, we get the same results under public-spirited behavior with arbitrary utilities as Benadè et al. [2021] get under the unit-sum assumption.

Theorem 16 (lower bound)**.** *For threshold approval votes with any threshold $\tau > 0$, every randomized rule f has distortion*

$$\mathrm{dist}_{\tau\text{-}\mathsf{th}}(f) \geqslant \frac{1}{2}\left(\left\lfloor \frac{\sqrt{m}}{2} \right\rfloor + 1\right).$$

Proof. We are borrowing the construction from Theorem 3.4 in Benadè et al. [2021]. Consider the case where each alternative has cost 1. We consider two cases. First suppose that $\tau \leqslant 1/\lfloor\sqrt{m}\rfloor$. Fix a set S of $\lfloor\sqrt{m}/2\rfloor + 1$ alternatives. Construct the input profile $\vec{\rho}$ where $\rho_i = S$ for all $i \in N$. There must exist $a^* \in S$ where $\Pr[a^*] \leqslant 1/|S|$. Consider the utility matrix U where for all $i \in N$, $u_i(a^*) = 1/2$ and for $a \in S \setminus \{a^*\}$, $u_i(a) = 2/\lfloor\sqrt{m}/2\rfloor$ and $u_i(a) = 0$ for $a \in A \setminus S$. Note that since voters have identical utilities, we have

$u_i(a) = v_i(a)$ for any alternative $a \in A$. We have $\mathsf{sw}(a^*) = n/2$ and for $a \in A \setminus \{a^*\}$, $\mathsf{sw}(a) \leqslant n/\sqrt{m}$. That gives us

$$\begin{aligned}
\mathrm{dist}_{\tau}\text{-}\mathsf{th}(f) &\geqslant \frac{\mathsf{sw}(a^*)}{\mathbb{E}_{a\sim f(\vec{\rho},c)}[\mathsf{sw}(a)]} \\
&\geqslant \frac{\frac{n}{2}}{\frac{1}{\lfloor\sqrt{m}/2\rfloor+1}\frac{n}{2} + \frac{\lfloor\sqrt{m}/2\rfloor}{\lfloor\sqrt{m}/2\rfloor+1}\frac{n}{\sqrt{m}}} \\
&\geqslant \frac{1}{\frac{1}{\lfloor\sqrt{m}/2\rfloor+1} + \frac{1}{\lfloor\sqrt{m}/2\rfloor+1}} \geqslant \frac{1}{2}\left(\left\lfloor\frac{\sqrt{m}}{2}\right\rfloor + 1\right).
\end{aligned}$$

On the other hand if $\tau > 1/\lfloor\sqrt{m}\rfloor$, construct the input profile $\vec{\rho}$ where $\rho_i = \emptyset$ for $i \in N$. In this case there exists $a^* \in A$ where $Pr[a^*] \leqslant 1/m$. Consider the utility matrix U where for every voter $u_i(a^*) = 1/\lfloor\sqrt{m}\rfloor$ and for $a \in A\setminus\{a^*\}$, $u_i(a) = (1-1/\lfloor\sqrt{m}\rfloor)/(m-1)$. We have $\mathsf{sw}(a^*) = n/\lfloor\sqrt{m}\rfloor$, and $sw(a) = n(1-1/\lfloor\sqrt{m}\rfloor)/(m-1)$ for $a \in A \setminus \{a^*\}$. That gives us:

$$\begin{aligned}
\mathrm{dist}_{\tau}\text{-}\mathsf{th}(f) &\geqslant \frac{\mathsf{sw}(a^*)}{\mathbb{E}_{a\sim f(\vec{\rho},c)}[\mathsf{sw}(a)]} \\
&\geqslant \frac{\frac{n}{\lfloor\sqrt{m}\rfloor}}{\frac{1}{m}\frac{n}{\lfloor\sqrt{m}\rfloor} + \frac{m-1}{m}\frac{n\left(1-\frac{1}{\lfloor\sqrt{m}\rfloor}\right)}{m-1}} \geqslant \frac{m}{\lfloor\sqrt{m}\rfloor} \geqslant \lfloor\sqrt{m}\rfloor,
\end{aligned}$$

which gives us the desired bound. □

Benadè et al. [2021] consider an additional source of randomness, whereby the designer samples a threshold τ from a distribution R over support $[0, 1]$, and then all voters are asked to submit their threshold approval votes using this value of τ (same for all voters). We refer to this ballot format as *randomized threshold approval votes* with threshold distribution D (D-rth). Note that $\mathcal{L}_{D\text{-}\mathsf{rth}} = \mathcal{L}_{\tau\text{-}\mathsf{th}} = 2^A$. Since randomness is already introduced, it makes sense to also allow the aggregation rule f to be randomized in this case. When defining the distortion of a randomized rule f, we take expectation over the sampling of threshold τ (before taking any worst case).

Theorem 17 (lower bound)**.** *For randomized threshold approval votes with the threshold sampled from any distribution D, every randomized rule f has distortion*

$$\mathrm{dist}_{D\text{-}\mathsf{rth}}(f) \geqslant \frac{1}{2}\left\lceil\frac{\log_2(m)}{\log_2(2\lceil\log_2(m)\rceil)}\right\rceil.$$

Proof. We are borrowing the construction directly from Theorem 3.6 in Benadè et al. [2021]. Consider the case where $c_a = 1$ for all $a \in A$, and let f be an arbitrary rule that both returns a threshold and a set of alternatives randomly.

Split up the $(1/m, 1]$ interval into $\lceil\log_2(m)/\log_2(2\log_2(m))\rceil$ parts I_j defined such that

$$I_j = \left(\frac{(2\log_2(m))^{j-1}}{m}, \min\left\{\frac{(2\log_2(m))^{j}}{m}, 1\right\}\right].$$

Define u_j and ℓ_j to be the largest and smallest points in I_j respectively. By construction, $u_j \leqslant 2\log_2(m)\ell_j$ for all j.

The key idea is to give utilities to alternatives within the interval that the threshold with least probability is contained in, so that with high probability, the alternatives are either all approved or all disapproved.

Indeed, let k be a value such that

$$\Pr(t \in I_k) \leqslant \lceil\log_2(m)/\log_2(2\log_2(m))\rceil^{-1},$$

which must exist by the pigeonhole principle.

Fix a subset $S \subseteq A$ of size $\lceil \log_2(m) \rceil$, and let $V = u_k/2 + (\lceil \log_2(m) \rceil - 1)\ell_k$.

We will give each voter the same utilities, so that $u(a) := u_i(a) = v_i(a)$ for all $i \in N, a \in A$. For all $a \in S$, assign utilities so that $\sum_{a \in S} u(a) = V$, for all $a \notin S$, let $u(a) = (1 - V)/(m - \lceil \log_2(m) \rceil)$.

We can verify that $\ell_k \leqslant \frac{1}{2\log_2(m)} u_k$ for all k. We can then see that the utilities sum to one, and are all positive as:

$$V = \frac{u_k}{2} + (\lceil \log_2(m) \rceil - 1)\ell_k \leqslant \frac{1}{2} + \frac{\lceil \log_2(m) \rceil - 1}{2\log_2(m)} \leqslant 1.$$

We construct this so that all alternatives in S have utilities contained in I_k. Thus, when $t \notin I_k$, all voters either approve S or disapprove S. Therefore, there must exist some $a^* \in S$ such that

$$\Pr(a^* \text{is returned} \mid t \notin I_k) \leqslant 1/\lceil \log_2(m) \rceil .$$

Now, we can assign $u(a^*) = u_k/2$ and $u(a) = \ell_k$ for $a \in S \setminus \{a^*\}$. Then, the optimal choice is a^* with social welfare $nu_k/2$, but instead, since $\ell_k > (1 - V)/(m - \log_2(m))$, we pick with high probability an alternative with at most $n\ell_k$ utility.

Indeed, the expected social welfare of f is:

$$\begin{aligned}
&\Pr(t \in I_k) \cdot \frac{nu_k}{2} + \Pr(t \notin I_k)\left(\frac{1}{\lceil \log_2(m) \rceil} \cdot \frac{nu_k}{2} + \frac{\lceil \log_2(m) \rceil - 1}{\lceil \log_2(m) \rceil} \cdot n\ell_k\right) \\
&\leqslant \left(\lceil \log_2(m)/\log_2(2\log_2(m)) \rceil^{-1} + \frac{1}{\lceil \log_2(m) \rceil} + \frac{\lceil \log_2(m) \rceil - 1}{\lceil \log_2(m) \rceil} \cdot \frac{1}{\log_2(m)}\right) \frac{nu_k}{2} \\
&\leqslant \left(\lceil \log_2(m)/\log_2(2\log_2(m)) \rceil^{-1}\right) nu_k.
\end{aligned}$$

The maximum social welfare that we can get is $nu_k/2$, so the distortion is:

$$\text{dist}_{D\text{-}\mathsf{rth}}(f) \geqslant \frac{\frac{nu_k}{2}}{nu_k \left\lceil \frac{\log_2(m)}{\log_2(2\log_2(m))} \right\rceil^{-1}} = \frac{1}{2} \left\lceil \frac{\log_2(m)}{\log_2(2\lceil \log_2(m) \rceil)} \right\rceil . \qquad \square$$

Theorems 16 and 17 are corollaries of Theorems 3.4 and 3.6 of Benadè et al. [2021], respectively. Their lower bound, derived under the unit-sum assumption, carries over to our more general setup. While they do not allow public-spirited behavior, in their construction the utility of each alternative is the same across all voters, ensuring that any level of public-spirited behavior does not affect their construction. The only reason we provide full proofs is that Benadè et al. [2021] derive only an asymptotic lower bound by making several simplifying assumptions, which we carefully remove to derive an exact lower bound.

5 PB with Ranking of Predefined Bundles

We have shown that for all commonly-studied PB ballot formats, all deterministic rules incur $\Omega(m)$ distortion — an issue in the practical case where m is large. This motivates our study of a novel ballot format, *ranking of predefined bundles* (rank-b($\mathcal{P}$)). In Sections 5.1-5.3, we will explore various ways to use rank-b($\mathcal{P}$) ballots, which differ in how $\mathcal{P}$ is chosen. For intuition, note that the lowest-distortion choice of $\mathcal{P}$ is simply $\mathcal{F}$; then, we are effectively in the single-winner setting and *Copeland* guarantees constant (in m) distortion. However, this choice of $\mathcal{P}$ comes at a steep elicitation cost, requiring voters to rank exponentially many bundles.

Our refined goal, therefore, is to design $\mathcal{P}$ to permit low distortion *while containing at most polynomial (or pseudopolynomial) bundles*. After designing and analyzing various such choices of $\mathcal{P}$, in Section 6 we explore the practicality of the resulting elictation protocols.

5.1 Sublinear Distortion

We first propose rank-b with *high-low bundles* ($\mathcal{P}$ = HLB), which we show permits sublinear distortion.

High-low bundles (HLB): Let $L = \{a \in A : c(a) \leqslant 1/\lceil\sqrt{m}\rceil\}$ be the set of *low-cost* alternatives and $H = A \setminus L$ be the set of *high-cost* alternatives. The *high-low bundling* rule (HLB) partitions L into at most $\lceil\sqrt{m}\rceil$ feasible bundles,[4] and H into an arbitrary partition of feasible bundles. Then it defines $\mathcal{P}$ to be the union of these partitions.

