The Distortion of Public Spirited Participatory Budgeting

MARK BEDAYWI, BAILEY FLANIGAN, MOHAMAD LATIFIAN, and NISARG SHAH

Governments worldwide are increasingly turning to participatory budgeting (PB) as a tool for democratically allocating limited budgets to public-good projects. In PB, constituents vote on their preferred projects from a provided list via specially-designed 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 ballot formats and aggregation rules that yield outcomes with low *distortion* (informally, those with high social welfare). Existing bounds, however, rely on strong assumptions that restrict voters' latent utilities. We prove that low distortion PB outcomes can be achieved without any assumptions on voters' utilities by leveraging the established idea that voters can be *public-spirited*: they may consider others' interests alongside their own when when voting.

Flanigan et al. [2023] formally introduce the framework of distortion under public-spirited *single-winner* voting, an important setting which can be viewed as the special case of PB where only a single project can be funded. Before moving on to PB, we completely close the gaps in their results to derive tight bounds on the optimal deterministic and randomized rules for this case.

Moving on to PB, we study various common ballot formats, such as rankings by value, rankings by value for money, k-approvals, knapsack votes, and threshold approval votes. We prove that rankings by value easily permit achieving distortion linear in the number of projects, and unfortunately, none of the other common ballot formats can break that barrier. Our main contribution is to design a novel and highly practical ballot format for PB which, we prove, allows achieving sublinear (and even logarithmic, if voting over two rounds is possible) distortion in the number of projects. These distortion bounds are significantly lower than those without public-spirited voting (even under restricted utilities), highlighting the potential of *democratic deliberation* — a practice believed to cultivate public spirit, and which is commonplace in real-world PB — to enable higher-welfare outcomes in PB elections.

Manuscript submitted for review to the 25th ACM Conference on Economics & Computation (EC'24).

1 INTRODUCTION

Governments at all scales regularly face the question: *With a limited budget, which public-good projects — e.g., building bike paths or installing streetlamps — should they fund?* To make such decisions democratically, governments are increasingly using *participatory budgeting* (PB), in which constituents vote on which projects they would like to see funded. In PB, the government supplies a budget B and a list of m potential projects $a \in \{1, \dots, m\}$ with corresponding costs $c_1, \dots, c_m$. Voters submit their preferences via *ballots*, and then these ballots are aggregated via an *aggregation rule* to select a set of projects to be funded, whose total cost must be at most B. PB is now used all over the world to decide allocations of public funds[1] [De Vries et al., 2022, Participedia, 2023, Wampler et al., 2021].

When designing the PB process described above, 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 combined. In using this measurement, 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 their 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 $\text{sw}(S) = \sum_{i \in N} u_i(S)$.

If voters' utilities were observable, 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*. For example, popular ballot formats in PB include *rankings by value*, where voters are asked to rank the individual projects, or k*-approval votes*, where voters are asked to approve their favorite k alternatives. It is not hard to see that such ballot formats lose far too much information about voters' utilities to allow deterministic selection of a high-welfare solution: suppose there are two projects, a and b, both costing B so we must simply choose one or the other to fund. If half the population has utilities $1, 0$ for a, b and the other half has utilities $0, 1000$ for a, b (so the welfare of b is 1000 times that of a). Although b has far higher social welfare, any ordinal ballot format where voters only compare sets of alternatives will produce symmetric ballots, leading to any deterministic aggregation rule—i.e., any deterministic mapping from n ballots to an allocation funds—to choose (without loss of generality) a; the best thing we can do here is to randomize uniformly over the two options.

This example illustrates a prohibitive impossibility: in the worst case, *any* deterministic aggregation rule over any ordinal PB ballot format will select an outcome with arbitrarily sub-optimal social welfare, simply because these PB ballot formats do not contain enough information about voters' cardinal preferences. Formally, this sub-optimality is captured with the *distortion*: the worst-case (over possible latent utilities) ratio of the best possible social welfare that of the outcome. Existing work sidesteps this impossibility by assuming that each voter's utilities are restricted to add up to 1 [Benadè et al., 2021]. Although this permits bounded distortion in theory, it remains unclear whether these bounds apply in practice: For example, this assumption may not hold in the likely case that the public goods will more greatly impact 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 behavior would be reliably present in the wild, as Flanigan *et al* point out, research suggests that public spirit can be cultivated via democratic deliberation — *a practice that is already commonplace in PB elections* [De Vries et al., 2022, Participedia, 2023]. The possibility of cultivating public spirit among PB participants motivates our main research question:

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

Question: *If voters are public-spirited, do there exist ballot formats and associated aggregation rules that achieve small distortion, without any restrictions on voters' utilities?*

An affirmative answer to this question would suggest a practicable approach — democratic deliberation — to achieving higher-welfare outcomes in PB elections. In the process of pursuing this question, we close an open question for the single-winner voting setting left open by Flanigan et al. [2023], and introduce a new ballot format which makes better use of voters' public spirit to break an important distortion barrier in the PB context. We overview these contributions below.