The rank-b(HLB) ballot asks voters to rank $|\mathcal{P}| \leqslant |H| + |L| = m$ bundles (and in fact, $|\mathcal{P}| \leqslant m - |L| \cdot (1 - 1/\lceil\sqrt{m}\rceil)$, so if there are many low-cost projects, $|\mathcal{P}| \ll m$). Finally, in Theorem 18 we show that if *Copeland* rule is applied to the voting profile elicited via rank-b(HLB) on $\mathcal{P}$, the rank-b ballot format dominates all the previous ballot formats by a factor of $O(\sqrt{m})$.

The main idea of our proof is that if A^* is the optimal bundle, then either $\mathsf{sw}(L \cap A^*)$ or $\mathsf{sw}(H \cap A^*)$ must have welfare at least $\mathsf{sw}(A^*)/2$. Then, there must be a bundle in $\mathcal{P}$ with welfare at least $(1/\lceil\sqrt{m}\rceil) \cdot \mathsf{sw}(L \cap A^*)$ because L was partitioned into at most $\lceil\sqrt{m}\rceil$ bundles, and also one with welfare at least $(1/\lceil\sqrt{m}\rceil) \cdot \mathsf{sw}(H \cap A^*)$ because $|H \cap A^*| \leqslant \lceil\sqrt{m}\rceil$.

Theorem 18. $\text{dist}_{\mathsf{rank\text{-}b}(\text{HLB})}(\text{Copeland}) = O(\sqrt{m}/\gamma_{\min}^2)$.

Proof. Let A^* be an optimal budget-feasible set of alternatives. Clearly, $\mathsf{sw}(A^*) = \mathsf{sw}(L \cap A^*) + \mathsf{sw}(H \cap A^*)$, implying that at least one of $\mathsf{sw}(L \cap A^*) \geqslant \frac{1}{2}\mathsf{sw}(A^*)$ and $\mathsf{sw}(H \cap A^*) \geqslant \frac{1}{2}\mathsf{sw}(A^*)$ must be true. In both cases, we claim that there exists a bundle $P^* \in \mathcal{P}$ for which $\mathsf{sw}(P^*) \geqslant \frac{\mathsf{sw}(A^*)}{2\lceil\sqrt{m}\rceil}$.

Suppose $\mathsf{sw}(L \cap A^*) \geqslant \frac{1}{2}\mathsf{sw}(A^*)$. Since L is partitioned into at most $\lceil\sqrt{m}\rceil$ bundles in $\mathcal{P}$, there exists $P^* \in \mathcal{P}$ such that $\mathsf{sw}(P^*) \geqslant \frac{\mathsf{sw}(L)}{\lceil\sqrt{m}\rceil} \geqslant \frac{\mathsf{sw}(L \cap A^*)}{\lceil\sqrt{m}\rceil} \geqslant \frac{\mathsf{sw}(A^*)}{2\lceil\sqrt{m}\rceil}$.

Next, suppose $\mathsf{sw}(H \cap A^*) \geqslant \frac{1}{2}\mathsf{sw}(A^*)$. Since each alternative in $H \cap A^*$ has cost more than $\frac{1}{\lceil\sqrt{m}\rceil}$ and lies in the budget-feasible set A^*, we have that $|H \cap A^*| \leqslant \lceil\sqrt{m}\rceil$. Thus, there exists an alternative $a^* \in H \cap A^*$ with $\mathsf{sw}(a^*) \geqslant \frac{\mathsf{sw}(H \cap A^*)}{\lceil\sqrt{m}\rceil} \geqslant \frac{\mathsf{sw}(A^*)}{2\lceil\sqrt{m}\rceil}$. Hence, for the bundle $P^* \in \mathcal{P}$ containing a^*, we have $\mathsf{sw}(P^*) \geqslant \frac{\mathsf{sw}(A^*)}{2\lceil\sqrt{m}\rceil}$.

Finally, if Copeland applied to the rank-b(HLB) ballots picks bundle P, using its distortion bound, we have

$$\mathsf{sw}(P) \geqslant \gamma_{\min}^2 \cdot \mathsf{sw}(P^*) \geqslant \gamma_{\min}^2 \cdot \frac{\mathsf{sw}(A^*)}{2\lceil\sqrt{m}\rceil},$$

yielding distortion at most $\frac{2\lceil\sqrt{m}\rceil}{\gamma_{\min}^2} = O(\sqrt{m}/\gamma_{\min}^2)$, as needed. □

While this is already a significant improvement on previous results, there is room for more: crafting predefined bundles with no information about (and thus no regard for) voters' preferences can be both theoretically lossy and practically unappealing. Thus, we next explore: *what distortion is possible when our bundling rule has some knowledge of voters' preferences?* We explore this question in Sections 5.2 and 5.3 by defining a two-round elicitation protocol: in Round 1, we elicit voter preferences using the canonical rank ballot format; then, in Round 2, we use this preference information to craft $\mathcal{P}$ and deploy ballot rank-b($\mathcal{P}$). We denote this two-round ballot format as rank → rank-b($\mathcal{P}$).

5.2 Logarithmic Distortion in Two Rounds

We now propose rank-b with *tiered-cost bundles* ($\mathcal{P}$ = TCB). At a high level, TCB partitions alternatives into $O(\log m)$ tiers by cost, and then uses *Iterative Copeland* to select a feasible bundle of $m/2^\ell$ alternatives from the tier containing alternatives with costs between $2^{\ell-1}/m$ and $2^\ell/m$.

Tiered-cost bundles (TCB): Set $L = \lceil \log_2 m \rceil$. For each $\ell \in [L]$, define the tier T_ℓ such that

$$T_\ell = \{a \in A : 2^{\ell-1}/m < c(a) \leqslant 2^\ell/m\} \quad \text{for all } \ell > 0;$$

[4] This is possible because $|L| \leqslant m$ and any subset of $\lceil\sqrt{m}\rceil$ alternatives from L is feasible.

let $T_0 = \{a \in A : c(a) \leqslant 1/m\}$. Then, use *Iterative Copeland* to pick a bundle $P_\ell \subseteq T_\ell$ of size $t_\ell = \lfloor \min(|T_\ell|, \max(1, m/2^\ell)) \rfloor$. Since $c(a) \leqslant 2^\ell/m$ for each $a \in T_\ell$, P_ℓ is budget-feasible. Set $\mathcal{P} = \{P_0, P_1, \ldots, P_L\}$. Then, TCB asks voters to rank $L \leqslant 1 + \lceil \log_2 m \rceil$ bundles.

After asking voters to rank a total of at most $m + 1 + \lceil \log_2 m \rceil$ objects over both rounds, aggregating via *Copeland* achieves distortion $O(\log(m)/\gamma_{\min}^4)$.

The key insight is that for the optimal bundle A^*, $\mathsf{sw}(A^*) = \sum_\ell \mathsf{sw}(A^* \cap T_\ell)$, so the best of the $1 + \lceil \log_2 m \rceil$ feasible bundles in the sum (call it $A^* \cap T_{\ell'}$) must be an $O(\log m)$ approximation of A^*. Then, the welfare of the best t_ℓ-sized subset $P_\ell^* \subseteq T_{\ell'}$ 2-approximates that of $A^* \cap T_{\ell'}$; P_ℓ constant-approximates the welfare of P_ℓ^* (by the distortion of *Iterative Copeland*); and the chosen bundle constant-approximates the welfare of P_ℓ (by the distortion of the final *Copeland* aggregation).

Theorem 19. $\mathrm{dist}_{\mathsf{rank}\to\mathsf{rank\text{-}b}(\mathrm{TCB})}(\mathrm{Copeland}) = O(\log(m)/\gamma_{\min}^4)$.

Proof. Let A^* be an optimal budget-feasible set of the alternatives. Choose $\ell \in \{0, 1, \ldots, L\}$ with the highest $\mathsf{sw}(A^* \cap T_\ell)$; note that, by the pigeonhole principle, $\mathsf{sw}(A^* \cap T_\ell) \geqslant \frac{\mathsf{sw}(A^*)}{1+L}$.

Let P_ℓ^* be the optimal t_ℓ-sized subset of T_ℓ; note that this is feasible due to the definition of t_ℓ. Further, since $A^* \cap T_\ell$ is feasible, we have $|A^* \cap T_\ell| \leqslant 2t_\ell$. Hence, $A^* \cap T_\ell$ can be partitioned into two t_ℓ-sized subsets of T_ℓ, the better of which must be a 2-approximation of $A^* \cap T_\ell$. Since P_ℓ^* is the best t_ℓ-sized subset of T_ℓ, we have $\mathsf{sw}(P_\ell^*) \geqslant \frac{1}{2}\mathsf{sw}(A^* \cap T_\ell)$.

Next, because we pick $P_\ell \subseteq T_\ell$ of size t_ℓ using the iterated Copeland rule, given its distortion bound of $(2\gamma_{\min}^{-1} - 1)^2$ from Lemma 4, and the the distortion of Copeland given in Theorem 3.3 of Flanigan et al. [2023], we have

$$\mathsf{sw}(P_\ell) \geqslant \frac{\mathsf{sw}(P_\ell^*)}{(2\gamma_{\min}^{-1} - 1)^2} \geqslant \frac{\mathsf{sw}(A^* \cap T_\ell)}{2 \cdot (2\gamma_{\min}^{-1} - 1)^2} \geqslant \frac{\mathsf{sw}(A^*)}{2 \cdot (1 + L) \cdot (2\gamma_{\min}^{-1} - 1)^2}.$$

Finally, since we pick a bundle $P \in \mathcal{P}$ using Copeland's rule, using its distortion bound again, we have

$$\mathsf{sw}(P) \geqslant \frac{\mathsf{sw}(P_\ell)}{(2\gamma_{\min}^{-1} - 1)^2} \geqslant \frac{\mathsf{sw}(A^*)}{2 \cdot (1 + L) \cdot (2\gamma_{\min}^{-1} - 1)^4},$$

yielding a distortion of at most $2 \cdot (1 + L) \cdot (2\gamma_{\min}^{-1} - 1)^4 = O(\log m/\gamma_{\min}^4)$. □

5.3 Constant Distortion in Two Rounds

Finally, we propose rank-b with *exhaustive bundles* ($\mathcal{P} = \mathrm{EB}$). EB uses the same tiers as TCB, but instead of having each bundle consist of alternatives from the same tier, it crafts bundles by using *Iterative Copeland* to choose a subset $S_\ell \subseteq T_\ell$ of size t_ℓ from *every* tier and putting them together as $\cup_\ell S_\ell$. Ideally, we want to explore every possible combination of values of $t_0, \ldots, t_{\lceil \log_2 m \rceil}$, so long as the resulting bundle is feasible, but a slight optimization is achieved by only choosing values that are powers of 2.