1.1 Results and contributions.

We study the distortion of PB with public-spirited participants by adopting Flanigan et al. [2023]'s model of public-spirited voting, extending it as needed to new ballot formats. In this model, each voter i evaluates each alternative a not just according to her her own utility $u_i(a)$, but by her *public-spirited (PS) value*: the convex combination of her utility for a and its social welfare. This convex combination is weighted by her *public spirit level* $\gamma_i \in [0, 1]$, where higher γ_i means she more strongly weighs the social welfare. As in Flanigan et al. [2023], our distortion bounds, summarized in Tables 1 and 2, are parameterized by $\gamma_{\min} = \min_i \gamma_i$, the minimum public spirit level of any voter.

Contribution 1: Tight bounds for single-winner voting with *ranking by value* ballots. Building directly from Flanigan et al. [2023], we begin by studying *ranking-by-value* ballots, where voters rank the alternatives in $[m]$ in decreasing order of their public-spirited values. Before analyzing the performance of this ballot format in the PB context, we first study it in the single-winner context — a significant strict restriction of the PB setting where all projects cost B. We begin with the single-winner setting because, although this is precisely the setting studied by Flanigan *et al*, there remain two important open questions, which we close in order to build upon their answers later.

1.1 What is the best distortion achievable by any ***deterministic*** *voting rule over ranking-by-value ballots?* The lowest-distortion voting rule identified by Flanigan *et al* is Copeland, achieving constant (in m) distortion of exactly $(1 + {}^{2(1-\gamma_{\min})}/_{\gamma_{\min}})^2$; in contrast, their results lead to a lower bound on any deterministic rule of at most $1 + {}^{2(1-\gamma_{\min})}/_{\gamma_{\min}}$. To close this gap, we design a nontrivial construction to prove a stronger lower bound. This lower bound is tight to known upper bounds in its dependency on both m and $\gamma_{\min}$, thereby closing the question of what level of distortion is possible in single-winner public-spirited deterministic voting. This analysis reveals that in fact, the rule Copeland is optimal (except when m is small relative to $1/\gamma_{\min}$, in which case Plurality is optimal).

1.2 What is the best distortion achievable by any ***randomized*** *voting rule over ranking-by-value ballots?* Flanigan et al. [2023] did not study randomized voting rules at all. Thus, here we must prove lower and upper bounds anew. Our lower bound arises from the same construction as described above. We identify a novel optimal voting rule for this case, whose distortion matches our lower bound in both m and $\gamma_{\min}$. Its distortion is $\Theta(\min\{m, 1/\gamma_{\min}\})$, the best distortion possible in single-winner public-spirited randomized voting.

Contribution 2: Distortion bounds for PB with *ranking-by-value* ballots. Next, we generalize our results from the single-winner to the PB setting, again with the goal of identifying optimal aggregation rules and proving matching lower bounds.

2.1 Lower bounds. First, we extend our lower bounds from the single-winner case to prove that in PB, the distortion of any deterministic rule must be in $\Omega(m/\gamma_{\min})$, and that of any randomized rule must be in $\Omega(\log m)$.

2.2 Upper bounds via reductions from single-winner voting to PB. For both deterministic and randomized rules, we prove our upper bounds via direct reductions relating any voting rule's distortion in the single-winner setting to its performance in the PB setting. Such a reduction was previously known for deterministic rules, incurring a factor of at most m in the distortion from single-winner to PB. Via this method, we find that Copeland is again optimal as before, with distortion matching our lower bound in both its dependency on m and $\gamma_{\min}$.

For randomized rules, no such reduction existed, so we extend the previous reduction to the randomized case. Via this reduction, we incur a factor of order at most $\log m$ in the distortion from single-winner to PB. Then, we apply this reduction to give an upper bound on our single-winner randomized rule above. In the PB setting, this voting rule achieves distortion with optimal dependency on m, and within a factor of at most $\gamma_{\min}$ of optimal dependency on $\gamma_{\min}$.