Exhaustive bundles (EB): Let $L = \lceil \log_2 m \rceil$ and define tiers $T_0, T_1, \ldots, T_L$ as in TCB. Fix $R = \lfloor \log_2 m \rfloor$. For each $\ell \in [L]$ and $r \in \{0, 2^0, 2^1, \ldots, 2^R\}$ such that $|T_\ell| \geqslant r$, choose $P_{\ell,r} \subseteq T_\ell$ of size r using *Iterative Copeland* applied to the rank ballots from Round 1 (if $p = 0$, simply choose $\emptyset$). Call a sequence $\vec{t} = (t_0, t_1, \ldots, t_L)$ *valid* if $t_\ell \in \{0, 2^0, 2^1, \ldots, 2^R\}$ for each $\ell \in [L]$, and for such a sequence define $P_{\vec{t}} = \cup_{\ell=0}^{L} P_{\ell,t_\ell}$. In other words, in a valid sequence, the ℓ-th element represents the number of alternatives that are selected from T_ℓ, and with this definition each valid sequence leads to a potential bundle. Finally, let $\mathcal{P} = \{P_{\vec{t}} : \vec{t} \text{ is valid} \wedge P_{\vec{t}} \in \mathcal{F}\}$, so all bundles in $\mathcal{P}$ are feasible. Note that $|\mathcal{P}|$ is at most the number of valid sequences, which is $(1 + R)^{1+L} = O((\log m)^{O(\log m)}) = O(m^{O(\log \log m)})$.[5]

In exchange for asking voters to rank quasipolynomially many objects across two rounds, using *Copeland* to aggregate, we achieve constant distortion! The key idea is to let A^* be an optimal bundle. Then, consider the sequence $\vec{t}$, where $t_\ell = 2^{\lfloor \log_2 |A^* \cap T_\ell| \rfloor - 1}$ for each $\ell \in \{0, 1, \ldots, L\}$. Since $t_\ell \leqslant |A^* \cap T_\ell|/2$, $c(P_{\ell,t_\ell}) \leqslant c(A^* \cap T_\ell)$,

[5] Although this is many bundles, they are similar, potentially decreasing cognitive load of ranking them: they consist of combinations of at most $(1 + R) \cdot (1 + L) = O(m)$ many bundles $P_{\ell,r}$.

$P_{\vec{t}}$ must be feasible. We then show that its welfare constant-approximates that of A^*. The only remaining issue is that when $|A^* \cap T_\ell| = 1$, we cannot set $t_\ell = 2^{\lfloor \log_2 |A^* \cap T_\ell| \rfloor - 1}$. This is addressed by taking two cases, depending on whether much of the welfare of A^* is contributed by those $A^* \cap T_\ell$ that have size 1 or those that have size at least 2.

Theorem 20. $\text{dist}_{\mathsf{rank} \to \mathsf{rank\text{-}b(EB)}}(\text{Copeland}) = O(1/\gamma_{\min}^4)$.

Proof. Let A^* be an optimal budget-feasible set of the alternatives. Let $U^* = \cup_{\ell \in \{0,1,\dots,L\} : |A^* \cap T_\ell| = 1}(A^* \cap T_\ell)$ be the set of alternatives in A^* such that there is no other alternative in A^* from their tier. First, we show that there exists a bundle $P^* \in \mathcal{P}$ such that $\mathsf{sw}(P^*) \geqslant \frac{1}{9\gamma_{\min}^2} \cdot \mathsf{sw}(A^*)$. We do so by splitting into two cases:

Case 1: $\mathsf{sw}(U^*) < \frac{1}{3}\mathsf{sw}(A^*)$. Here, we seek a bundle in $\mathcal{P}$ that provides a good approximation to $B^* = A^* \setminus U^*$ because $\mathsf{sw}(B^*) \geqslant \frac{2}{3}\mathsf{sw}(A^*)$. We consider two sub-cases:

Case 1a: $\mathsf{sw}(B^* \cap T_0) \geqslant \frac{2}{9}\mathsf{sw}(A^*)$. In this case, consider $P_{0,r}$ for the greatest feasible r, and note that $r \geqslant |T_0|/2$. The best r-sized subset of T_0 is $1/2$-approximation of T_0, so, by the distortion of iterated Copeland,

$$\mathsf{sw}(P_{0,r}) \geqslant \frac{\mathsf{sw}(T_0)}{2 \cdot (2\gamma_{\min}^{-1} - 1)^2} \geqslant \frac{\mathsf{sw}(B^* \cap T_0)}{2 \cdot (2\gamma_{\min}^{-1} - 1)^2} \geqslant \frac{\mathsf{sw}(A^*)}{9 \cdot (2\gamma_{\min}^{-1} - 1)^2}.$$

Case 1b: $\mathsf{sw}(B^* \cap T_0) < \frac{2}{9}\mathsf{sw}(A^*)$. Hence, $\mathsf{sw}(B^* \cap \cup_{\ell \in [L]} T_\ell) = \mathsf{sw}(B^*) - \mathsf{sw}(B^* \cap T_0) \geqslant \frac{4}{9}\mathsf{sw}(A^*)$. Define $t_0 = 0$, and for each $\ell \in [L]$, define

$$t_\ell = \begin{cases} 0, & \text{if } |A^* \cap T_\ell| = 1, \\ 2^{\lfloor \log_2 |A^* \cap T_\ell| \rfloor - 1}, & \text{if } |A^* \cap T_\ell| \geqslant 2. \end{cases}$$

First, note that $\vec{t}$ is a valid sequence. Next, we prove that $P_{\vec{t}} = \cup_{\ell=0}^{L} P_{\ell,t_\ell}$ is feasible. Note that $t_0 = 0$, so $P_{0,t_0} = \emptyset$. For each $\ell \in [L]$, $|P_{\ell,t_\ell}| = t_\ell \leqslant \frac{|A^* \cap T_\ell|}{2}$. Since alternatives in T_ℓ differ from each other in cost by a factor of at most 2, this implies $c(P_{\ell,t_\ell}) \leqslant c(A^* \cap T_\ell)$. Hence, $c(P_{\vec{t}}) \leqslant \sum_{\ell=1}^{T} c(A^* \cap T_\ell) \leqslant c(A^*) \leqslant 1$; hence, $P_{\vec{t}} \in \mathcal{P}$.

Finally, for each $\ell \in [L]$ such that $|A^* \cap T_\ell| \geqslant 2$, note that $t_\ell \geqslant \frac{1}{4}|A^* \cap T_\ell|$; hence, the best t_ℓ-sized subset of T_ℓ is a $(1/4)$-approximation of $A^* \cap T_\ell$, and applying the distortion guarantee of iterated Copeland, we have

$$\mathsf{sw}(P_{\ell,t_\ell}) \geqslant \frac{\mathsf{sw}(A^* \cap T_\ell)}{4 \cdot (2\gamma_{\min}^{-1} - 1)^2}.$$

Summing over $\ell \in [L]$ such that $|A^* \cap T_\ell| \geqslant 2$, we have

$$\mathsf{sw}(P_{\vec{t}}) \geqslant \frac{\mathsf{sw}(B^* \cap \cup_{\ell \in [L]} T_\ell)}{4 \cdot (2\gamma_{\min}^{-1} - 1)^2} \geqslant \frac{\mathsf{sw}(A^*)}{9 \cdot (2\gamma_{\min}^{-1} - 1)^2}.$$

Case 2: $\mathsf{sw}(U^*) \geqslant \frac{1}{3}\mathsf{sw}(A^*)$. In this case, we seek a bundle in $\mathcal{P}$ that provides a good approximation of U^*. To do this, we consider three bundles, and prove that at least one of which must be a sufficiently good approximation.

Case 2a: $\mathsf{sw}(U^* \cap T_L) \geqslant \frac{1}{3}\mathsf{sw}(U^*)$. Then, since $|U^* \cap T_L| = 1$, we can take $P_{L,1}$, which also has size $|P_{L,1}| = 1$. Further, by the distortion guarantee of iterated Copeland, we have

$$\mathsf{sw}(P_{L,1}) \geqslant \frac{\mathsf{sw}(U^* \cap T_L)}{(2\gamma_{\min}^{-1} - 1)^2} \geqslant \frac{\mathsf{sw}(A^*)}{9 \cdot (2\gamma_{\min}^{-1} - 1)^2}.$$

Case 2b: $\mathsf{sw}(U^* \cap T_{L-1}) \geqslant \frac{1}{3}\mathsf{sw}(U^*)$. Similarly, since $|U^* \cap T_{L-1}| = 1$, we can take $P_{L-1,1}$, netting

$$\mathsf{sw}(P_{L-1,1}) \geqslant \frac{\mathsf{sw}(U^* \cap T_{L-1})}{(2\gamma_{\min}^{-1} - 1)^2} \geqslant \frac{\mathsf{sw}(A^*)}{9 \cdot (2\gamma_{\min}^{-1} - 1)^2}.$$

Case 2c: $\mathsf{sw}(U^* \setminus (T_L \cup T_{L-1})) \geqslant \frac{2}{3}\mathsf{sw}(U^*)$. Hence, $\mathsf{sw}(U^* \cap (\cup_{\ell=0}^{L-2} T_\ell)) \geqslant \frac{1}{3}\mathsf{sw}(U^*)$. Take $\vec{t}$ where $t_\ell = 1$ for each $\ell \in \{0, 1, \ldots, L-2\}$ and $t_L = 0$. Note that this is a valid sequence. Further, $c(P_{\vec{t}}) \leqslant \sum_{\ell=0}^{L-2} \frac{2^\ell}{m} = \frac{2^{L-1}-1}{m} \leqslant 1$. By the distortion of iterated Copeland, we have

$$\mathsf{sw}(P_{\vec{t}}) \geqslant \frac{\mathsf{sw}(U^* \cap \cup_{\ell=0}^{L-1} T_\ell)}{(2\gamma_{\min}^{-1} - 1)^2} \geqslant \frac{\mathsf{sw}(A^*)}{9 \cdot (2\gamma_{\min}^{-1} - 1)^2}.$$

Finally, applying the distortion bound of the final Copeland aggregation, the bundle P picked must satisfy

$$\mathsf{sw}(P) \geqslant \frac{\mathsf{sw}(P^*)}{(2\gamma_{\min}^{-1} - 1)^2} \geqslant \frac{\mathsf{sw}(A^*)}{9 \cdot (2\gamma_{\min}^{-1} - 1)^4},$$

which implies a distortion of $O(1/\gamma_{\min}^4)$, as desired. □

6 Discussion

On the practicality of rank-b ballots. We propose that our *ranking of predefined bundles* ballot format has three main practical advantages in addition to low distortion. (1) It is fully ordinal in contrast to, e.g., threshold approval votes [Benadè et al., 2021], which ask voters to compare projects via precise numeric utility values. (2) Comparing entire feasible bundles may provide voters more context about cost trade-offs than ballots where they compare individual projects. (3) The aggregation rules we study always select a bundle that is on the ballot, allowing every vote favorably ranking the winning bundle to be interpreted as a direct endorsement of the final outcome.

While one may worry that doing two rounds of elicitation is impracticable, PB participants often meet several times, so doing so is likely feasible. In fact, the flexibility in the PB process may permit a variety multi-round protocols with even more favorable trade-offs. A second potential worry is that low-distortion rank-b ballots may ask voters to rank too many bundles. We now show that in 1244 real PB elections from https://pabulib.org (with some randomized imputation of incomplete preferences), even our rank $\to$ rank-b(EB) ballot (with some minor heuristic tweaks that maintain constant distortion) typically requires voters to rank *far less than m bundles*. Details on data and implementation, plus some supplemental results, are found in Appendix E.

Future directions. rank-b ballots represent an exciting future direction in PB: there is an expansive design space of bundles and multi-round protocols that can potentially drive down query complexity, guarantee low distortion, satisfy desirable axioms, and be well-received in real-world experiments. Beyond rank-b ballots, there are many questions remaining about what public spirit looks like in real democratic contexts, and how this can be incorporated into social choice theory. For instance: (1) In what other social choice contexts — such as matching [Filos-Ratsikas et al., 2014] and fair division [Halpern and Shah, 2021] — is public spirit a reasonable assumption?, (2) Can we measure the degree and nature of public spirit resulting from different democratic processes, deliberative or otherwise?, and (3) If voters account for a welfare notion other than utilitarian social welfare (or likewise, we care about other objectives like Nash welfare or proportional fairness [Ebadian et al., 2022]), can one prove similar guarantees?

References

E. Anshelevich, O. Bhardwaj, E. Elkind, J. Postl, and P. Skowron. Approximating optimal social choice under metric preferences. *Artificial Intelligence*, 264:27–51, 2018.

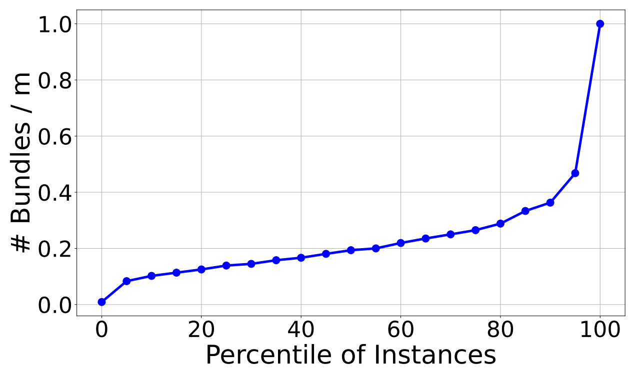

Figure 1: The number of bundles rank-b(EB) asks voters to rank, per alternative, in 1244 instances ordered by quantile.

Elliot Anshelevich, Aris Filos-Ratsikas, Nisarg Shah, and Alexandros A Voudouris. Distortion in social choice problems: The first 15 years and beyond. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 4294–4301, 2021. Survey Track.

Haris Aziz and Nisarg Shah. Participatory budgeting: Models and approaches. In Tamás Rudas and Gábor Péli, editors, *Pathways Between Social Science and Computational Social Science: Theories, Methods, and Interpretations*, pages 215–236. Springer, 2021.

Dorothea Baumeister, Linus Boes, and Tessa Seeger. Irresolute approval-based budgeting. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, pages 1774–1776, 2020.

G. Benadè, S. Nath, A. D. Procaccia, and N. Shah. Preference elicitation for participatory budgeting. *Management Science*, 65(5):2813–2827, 2021.

Allan Borodin, Daniel Halpern, Mohamad Latifian, and Nisarg Shah. Distortion in voting with top-t preferences. In *Proceedings of the 31st International Joint Conference on Artificial Intelligence (IJCAI)*, pages 116–122, 2022.

C. Boutilier, I. Caragiannis, S. Haber, T. Lu, A. D. Procaccia, and O. Sheffet. Optimal social choice functions: A utilitarian view. *Artificial Intelligence*, 227:190–213, 2015.

Markus Brill, Stefan Forster, Martin Lackner, Jan Maly, and Jannik Peters. Proportionality in approval-based participatory budgeting. In *Thirty-Seventh AAAI Conference on Artificial Intelligence*, pages 5524–5531, 2023.

I. Caragiannis and A. D. Procaccia. Voting almost maximizes social welfare despite limited communication. *Artificial Intelligence*, 175(9–10):1655–1671, 2011.

Ioannis Caragiannis, Swaprava Nath, Ariel D. Procaccia, and Nisarg Shah. Subset selection via implicit utilitarian voting. *Journal of Artificial Intelligence Research*, 58:123–152, 2017.

Michiel S De Vries, Juraj Nemec, and David Špaček. International trends in participatory budgeting. *Cham: Palgrave Macmillan*, 2022.

Soroush Ebadian, Anson Kahng, Dominik Peters, and Nisarg Shah. Optimized distortion and proportional fairness in voting. In *Proceedings of the 23rd ACM Conference on Economics and Computation (EC)*, pages 563–600, 2022.

Brandon Fain, Kamesh Munagala, and Nisarg Shah. Fair allocation of indivisible public goods. In *Proceedings of the 19th ACM Conference on Economics and Computation (EC)*, pages 575–592, 2018.

Piotr Faliszewski, Jarosław Flis, Dominik Peters, Grzegorz Pierczyński, Piotr Skowron, Dariusz Stolicki, Stanisław Szufa, and Nimrod Talmon. Participatory budgeting: data, tools, and analysis. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 2667–2674, 2023.

Aris Filos-Ratsikas and Peter Bro Miltersen. Truthful approximations to range voting. In *Proceedings of the 10th Conference on Web and Internet Economics (WINE)*, pages 175–188, 2014.

Aris Filos-Ratsikas, Søren Kristoffer Stiil Frederiksen, and Jie Zhang. Social welfare in one-sided matchings: Random priority and beyond. In *International Symposium on Algorithmic Game Theory*, pages 1–12, 2014.

Peter C Fishburn. Probabilistic social choice based on simple voting comparisons. *The Review of Economic Studies*, 51(4):683–692, 1984.

David C Fisher and Jennifer Ryan. Tournament games and positive tournaments. *Journal of Graph Theory*, 19(2):217–236, 1995.

Bailey Flanigan, Ariel D Procaccia, and Sven Wang. Distortion under public-spirited voting. In *Proceedings of the 24th ACM Conference on Economics and Computation*, page 700, 2023.

J. Gastil, C. Bacci, and M. Dollinger. Is deliberation neutral? patterns of attitude change during the deliberative polls. *Journal of public deliberation*, 6(2), 2010.

Ashish Goel, Reyna Hulett, and Anilesh K. Krishnaswamy. Relating metric distortion and fairness of social choice rules. In *Proceedings of the 13th Workshop on Economics of Networks, Systems and Computation*, NetEcon '18, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450359160. doi: 10.1145/3230654.3230658. URL https://doi.org/10.1145/3230654.3230658.

Daniel Halpern and Nisarg Shah. Fair and efficient resource allocation with partial information. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 224–230, 2021.

Ararat Harutyunyan, Tien-Nam Le, Alantha Newman, and Stéphan Thomassé. Domination and fractional domination in digraphs. *arXiv preprint arXiv:1708.00423*, 2017.

D. R. Kinder and D. R. Kiewiet. Sociotropic politics: the american case. *British Journal of Political Science*, 11(2), 1981.

Germain Kreweras. Aggregation of preference orderings. In *Mathematics and Social Sciences I: Proceedings of the seminars of Menthon-Saint-Bernard*, pages 73–79, 1965.

Gilbert Laffond, Jean-Francois Laslier, and Michel Le Breton. The bipartisan set of a tournament game. *Games and Economic Behavior*, 5(1):182–201, 1993.

Participedia. https://participedia.net/search?selectedCategory=all&query=participatory%20budgeting, 2023. Accessed: 2025-01-27.

Dominik Peters, Grzegorz Pierczyński, and Piotr Skowron. Proportional participatory budgeting with additive utilities. *Advances in Neural Information Processing Systems*, 34:12726–12737, 2021.

A. D. Procaccia and J. S. Rosenschein. The distortion of cardinal preferences in voting. In *Proceedings of the 10th International Workshop on Cooperative Information Agents (CIA)*, pages 317–331, 2006.

Simon Rey and Jan Maly. The (computational) social choice take on indivisible participatory budgeting. arXiv:2303.00621, 2023.

Simon Rey, Ulle Endriss, and Ronald de Haan. Designing participatory budgeting mechanisms grounded in judgment aggregation. In *Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning*, pages 692–702, 2020.

Ronald L Rivest and Emily Shen. An optimal single-winner preferential voting system based on game theory. In *Proc. of 3rd International Workshop on Computational Social Choice*, pages 399–410, 2010.

Nimrod Talmon and Piotr Faliszewski. A framework for approval-based budgeting methods. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI)*, pages 2181–2188, 2019.

Brian Wampler, Stephanie McNulty, and Michael Touchton. *Participatory budgeting in global perspective*. Oxford University Press, 2021.

R. Wang, J. S. Fishkin, and R. C. Luskin. Does deliberation increase public-spiritedness? *Social Science Quarterly*, 101(6):2163–2182, 2020.

Appendix

Table of Contents

A Constructions of the Lower Bounds in Section 3

A.1 Proof of Theorem 2

Theorem 2. *Every deterministic rule f has distortion*

$$\mathrm{dist}_{\mathsf{rank}}(f) \in \Omega\left(m/\gamma_{\min}\right).$$

Proof. Consider an instance with $A = \{a, b_1, \ldots b_{m-1}\}$, where a costs 1 and every other alternative costs $1/(m-1)$. Define $p = \frac{1-\gamma_{\min}}{1-\gamma_{\min}+m^2}$. Let N_1 be a set of $n(1-p)$ voters and $N_2 = N \setminus N_1$. Suppose that members of N_1 submit ranking $(a \succ b_1 \succ \ldots \succ b_{m-1})$ and members of N_2 vote $(b_1 \succ \ldots \succ b_{m-1} \succ a)$.

Now consider two cases.

Case 1: If the aggregation rule selects a, consider utility matrix U where members of N_1 have utility of $\frac{\gamma_{\min} p}{1-p\gamma_{\min}}$ for a and 0 for the rest, while members of N_2 have utility of 0 for a and 1 for the rest of the alternatives. This means $\mathsf{sw}(a) = n(1-p)\frac{\gamma_{\min} p}{1-\gamma_{\min} p}$, and $\mathsf{sw}(b) = np$ for $b \in A \setminus \{a\}$. Alongside with the PS-vector $\vec{\gamma} = [\gamma_{\min}]^n$ we have value matrix $V_{\vec{\gamma},U}$ first of all we have to make sure that this is consistent with the input profile. For $i \in N_1$,

$$\begin{aligned} v_i(a) &= (1-\gamma_{\min})\frac{\gamma_{\min} p}{1-\gamma_{\min} p} + \gamma_{\min}(1-p)\frac{\gamma_{\min} p}{1-\gamma_{\min} p} \\ &= (1-\gamma_{\min} p)\frac{\gamma_{\min} p}{1-\gamma_{\min} p} = \gamma_{\min} p, \end{aligned}$$

and $v_i(b_j) = (1-\gamma_{\min}) \cdot 0 + \gamma_{\min} p = \gamma_{\min} \cdot p$. Therefore, the value matrix is consistent with the ranking of the members of N_1. On the other hand for $i \in N_2$ we have, $v_i(a) = \gamma_{\min}(1-p)\frac{\gamma_{\min} p}{1-\gamma_{\min} p}$, and $v_i(b_j) = 1-\gamma_{\min}+\gamma_{\min} p$, where for $p = \frac{1-\gamma_{\min}}{1-\gamma_{\min}+m^2}$ we have:

$$\begin{aligned} v_i(a) &= \frac{\gamma_{\min}^2 m^2 \left(1-\gamma_{\min}\right)}{\left(m^2+1-\gamma_{\min}\right)\left(m^2+(1-\gamma_{\min})^2\right)}, \\ v_i(b_j) &= \frac{(m^2+1)(1-\gamma_{\min})}{m^2+1-\gamma_{\min}}. \end{aligned}$$

This gives us:

$$\begin{aligned} \frac{v_i(a)}{v_i(b_j)} &= \frac{\gamma_{\min}^2 m^2}{\left(m^2+1\right)\left(m^2+(1-\gamma_{\min})^2\right)} \leqslant 1 \\ \implies v_i(b_j) &\geqslant v_i(a), \end{aligned}$$

and therefore the votes of voters in N_2 are consistent with the value matrix $V_{\vec{\gamma},U}$.