Contribution 3: Approval-style ballot formats. A practically important type of ballot format in the PB context are k-approval ballots. In our model, this means voters submit the set of k alternatives for which they have the highest public-spirited values. Due to their practical importance, we now repeat our analysis for this entirely new ballot format. Our first key finding is that if k is larger than one or more maximal budget-feasible sets of projects, the distortion can be unbounded because voters' approval sets can be budget-infeasible, thus giving us no information about their preferences over budget-feasible sets. This is clearly avoided when $k = 1$; accordingly, we give matching lower and upper bounds on the distortion of 1-approval ballots of $\Theta(m^2/\gamma_{\min})$

The issue of k-approval ballots permitting budget-infeasible approval sets motivates another ballot format often considered in the PB literature — *knapsack* ballots. Knapsack ballots again allow each voter to approve a set of items, but *only if that set is budget-feasible.* Perhaps the most striking finding in our analysis of knapsack ballots is that while they have at best exponential distortion $\Omega(2^m/\sqrt{m})$ under the unit sum utilities assumption, we show via a novel approach of comparing entire subsets of alternatives that under public-spirited voting, these ballots have polynomial distortion of at most order $O(m^3)$.

Contribution 4: Ballot formats that breaks the *m* distortion barrier. In the previous sections, our lower bounds show us that across the ballot formats we study — plus two others whose analysis we relegate to the appendix — *no ballot format can achieve distortion with sublinear dependency on m with deterministic aggregation rules (which is the practical case of interest).* This barrier also exists under the unit-sum utilities assumption [Benadè et al., 2021]. Motivated by this, we ask: *is public spirit powerful enough to permit a* ***any*** *practical ballot format to break this barrier?*

We find that in fact, the answer is yes. We define a new, simple ballot format, which pre-partitions the alternatives into (at most m) feasible sets of alternatives, and requires voters to rank them rather than the individual alternatives. We show that by carefully bundling the alternatives in the ballot, we can get $O(\sqrt{m}/\gamma_{\min}^2)$ distortion. If a second stage of elicitation is allowed, we show that the distortion can be further reduced to $O(\log m/\gamma_{\min}^4)$ using this ballot format. These results show that our new ballot format is significantly more efficient, while also being thrifty and practical. These results also point to the exciting open question of whether any ordinal ballot that only asks voters to compare polynomially many sets of alternatives can reduce the distortion all the way down to a constant.

1.2 Related work

Our work directly builds on the works of Benadè et al. [2021], who analyzed distortion in PB, and Flanigan et al. [2023], who introduced the public-spirit model. Our results eliminate the unit-sum assumption made in the former work, and generalize the latter work from single-winner elections

		Public-Spirit	Unit-Sum
SW	Deterministic	$\Theta\left({}^{1}/_{\gamma_{\min}} \cdot \min\{m, {}^{1}/_{\gamma_{\min}}\}\right)$	$\Theta(m^2)$
	Randomized	$\Theta\left(\min\{m, {}^{1}/_{\gamma_{\min}}\}\right)$	$\Theta(\sqrt{m})$
PB	Deterministic	$\Omega\left({}^{m}/_{\gamma_{\min}}\right), O\left({}^{m}/_{\gamma_{\min}} \cdot \min\{m, {}^{1}/_{\gamma_{\min}}\}\right)$	$\Theta(m^2)$
	Randomized	$\Omega\left(\log m\right), O\left(\min\{m, {}^{(\log m)}/_{\gamma_{\min}}\}\right)$	$\Omega(\sqrt{m}), O\left(\sqrt{m}\log m\right)$

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

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

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

(selecting a single alternative) to the more general problem of PB, where multiple alternatives are selected subject to a budget constraint and there are multiple reasonable ballot formats to consider.

Procaccia and Rosenschein [2006] introduce the distortion framework in single-winner elections under the unit-sum assumption. We now know that the best distortions achievable by deterministic and randomized rules for this special case are $\Theta(m^2)$[Caragiannis et al., 2017, Caragiannis and Procaccia, 2011] and $\Theta(\sqrt{m})$ [Boutilier et al., 2015, Ebadian et al., 2022], respectively. Optimal distortion bounds have also been identified for k-committee selection [Borodin et al., 2022, Caragiannis et al., 2017], which still remains a special case of PB. As an alternative to the unit-sum assumption, unit-range utilities or metric costs have been studied [Anshelevich et al., 2018, Filos-Ratsikas and Miltersen, 2014], but all of these place some restriction on voter preferences. For further details, we suggest 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 Baumeister et al. [2020], Rey et al. [2020], Talmon and Faliszewski [2019]. Another important consideration in PB is whether the allocation of funds is fair with respect to (groups of) voters [Brill et al., 2023, Fain et al., 2018, Peters et al., 2021]. For further details, we suggest the survey of Rey and Maly [2023] and the book chapter of Aziz and Shah [2021].