By picking budget-feasible set $\{b_1, \ldots, b_{m-1}\}$ we can get a social welfare of $n(m-1)p$, while instead we got $n(1-p)\frac{\gamma_{\min} p}{1-p\gamma_{\min}}$ by choosing a. This leaves us with a distortion of

$$\frac{(m-1)(1-p\gamma_{\min})}{(1-p)\gamma_{\min}}.$$

Since $p \geqslant 0$ and $\gamma_{\min} \leqslant 1$, we know $p \geqslant p\gamma_{\min}$, and so $1-p\gamma_{\min} \geqslant 1-p$. Therefore, we get the desired distortion:

$$\frac{(m-1)(1-p\gamma_{\min})}{(1-p)\gamma_{\min}} \geqslant \frac{m-1}{\gamma_{\min}}.$$

Case 2: If the aggregation rule does not select a, consider the utility matrix U where members of N_1 have utility of 1 for a and 0 for the rest, while members of N_2 have utility of 0 for a and $\frac{\gamma_{\min}(1-p)}{1-\gamma_{\min}(1-p)}$ for the rest of the alternatives. This gives us $\mathsf{sw}(a) = n(1-p)$, and $\mathsf{sw}(b) = np\frac{\gamma_{\min}(1-p)}{1-\gamma_{\min}(1-p)}$ for $b \in A \setminus \{a\}$. Again we have to check that the value matrix $V_{\vec{\gamma},U}$ is consistent with the input profile. For $i \in N_1$ we have: $v_i(a) = 1 - \gamma_{\min} + \gamma_{\min}(1-p) = 1 - \gamma_{\min}p$, and $v_i(b_j) = \gamma_{\min}p\frac{\gamma_{\min}(1-p)}{1-\gamma_{\min}(1-p)}$.

The value matrix is consistent with the ranking of the members of N_1, i.e. $v_i(a) \geqslant v_i(b_j)$, as:

$$\begin{aligned}\gamma_{\min} \leqslant 1 &\implies 0 \leqslant \gamma_{\min}p \leqslant 1 - \gamma_{\min}(1-p)\\ &\implies \gamma_{\min}p\frac{1}{1-\gamma_{\min}(1-p)} \leqslant 1\\ &\implies \gamma_{\min}p\frac{\gamma_{\min}(1-p)}{1-\gamma_{\min}(1-p)} \leqslant 1 - \gamma_{\min}p.\end{aligned}$$

Moreover, for $i \in N_2$ we have: $v_i(a) = \gamma_{\min}(1-p)$, and

$$\begin{aligned}v_i(b_j) &= (1-\gamma_{\min})\frac{\gamma_{\min}(1-p)}{1-\gamma_{\min}(1-p)} + \gamma_{\min}p\frac{\gamma_{\min}(1-p)}{1-\gamma_{\min}(1-p)}\\ &= (1-\gamma_{\min}(1-p))\frac{\gamma_{\min}(1-p)}{1-\gamma_{\min}(1-p)} = \gamma_{\min}(1-p).\end{aligned}$$

So we have $v_i(a) = v_i(b_j)$ which means that the value matrix is consistent with the ranking of the members of N_2 as well.

Since a is not picked by the aggregation rule, we get a maximum social welfare of $(m-1)np\frac{\gamma_{\min}(1-p)}{1-\gamma_{\min}(1-p)}$ when we could have gotten a social welfare of np from a, meaning a distortion of:

$$\mathrm{dist}_{\mathsf{rank}}(f) \geqslant \frac{1-\gamma_{\min}(1-p)}{\gamma_{\min}p(m-1)} \geqslant \frac{m-1}{\gamma_{\min}}.$$

All the conditions above hold for $m \geqslant 2$, so we have a distortions of at least: $\frac{m-1}{\gamma_{\min}}$. □

A.2 Proof of Theorem 3

Theorem 3. *For all deterministic single-winner rules f,*

$$\mathrm{dist}^{\text{single-win}}_{\mathsf{rank}}(f) \in \Omega(\min\{m/\gamma_{\min},\, 1/\gamma^2_{\min}\}).$$

Proof. Suppose we have m alternatives $a_1, \ldots, a_m$ and n voters each with the same PS-value of $\gamma = \gamma_{\min}$. For ease of exposition, let n be divisible by m. Our construction consists of m types of voters, equally distributed with n/m voters of each type. Let N_k be the set of voters of type k. Suppose each voter type votes as follows,

$$\begin{array}{ccccccccccc} N_1 & : & a_1 & \succ & a_2 & \succ & \ldots & \succ & a_{m-1} & \succ & a_m\\ N_2 & : & a_2 & \succ & a_3 & \succ & \ldots & \succ & a_m & \succ & a_1\\ \vdots & & & & & & & & & & \\ N_{m-1} & : & a_{m-1} & \succ & a_m & \succ & \ldots & \succ & a_{m-3} & \succ & a_{m-2}\\ N_m & : & a_m & \succ & a_1 & \succ & \ldots & \succ & a_{m-2} & \succ & a_{m-1} \end{array}$$

so that N_i prefers alternative a_i most, and cycles through the rest.

Without the loss of generality, suppose the voting rule picks a_1. We will set the utilities so that $\mathsf{sw}(a_m) > \mathsf{sw}(a_{m-1}) > \cdots > \mathsf{sw}(a_2) > \mathsf{sw}(a_1)$. To do so, set for all voters i,

$$u_i(a_m) = \begin{cases} 1 & \text{if } i \in N_m\\ 0 & \text{if } i \in N_1\\ u_i(a_1) & \text{otherwise}\end{cases}.$$

For all k from 1 to $m-1$ and for all $i \in N_1$,

$$u_i(a_k) = \frac{\gamma}{1-\gamma}\left(\frac{\mathsf{sw}(a_m) - \mathsf{sw}(a_k)}{n}\right),$$

and for all j from 2 to m, for all $i \in N_j$, for k from 1 to $m-1$, when $k < j-1$:

$$u_i(a_k) = \frac{\gamma}{1-\gamma}\left(\frac{\mathsf{sw}(a_{j-1}) - \mathsf{sw}(a_k)}{n}\right),$$

and when $k \geqslant j$:

$$u_i(a_k) = \frac{\gamma}{1-\gamma}\left(\frac{\mathsf{sw}(a_m) - \mathsf{sw}(a_k)}{n} + \frac{\mathsf{sw}(a_{j-1}) - \mathsf{sw}(a_1)}{n}\right),$$

and $u_i(a_{j-1}) = 0$.

Then, for k from 1 to $m-1$,

$$\begin{aligned}
\mathsf{sw}(a_k) &= \sum_{j=1}^{m}\sum_{i\in N_j} u_i(a_k) \\
&= \frac{\gamma}{1-\gamma}\cdot\frac{1}{n}\Bigg(\sum_{i\in N_1}\Big(\mathsf{sw}(a_m) - \mathsf{sw}(a_k)\Big) + \sum_{j=2}^{k}\sum_{i\in N_j}\Big(\mathsf{sw}(a_m) - \mathsf{sw}(a_k) + \mathsf{sw}(a_{j-1}) - \mathsf{sw}(a_1)\Big) + 0 \\
&\quad + \sum_{j=k+2}^{m}\sum_{i\in N_j}\Big(\mathsf{sw}(a_{j-1}) - \mathsf{sw}(a_k)\Big)\Bigg) \\
&= \frac{\gamma}{1-\gamma}\cdot\frac{1}{n}\cdot\frac{n}{m}\left((k-1)(\mathsf{sw}(a_m) - \mathsf{sw}(a_1)) - (m-1)\mathsf{sw}(a_k) + \sum_{j=1, j\neq k}^{m}\mathsf{sw}(a_j)\right) \\
&= \frac{\gamma}{1-\gamma}\cdot\frac{1}{m}\left((k-1)(\mathsf{sw}(a_m) - \mathsf{sw}(a_1)) - m\cdot\mathsf{sw}(a_k) + \sum_{j=1}^{m}\mathsf{sw}(a_j)\right).
\end{aligned}$$

Let $S = \sum_{j=1}^{m}\mathsf{sw}(a_j)$. Adding $\frac{\gamma}{1-\gamma}\mathsf{sw}(a_k)$ to both sides of the above and rearranging, we get:

$$\mathsf{sw}(a_k) = \frac{\gamma}{m}\left((k-1)(\mathsf{sw}(a_m) - \mathsf{sw}(a_1)) + S\right).$$

In particular, $\mathsf{sw}(a_1) = \frac{\gamma}{m}S$, so

$$\mathsf{sw}(a_k) = \frac{\gamma}{m}\left((k-1)\mathsf{sw}(a_m) + S\cdot\frac{m-(k-1)\gamma}{m}\right).$$

Via the same reasoning,

$$\begin{aligned}
\mathsf{sw}(a_m) &= \sum_{j=1}^{m}\sum_{i\in N_j} u_i(a_m) \\
&= \frac{\gamma}{1-\gamma}\cdot\frac{1}{n}\Bigg(\sum_{j=2}^{m-1}\sum_{i\in N_j}\Big(\mathsf{sw}(a_{j-1})-\mathsf{sw}(a_1)\Big)\Bigg)+\frac{n}{m} \\
&= \frac{\gamma}{1-\gamma}\cdot\frac{1}{m}\Bigg(\sum_{j=2}^{m-1}\Big(\mathsf{sw}(a_{j-1})-\mathsf{sw}(a_1)\Big)\Bigg)+\frac{n}{m} \\
&= \frac{\gamma}{1-\gamma}\cdot\frac{1}{m}\Big(S-(m-2)\mathsf{sw}(a_1)-\mathsf{sw}(a_m)-\mathsf{sw}(a_{m-1})\Big)+\frac{n}{m} \\
&= \frac{\gamma}{1-\gamma}\cdot\frac{1}{m}\Bigg(S-\frac{\gamma(m-2)}{m}S-\mathsf{sw}(a_m)-\frac{\gamma}{m}\bigg((m-2)\mathsf{sw}(a_m)+S\cdot\frac{m-(m-2)\gamma}{m}\bigg)\Bigg)+\frac{n}{m} \\
&= \frac{\gamma}{1-\gamma}\cdot\frac{1}{m}\bigg(\frac{m-(m-2)\gamma}{m}\cdot\frac{m-\gamma}{m}S-\frac{m+\gamma(m-2)}{m}\mathsf{sw}(a_m)\bigg)+\frac{n}{m} \\
&= \frac{\gamma}{1-\gamma}\cdot\frac{1}{m}\bigg(\frac{m-(m-2)\gamma}{m}\cdot\frac{m-\gamma}{m}S\bigg)+\frac{n}{m}-\frac{\gamma(m+\gamma(m-2))}{(1-\gamma)m^2}\mathsf{sw}(a_m).
\end{aligned}$$

Adding $\frac{\gamma(m+\gamma(m-2))}{(1-\gamma)m^2}\mathsf{sw}(a_m)$ to both sides and rearranging:

$$\begin{aligned}
\mathsf{sw}(a_m) &= \frac{(1-\gamma)m^2}{(1-\gamma)m^2+\gamma(m+\gamma(m-2))}\left(\frac{\gamma}{1-\gamma}\cdot\frac{1}{m}\bigg(\frac{m-(m-2)\gamma}{m}\cdot\frac{m-\gamma}{m}S\bigg)+\frac{n}{m}\right) \\
&= \frac{\gamma m}{(1-\gamma)m^2+\gamma(m+\gamma(m-2))}\bigg(\frac{m-(m-2)\gamma}{m}\cdot\frac{m-\gamma}{m}S\bigg)+\frac{(1-\gamma)mn}{(1-\gamma)m^2+\gamma(m+\gamma(m-2))} \\
&= \frac{\gamma(m-(m-2)\gamma)}{(1-\gamma)m^2+\gamma(m+\gamma(m-2))}\cdot\frac{m-\gamma}{m}S+\frac{(1-\gamma)nm}{(1-\gamma)m^2+\gamma(m+\gamma(m-2))}.
\end{aligned}$$

Now, we can finally solve for S:

$$\begin{aligned}
S &= \sum_{k=1}^{m}\mathsf{sw}(a_k) \\
&= \mathsf{sw}(a_m)+\frac{\gamma}{m}\sum_{k=1}^{m-1}\bigg((k-1)\mathsf{sw}(a_m)+S\frac{m-(k-1)\gamma}{m}\bigg) \\
&= \mathsf{sw}(a_m)+\frac{\gamma(m-1)(m-2)}{2m}\mathsf{sw}(a_m)+\frac{\gamma}{m^2}S\sum_{k=1}^{m-1}(m-(k-1)\gamma) \\
&= \frac{2m+\gamma(m-1)(m-2)}{2m}\mathsf{sw}(a_m)+\frac{\gamma}{m^2}S\cdot\frac{(m-1)(2\gamma+m(2-\gamma))}{2} \\
&= \frac{2m+\gamma(m-1)(m-2)}{2m}\bigg(\frac{\gamma(m-(m-2)\gamma)}{(1-\gamma)m^2+\gamma(m+\gamma(m-2))}\cdot\frac{m-\gamma}{m}S+\frac{(1-\gamma)nm}{(1-\gamma)m^2+\gamma(m+\gamma(m-2))}\bigg) \\
&\quad + S\cdot\frac{\gamma(m-1)(2\gamma+m(2-\gamma))}{2m^2}.
\end{aligned}$$

After simplifying this, we get:

$$S = n\frac{2\gamma+\gamma m^2+(2-3\gamma)m}{2(1-\gamma)m^2+2\gamma(\gamma+1)m-4\gamma^2}.$$

This then implies that

$$\mathsf{sw}(a_m) = \frac{n}{m} \cdot \frac{2m^2(1-\gamma) + \left(m(2-3\gamma) + 2\gamma + m^2\gamma\right)\gamma}{2(1-\gamma)m^2 + 2\gamma(\gamma+1)m - 4\gamma^2},$$

and so we ultimately get the following social welfare for each alternative, for k from 1 to $m-1$:

$$\mathsf{sw}(a_k) = \frac{n}{m} \cdot \frac{\gamma\left(2(1-\gamma)km + \gamma\left(m^2 - m + 2\right)\right)}{2(1-\gamma)m^2 + 2\gamma(\gamma+1)m - 4\gamma^2}.$$

The chain of inequalities $\mathsf{sw}(a_m) > \cdots > \mathsf{sw}(a_1)$ does indeed hold, and knowing this, we can verify that the above utilities are non-negative. This gives distortion, after simplifying:

$$\frac{\mathsf{sw}(a_m)}{\mathsf{sw}(a_1)} = 1 + \frac{2(1-\gamma)m^2}{\gamma\left(2\gamma + \gamma m^2 + (2-3\gamma)m\right)}.$$

To show that this is asymptotically as desired, we can write this as:

$$1 + \frac{2(1-\gamma)}{\gamma}\left(\frac{2\gamma + \gamma m^2 + (2-3\gamma)m}{m^2}\right)^{-1}.$$

Since, for any positive a, b, we have that $(a+b)^{-1} \geqslant \frac{1}{2}\min\{a^{-1}, b^{-1}\}$, this expression is in:

$$\Omega\left(1 + \frac{1-\gamma}{\gamma}\min\left\{\frac{m^2}{\gamma(m^2+2)}, \frac{m^2}{m(2-3\gamma)}\right\}\right) = \Omega\left(1 + \frac{1-\gamma}{\gamma}\min\left\{\frac{1}{\gamma}, m\right\}\right),$$

which in the $\gamma \to 0$ regime is asymptotic in $\Omega\left(\frac{\min\{1/\gamma, m\}}{\gamma}\right)$. □

A.3 Proof of Theorem 7

Theorem 7. *For all randomized single-winner rules f,*

$$\mathrm{dist}^{\text{single-win}}_{\mathsf{rank}}(f) \in \Omega\left(\min\left\{m, 1/\gamma_{\min}\right\}\right).$$

Proof. Use the same input profile $\vec{\rho}$ as in the proof of Theorem 3. Let $p(a_i)$ be the probability that a_i is picked by rule f and without the loss of generality, suppose that $a_{\min} = \arg\min_a p(a)$.

Then, for any j, $1 = \sum_i p(a_i) \geqslant p(a_j) + (m-1)p(a_{\min})$, so $p(a_j) \leqslant 1 - (m-1)p(a_j)$

By the proof of Theorem 3, $\mathsf{sw}(a_1) \leqslant \mathsf{sw}(a_2) \leqslant \cdots \leqslant \mathsf{sw}(a_m)$, and so we can maximize social welfare by picking a_m.

The expected social welfare of f is at most:

$$\begin{aligned}
\mathbb{E}_{a\sim f(\vec{\rho})}[\mathsf{sw}(a)] &= \frac{1}{m}\mathsf{sw}(a_m) + \frac{m-1}{m}\max_{k=1}^{m-1}\mathsf{sw}(a_k) \\
&= \frac{n}{m(2(1-\gamma)m^2 + 2\gamma(\gamma+1)m - 4\gamma^2)} \cdot \left(\frac{2m^2(1-\gamma) + \left(m(2-3\gamma) + 2\gamma + m^2\gamma\right)\gamma}{m}\right. \\
&\qquad \left. + \frac{m-1}{m} \cdot \left(\gamma\left(2(1-\gamma)(m-1)m + \gamma\left(m^2 - m + 2\right)\right)\right)\right) \\
&= \frac{n}{m} \cdot \frac{\gamma(\gamma-2)(m-2)(m-1) - 2m}{2((1-\gamma)m + 2\gamma)(m-\gamma)}.
\end{aligned}$$

So, the distortion is:

$$\frac{\mathsf{sw}(a_m)}{\mathbb{E}_{a\sim f(\vec{\rho})}[\mathsf{sw}(a)]} = \frac{n}{m}\cdot\frac{2m^2(1-\gamma)+\big(m(2-3\gamma)+2\gamma+m^2\gamma\big)\gamma}{2(1-\gamma)m^2+2\gamma(\gamma+1)m-4\gamma^2}$$
$$\cdot\left(\frac{n}{m}\cdot\frac{\gamma(\gamma-2)(m-2)(m-1)-2m}{2((1-\gamma)m+2\gamma)(m-\gamma)}\right)^{-1}$$
$$= 1+\frac{2(1-\gamma)(m-1)((1-\gamma)m+2\gamma)}{\gamma(2-\gamma)(m-2)(m-1)+2m}$$
$$\geqslant 1+\frac{2(1-\gamma)^2(m-1)m}{\gamma(2-\gamma)(m-2)(m-1)+2m}.$$

Since, for any positive a, b, we have that $(a+b)^{-1} \geqslant \frac{1}{2}\min\{a^{-1}, b^{-1}\}$:

$$\frac{\mathsf{sw}(a_m)}{\mathbb{E}_{a\sim f(\vec{\rho})}[\mathsf{sw}(a)]} \in \Omega\left((1-\gamma)^2\min\left\{\frac{2(m-1)m}{\gamma(2-\gamma)(m-2)(m-1)}, \frac{2(m-1)m}{2m}\right\}\right)$$
$$\in \Omega\left((1-\gamma)^2\min\left\{\frac{1}{\gamma}, m\right\}\right),$$

which in the $\gamma \to 0$ regime, is $\Omega\left(\min\left\{{}^{1}/_{\gamma}, m\right\}\right)$. □

B Distortion with Unrestricted Utilities and No Public Spirit

In this section, we consider the distortion that can be achieved under various ballot formats without an assumption of public-spirited voters, or equivalently, when $\gamma_i = 0$ for every voter $i \in N$. This serves as a benchmark and motivates the need for cultivating public spirit among voters. It is also interesting to note that without any public spirit, the information in the ballots is useless as rules that ignore the ballots altogether turn out to be worst-case optimal. In contrast, the worst-case optimal rules in the presence of even a little bit of public spirit are both qualitatively and quantitatively fairer.

Proposition 4. *For any ballot format $\mathsf{X} \in \{\mathsf{rank}, \mathsf{vfm}, \mathsf{knap}, \tau\text{-}\mathsf{th}, D\text{-}\mathsf{rth}\}$ (with any threshold τ and threshold distribution D), every deterministic rule has unbounded distortion when $\gamma_i = 0$ for all $i \in N$.*

Proof. First, consider the ballot formats other than randomized threshold approval votes. For deterministic threshold approval votes, pick any threshold $\tau \in [0, 1]$. Let n be even.

Consider an instance as follows. The cost of each alternative is 1, i.e., $c(a) = 1$ for each $a \in A$. Pick any two alternatives $a_1, a_2 \in A$, and let the input profile be as follows. Partition the voters into two equal-sized groups N_1, N_2.

- Under $\mathsf{X} \in \{\mathsf{rank}, \mathsf{vfm}\}$, each voter in N_1 ranks a_1 at the top, a_2 next, and the remaining alternatives afterwards (arbitrarily); and each voter in N_2 ranks a_2 at the top, a_1 next, and the remaining alternatives afterwards (arbitrarily).
- Under $\mathsf{X} \in \{\mathsf{knap}, \tau\text{-}\mathsf{th}\}$ (where $\tau \neq 0$), each voter in N_1 submits $\{a_1\}$ and each voter in N_2 submits $\{a_2\}$.
- Under $\mathsf{X} = \tau\text{-}\mathsf{th}$ with $\tau = 0$, every voter approves all the alternatives.

Fix any of the above ballot formats X and consider any deterministic rule f_{X}. Suppose it picks alternative a. Note that at least one of a_1 and a_2 is not picked by f_{X}. Due to the symmetry, assume without loss of generality that it is a_1. Then, for an arbitrarily chosen $\epsilon \in (0, 1)$, consider the following consistent utility matrix U.

- Each voter in N_1 has utility 1 for a_1 and 0 for all other alternatives.
- Each voter in N_2 has utility ϵ for a_2 and 0 for all other alternatives.

Then, the distortion of f_{X} is at least

$$\frac{\mathsf{sw}(a_1, U)}{\mathsf{sw}(a, U)} = \frac{n/2}{\epsilon \cdot n/2} = \frac{1}{\epsilon}.$$

Because $\epsilon \in (0, 1)$ was chosen arbitrarily, we can take the worst case by letting $\epsilon \to 0$, which establishes unbounded distortion.

For randomized threshold approval votes with any threshold distribution D, we cannot fix the input profile upfront as it depends on the threshold τ sampled from D. However, we can assume that for each τ the rule sees the profile described above for τ-th. The proof continues to work because the consistent utility matrix U described above is independent of the value of τ (and hence, can be set upfront without knowing the value of τ). □

Proposition 5. *For any ballot format* $\mathsf{X} \in \{\mathsf{rank}, \mathsf{vfm}, \mathit{knap}, \tau\text{-}\mathsf{th}, D\text{-}\mathsf{rth}\}$ *(with any threshold τ and threshold distribution D), every randomized rule has distortion at least m when $\gamma_i = 0$ for all $i \in N$ and this is tight.*

Proof. For the upper bound under all ballot formats, it suffices to show that the trivial randomized rule f, which does not take any ballots as input and simply returns a single alternative chosen uniformly at random, achieves distortion at most m. Fix any instance U and let A^* be an optimal budget-feasible set of alternatives. Then, the expected social welfare under f is

$$\frac{1}{m} \sum_{a \in A} \mathsf{sw}(a, U) \geqslant \frac{1}{m} \mathsf{sw}(A^*, U),$$

which implies the desired upper bound of m on the distortion of f.

For the lower bound, we simply extend the argument from the proof of Proposition 4. Define an instance with m alternatives $a_1, a_2, \ldots, a_m$, all with cost 1 (i.e., $c(a_j) = 1$ for all $j \in [m]$). Fix any randomized rule f_{X} for each ballot X in the statement of the proposition.

Let us first consider ballot formats other than randomized threshold approval votes. Consider the following symmetric profiles for each ballot format. Suppose n divides m and voters are partitioned into m equal-size groups $N_1, \ldots, N_m$. Then:

- for $\mathsf{X} \in \{\mathsf{rank}, \mathsf{vfm}\}$, for each $j \in [m]$, every voter in N_j submits the ranking $a_j \succ a_{j+1} \succ \cdots \succ a_m \succ a_1 \succ \cdots \succ a_{j-1}$, and
- for $\mathsf{X} = \{\mathsf{knap}, \tau\text{-}\mathsf{th}\}$ (for any τ), for each $j \in [m]$, every voter in N_j submits the set of alternatives $\{a_j\}$.

For τ-threshold approval votes, there is an edge case where this profile may not be feasible with $\tau = 0$, in which case we can set the profile to have every voter approving all alternatives. The utility matrix defined below would still remain consistent in this case.

For each ballot format X, the corresponding rule must pick at least one alternative with probability $p_{\mathsf{X}} \leqslant 1/m$. Due to the symmetry, we can assume without loss of generality that this alternative is a_1.

Fix any $\epsilon \in (0, 1)$. We define a consistent utility matrix U that works for all of the above ballot formats:

- Every voter in N_1 has utility 1 for a_1 and 0 for all other alternatives.
- For each $j \in \{2, \ldots, m\}$, every voter in N_j has utility ϵ for a_j and 0 for all other alternatives.

Finally, notice that the maximum possible social welfare is $\mathsf{sw}(a_1, U) = 1$, whereas the expected social welfare under the rule f_{X} is $p_{\mathsf{X}} \cdot 1 + (1 - p_{\mathsf{X}}) \cdot \epsilon \leqslant 1/m + (1 - 1/m) \cdot \epsilon$. Thus, the distortion of f_{X} is at least $\frac{1}{1/m + (1-1/m) \cdot \epsilon}$. Since $\epsilon \in (0, 1)$ was chosen arbitrarily, we can take the worst case by letting $\epsilon \to 0$, in which case we get that the distortion must be at least m.

For randomized threshold approval votes with threshold distribution D, we cannot fix the input profile as the input profile depends on the threshold τ sampled from D. However, we can assume that the rule sees the generic input profile described above (where each voter approves only her most favorite alternative) for any $\tau \neq 0$ and the edge-case input profile (where every voter approves all the alternatives). Due to the symmetry, the rest of the argument goes through as the final utility matrix U constructed above is consistent with these input profiles for all τ. $\square$

C Predefined Bundles under Unit-sum Utilities

Ranking predefined bundles is a generalization of ranking by value. When all alternatives have cost equal to the budget, Theorem 20 simply returns the Copeland winner. In the unit-sum model, Copeland must incur $\Omega(m)$ distortion, and so this rule must incur $\Omega(m)$ distortion for unit-sum utilities in the worst case.

In general, we can induce a predefined bundles rule from any single-winner voting rule. If the distortion is d in the single-winner case, our Theorem 20 gives at least d distortion for participatory budgeting. Because all ordinal single-winner voting rules must incur $\Omega(m)$ distortion, this forces our ballot format to incur $\Omega(m)$ distortion in the unit-sum case, regardless of the single-winner rule it is based on.

D The Robustness of Each Voting Rule

In this section, we justify that all upper bounds are robust to variations in the public spirit of voters. All of this stems from the robust Lemma 7. We start by proving a stronger version of Lemma 1 that helps us with the robustness results in this section. Note that Lemma 1 is the special case of Lemma 7 with $c = 0$.

Lemma 7. *(Robust Lemma 1) Let $A_1, A_2 \subseteq A$ be any two subsets of alternatives. Fix any $\alpha \geqslant 0$ and define $N_{A_1 \succ A_2} = \{i \in N : \alpha \cdot v_i(A_1) \geqslant v_i(A_2)\}$. For any $c < 1$, fix an arbitrary subset of voters $N'_{A_1 \succ A_2} \subseteq N_{A_1 \succ A_2}$ of size $\left|N'_{A_1 \succ A_2}\right| = c \cdot |N_{A_1 \succ A_2}|$. Suppose that for all voters $i \in N'_{A_1 \succ A_2}$ public spirit is small with $\gamma_i < \gamma_{\min}$, and for all voters $i \in N_{A_1 \succ A_2} \setminus N'_{A_1 \succ A_2}$ public spirit is large with $\gamma_i \geqslant \gamma_{\min}$. Then:*

$$\frac{\mathsf{sw}(A_2)}{\mathsf{sw}(A_1)} \leqslant \alpha \cdot \left(\frac{1-\gamma_{\min}}{\gamma_{\min}} \frac{n}{|N_{A_1 \succ A_2}|\,(1-c)} + 1 \right).$$

Proof. The proof is the same as the proofs of Lemmas 3.1 and 5.1 by Flanigan et al. [2023]. Let $\tilde{N}_{A_1 \succ A_2} = N_{A_1 \succ A_2} \setminus N'_{A_1 \succ A_2}$. Indeed, for each voter $i \in \tilde{N}_{A_1 \succ A_2}$, we know that $\alpha v_i(A_1) \geqslant v_i(A_2)$, and so:

$$\alpha \left((1-\gamma_i)u_i(A_1) + \gamma_i \frac{\mathsf{sw}(A_1)}{n} \right) \geqslant (1-\gamma_i)u_i(A_2) + \gamma_i \frac{\mathsf{sw}(A_2)}{n} \geqslant \gamma_i \frac{\mathsf{sw}(A_2)}{n}.$$

Dividing by γ_i and using the fact that $\frac{1-\gamma_i}{\gamma_i}$ is decreasing in γ_i we have:

$$\alpha \left(\frac{1-\gamma_{\min}}{\gamma_{\min}} \cdot u_i(A) + \frac{\mathsf{sw}(A_1)}{n} \right) \geqslant \frac{\mathsf{sw}(A_2)}{n}.$$

Summing over all voters in $\tilde{N}_{A_1 \succ A_2}$,

$$\alpha \left(\frac{1-\gamma_{\min}}{\gamma_{\min}} \sum_{i \in \tilde{N}_{A_1 \succ A_2}} u_i(A_1) + \frac{\mathsf{sw}(A_1)\left|\tilde{N}_{A_1 \succ A_2}\right|}{n} \right) \geqslant \frac{\mathsf{sw}(A_2)\left|\tilde{N}_{A_1 \succ A_2}\right|}{n}.$$

Using the fact that $\sum_{i \in \tilde{N}_{A_1 \succ A_2}} u_i(A_1) \leqslant \sum_{i \in N} u_i(A_1) = \mathsf{sw}(A_1)$,

$$\alpha \left(\frac{1-\gamma_{\min}}{\gamma_{\min}} \mathsf{sw}(A_1) + \frac{\mathsf{sw}(A_1)\left|\tilde{N}_{A_1 \succ A_2}\right|}{n} \right) \geqslant \frac{\mathsf{sw}(A_2)\left|\tilde{N}_{A_1 \succ A_2}\right|}{n}.$$

We know that $\left|\tilde{N}_{A_1 \succ A_2}\right| \geqslant (1-c)\left|N_{A_1 \succ A_2}\right|$. So, after some simplification, we finally get the desired upper bound:

$$\frac{\mathsf{sw}(A_2)}{\mathsf{sw}(A_1)} \leqslant \alpha\left(\frac{1-\gamma_{\min}}{\gamma_{\min}} \frac{n}{\left|N_{A_1 \succ A_2}\right|(1-c)}+1\right). \qquad \square$$

The notion of robustness we rely on is as follows.

Definition 3. Let $c \in [0,1]$. An instance is $(\gamma_{\min}, c)$*-robust* when there exists a subset of N' of N such that $|N'| = c \cdot |N|$, all voters $i \notin N'$ have large public spirit with $\gamma_i \geqslant \gamma_{\min}$ and all voters $i \in N'$ have small public spirit with $\gamma_i < \gamma_{\min}$.

Under this notion of robustness, the rule we rely on in proving every upper bounds, Copeland, has the following guarantee proven in Corollary 5.3 of Flanigan et al. [2023].

Proposition 6. *[Flanigan et al., 2023] For* $(\gamma_{\min}, c)$*-robust instances with* $c < 1/2$*,*

$$\operatorname{dist}^{\text{single-win}}_{\mathsf{rank}}(\mathit{Copeland}) = \left(\frac{2\gamma_{\min}^{-1} - (1+2c)}{1-2c}\right)^2 \in \mathcal{O}\left((1-2c)^{-2}\gamma_{\min}^{-2}\right).$$

As a direct corollary of this, iterated Copeland has the same distortion bound, and so committee selection done by repeatedly applying a single-winner voting rule is similarly robust. We need the majority of voters to have a sufficiently large public spirit value, more than 50%, for Copeland to be robust. Predefined bundle rules only rely on the combinatorial structure of the partitioning, which doesn't depend on public spirit values, and on comparisons between bundles, which Lemma 7 shows can be done robustly.

This allows us to rerun the proofs in Section 5, replacing the iterated Copeland bound with its robust formulation. For $(\gamma_{\min}, c)$-robust instances with $c < 1/2$, we then get the following bounds:

- **For high-low bundling**, modifying Theorem 18 appropriately, we get that $\operatorname{dist}_{\mathsf{rank\text{-}b(HLB)}}(\text{Copeland}) = O(\sqrt{m}/((1-2c)\gamma_{\min})^2)$,
- **For tiered-cost bundling**, modifying Theorem 19 appropriately, we get that $\operatorname{dist}_{\mathsf{rank} \to \mathsf{rank\text{-}b(TCB)}}(\text{Copeland}) = O(\log(m)/((1-2c)\gamma_{\min})^4)$,
- **For exhaustive bundling**, modifying Theorem 20 appropriately, we get that $\operatorname{dist}_{\mathsf{rank} \to \mathsf{rank\text{-}b(EB)}}(\text{Copeland}) = O(1/((1-2c)\gamma_{\min})^4)$.

When a constant proportion—larger than half—of the voters have sufficiently large public spirit, the asymptotic behavior of the distortion doesn't change. We can prove similar robustness bounds for each of the other ballot formats, though they may be less robust to large numbers of voters with low public spirit.

For ease of exposition, we first show a Corollary of Lemma 7 that allows any subset of voters to have zero utility.

Corollary 2. *Let* $A_1, A_2 \subseteq A$ *be any two subsets of alternatives. Fix any* $\alpha \geqslant 0$ *and define* $N_{A_1 \succ A_2} = \{i \in N : \alpha \cdot v_i(A_1) \geqslant v_i(A_2)\}$*. For any constant* $c < |N_{A_1 \succ A_2}|/n$*, fix an arbitrary subset of voters* $N' \subseteq N$ *of size* $|N'| \leqslant cn$*. Suppose that for all voters* $i \in N'$ *public spirit is small with* $\gamma_i < \gamma_{\min}$*, and for all voters* $i \in N_{A_1 \succ A_2} \setminus N'$ *public spirit is large with* $\gamma_i \geqslant \gamma_{\min}$*. Then:*

$$\frac{\mathit{sw}(A_2)}{\mathit{sw}(A_1)} \leqslant \alpha \cdot \left(\frac{1-\gamma_{\min}}{\gamma_{\min}} \frac{n}{\left|N_{A_1 \succ A_2}\right| - cn}+1\right).$$

Proof. Let $N'_{A_1 \succ A_2} = N_{A_1 \succ A_2} \cap N'$ be the set of voters in $N_{A_1 \succ A_2}$ with low public spirit. Necessarily,

$\left|N'_{A_1 \succ A_2}\right| \leqslant |N'| \leqslant cn < |N_{A_1 \succ A_2}|$. Define $c' = \left|N'_{A_1 \succ A_2}\right| / |N_{A_1 \succ A_2}| < 1$. By Lemma 7,

$$\begin{aligned}
\frac{\mathsf{sw}(A_2)}{\mathsf{sw}(A_1)} &\leqslant \alpha \cdot \left(\frac{1-\gamma_{\min}}{\gamma_{\min}} \frac{n}{|N_{A_1 \succ A_2}|(1-c')} + 1 \right) \\
&\leqslant \alpha \cdot \left(\frac{1-\gamma_{\min}}{\gamma_{\min}} \frac{n}{|N_{A_1 \succ A_2}|(1-\left|N'_{A_1 \succ A_2}\right| / |N_{A_1 \succ A_2}|)} + 1 \right) \\
&\leqslant \alpha \cdot \left(\frac{1-\gamma_{\min}}{\gamma_{\min}} \frac{n}{|N_{A_1 \succ A_2}|(1-|N'| / |N_{A_1 \succ A_2}|)} + 1 \right) \\
&\leqslant \alpha \cdot \left(\frac{1-\gamma_{\min}}{\gamma_{\min}} \frac{n}{|N_{A_1 \succ A_2}| - cn} + 1 \right),
\end{aligned}$$

as desired. □

As Corollary 2 shows, the number of voters that need to prefer one alternative over another in a voting rule, $N_{A_1 \succ A_2}$, determines how robust it is to voters with low PS. As long as a constant proportion of the voters determining any decision we make are sufficiently public spirited, the distortion bounds will continue to hold.

For $(\gamma_{\min}, c)$-robust instances, we get the following bounds.

- **For deterministic ranking by value**, the Copeland rule is used, so by Proposition 6, $\text{dist}_{\mathsf{rank}}(\textit{Copeland}) \in \mathcal{O}\left(m/((1-2c)\gamma_{\min})^2\right)$ when $c < 1/2$.

- **For stochastic ranking by value**, we first must bound the robust distortion of the maximal lottery rule. Here, the comparisons made require $n/2$ voters to prefer one alternative over another, so we get $\mathcal{O}(1/((1-2c)\gamma_{\min}))$ distortion when $c < 1/2$.

  By the reduction from single-winner rules to randomized participatory budgeting rules in Lemma 3 and Lemma 4, we get the robust distortion bound of $\text{dist}_{\mathsf{rank}}(\textit{Maximal Lottery}) \in \mathcal{O}\left(\log(m)/(1-2c)\gamma_{\min}\right)$ when $c < 1/2$.

- **For stochastic ranking by value per money** By the exact same argument, modifying Theorem 9, it follows that we get a robust upper bound of $\text{dist}_{\mathsf{vfm}}(f) \leqslant \mathcal{O}\left(\log(m)/(1-2c)\gamma_{\min}\right)$ when $c < 1/2$.

- **For one-approvals**, in Proposition 1, a plurality rule is used: an alternative is chosen when at least $1/m$ voters prefer one alternative over another. This gives us a robust upper bound of $\text{dist}_{\mathsf{1\text{-}app}}(f) \in \mathcal{O}\left(\frac{m^2}{(1-cm)\gamma_{\min}}\right)$ when $c < 1/m$.

- **For threshold approvals**, in Theorem 14, plurality is used, netting a robust upper bound of $\text{dist}_{(1/m)\text{-}\mathsf{th}}(f) \in \mathcal{O}\left(\frac{m^2}{(1-cm)\gamma_{\min}}\right)$ when $c < 1/m$.

- **For knapsack**, in Theorem 12, $|N_{A_1 \succ A_2}|$ could be as low as $1/2m^2$. This gives us a robust upper bound of $\text{dist}_{\mathsf{knap}}(f) \in O\left(\frac{m^3}{(1-2m^2c)^2\gamma_{\min}^2}\right)$ when $c < 1/2m^2$.

- **For committee selection knapsack**, in Theorem 13, $|N_{A_1 \succ A_2}|$ could be as low as $2/m$, resulting in distortion $1 + \frac{m}{2} + 2\frac{1-\gamma_{\min}}{\gamma_{\min}(2-mc)}m^2$ when $c < 2/m$.

E Experiments

In this section, we discuss in detail the analysis of the cognitive load on voters for ranking pre-defined bundles.

The database of all participatory budgeting elections taken from the Pabulib database [Faliszewski et al., 2023] on January 2025 is used to perform the analysis in this section. To focus the analysis on participatory

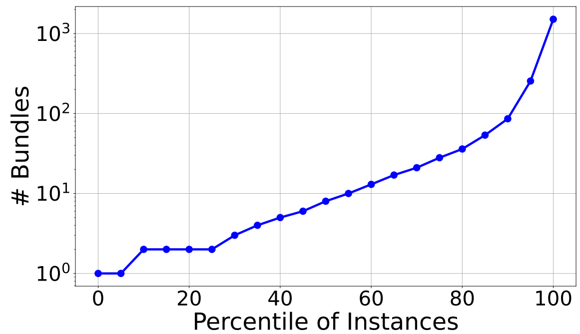

Figure 2: A quantile plot of the number of bundles voters have to rank using the constant distortion two round voting rule given in Section 5.3, on a log scale.

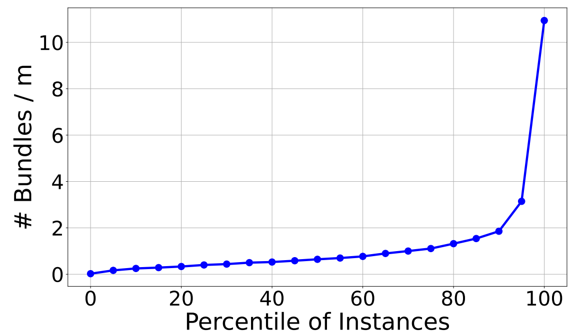

Figure 3: A quantile plot of the number of bundles voters have to rank per alternative using the constant distortion two round voting rule given in Section 5.3.

budgeting elections in practice, this excludes the elections run on Mechanical Turk. All in all, there are 1244 elections in the dataset.

Complete ordinal rankings are required for a constant distortion rule, so when this is not available, we fill in each voter's missing preferences uniformly randomly. To get a full ranking over alternatives when the vote type is cumulative or scoring based, we use the reported alternatives in the order and append the remaining alternatives to the end shuffled uniformly randomly. When the vote type is approval, we also shuffle the order of the reported alternatives uniformly randomly before appending the remaining unreported ones for each voter. Each election is run three times, and the number of bundles remaining is averaged. The seeds used to fill in missing voter preferences uniformly in these experiments are randomly chosen at the beginning.

The experiments were conducted on a system running Windows 11, with 32 GB of RAM, and on a 13th Gen Intel(R) Core(TM) i9-13900H @ 2600 Mhz CPU.

The number of bundles voters rank in the second round of voting is shown in Figure 2 and Figure 3. While already small the majority of the time, there are elections where the number of bundles voters rank are up to 10 times more than the number of bundles. There is, however, redundancy in real world elections we can exploit here that allows us to prune bundles we know are dominated from voter preferences before the second round of voting. This can be done while maintaining constant distortion.

Proposition 7. *Consider the total ordering $\succ_C$ of the alternatives given by the iterated Copeland rule, and two bundles $B_1, B_2 \subseteq A$. If there exists a one-to-one function $f : B_1 \to B_2$ such that $f(b) \succ_C b$ for all $b \in B_1$, then*

$$\frac{\mathsf{sw}(B_1)}{\mathsf{sw}(B_2)} \leqslant 2\gamma_{\min}^{-1} - 1. \tag{4}$$

Proof. When $a \succ_C b$, by the distortion of Copeland given in Theorem 3.3 of Flanigan et al. [2023], $\mathsf{sw}(b) \leqslant (2\gamma_{\min}^{-1} - 1)\mathsf{sw}(a)$. Therefore, for each $b \in B_1$, we know that $\mathsf{sw}(f(b)) \leqslant (2\gamma_{\min}^{-1} - 1)\mathsf{sw}(b)$. Summing over all $b \in B_1$, we get that $\mathsf{sw}(B_1) \leqslant (2\gamma_{\min}^{-1} - 1)\sum_{b\in B_1}\mathsf{sw}(b)$. Because f is one-to-one, we know that $\sum_{b\in B_1}\mathsf{sw}(f(b)) \leqslant \mathsf{sw}(B_2)$, which finally implies

$$\frac{\mathsf{sw}(B_1)}{\mathsf{sw}(B_2)} \leqslant 2\gamma_{\min}^{-1} - 1. \qquad \square$$

Proposition 7 allows us to prune bundles before the second round of voting with a constant $1/\gamma_{\min}^2$ factor increase in distortion in the worst case. Because Copeland gives us a total ordering, this induces a partial ordering over the bundles, the maximums of which in our preselected bundles we ask voters to rank in the second round of voting.

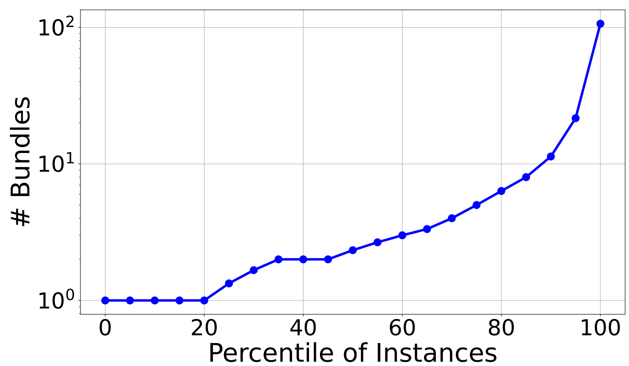

Figure 4: A quantile plot over 1244 real-world PB instances showing the number of bundles voters have to rank after pruning dominated bundles, on a log scale.

The number of bundles voters rank in the second round of voting after pruning dominated bundles is shown in Figure 4 and Figure 1