2 PRELIMINARIES

We introduce the most general framework of participatory budgeting (PB) first, and later introduce single-winner and multiwinner voting as its special cases.

There is a set N of n *voters* and a set A of m *alternatives* (projects). 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, we use $c(S) = \sum_{a \in S} c_a$ as the total cost of alternatives in S. Let $\mathcal{F} = \{S \subseteq A \colon c(S) \leqslant B\}$ be the set of *budget-feasible* subsets of alternatives. The goal is to select such a budget-feasible subset by eliciting and aggregating voter preferences.

Special cases. We note that *k-committee selection* is a special case of PB, where the cost of each alternative is $1/k$, so $\mathcal{F}$ consists of all subsets of alternatives of size k. We use "k-committee rule" to refer to a rule for this special case. Further, *single-winner selection* is a special case of k-committee selection where $k = 1$; we use "single-winner rule" to refer to a rule for this special case.

Utilities. 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 $U \in \mathbb{R}_{\geqslant 0}^{n \times m}$. Define the *social welfare* of an alternative $a \in A$ w.r.t. utility matrix U as $\text{sw}(a, U) = \sum_{i \in N} u_i(a)$; for a subset of alternatives $S \subseteq A$, define $\text{sw}(S, U) = \sum_{a \in S} \text{sw}(a, U)$. We use $\text{sw}(a)$ or $\text{sw}(S)$ when U is clear from context.

PS-values. Following the model introduced by 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 $\vec{\gamma} \in [0,1]^n$. Our results depend on the minimum public spirit level of the voters $\gamma_{\min} \triangleq \min_{i \in N} \gamma_i$.

Each voter submits her preferences according to not her personal utilities, but her *PS-values*, which she computes by taking a γ_i-weighted convex combination of her personal utilities and the average utility of all voters. Formally, the *PS-value* of voter i for alternative a is

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

Together, these PS-values form the *PS-value matrix* $V_{\vec{\gamma},U} \in \mathbb{R}_{\geqslant 0}^{n \times m}$. PS-values are additive across alternatives, so that for each $S \subseteq A$, $v_i(S) = \sum_{a \in S} v_i(a)$.

Note that PS-values have the same scale as utilities because $\text{sw}(a) = \sum_{i \in N} u_i(a) = \sum_{i \in N} v_i(a)$ for each $a \in A$. We show that this transformation allows us to get rid of the unit-sum assumption ($\sum_{i \in N} u_i(a) = 1, \forall a \in A$) required by much of the prior work [Benadè et al., 2021].

Elicitation. Since it is cognitively burdensome for voters to report numeric PS-values, it is common to elicit their preferences using discrete ballots. Following the model of Benadè et al. [2021], a *ballot format* $\mathsf{X} : \mathbb{R}_{\geqslant 0}^{m} \times [0,1]^m \to \mathcal{L}_{\mathsf{X}}$ 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.

We study four ballot formats also studied by Benadè et al. [2021], namely rankings by value, rankings by value for money, knapsack votes, and threshold approval votes, as well as a new ballot format we introduce, namely ranking of predefined bundles; we define them in their respective sections.

Aggregation Rules. Let $\Delta(\mathcal{F})$ be the set of all distributions over $\mathcal{F}$. A (randomized) *aggregation rule* $f : \mathcal{L}_{\mathsf{X}}^n \times [0,1]^m \to \Delta(\mathcal{F})$ for ballot format X takes an input profile $\vec{\rho} \in \mathcal{L}_{\mathsf{X}}^n$ and a cost function over alternatives $c \in [0,1]^m$ as input, and outputs a distribution over feasible sets of alternatives in $\mathcal{F}$. We say that f is deterministic if its output always has singleton support.

Distortion. The *distortion* measures the efficiency of a voting system, composed of a ballot format and an aggregation rule for that ballot format. For a ballot format X and minimum public spirit level $\gamma_{\min} \in [0,1]$, the distortion of an aggregation rule f on input profile $\vec{\rho}$ in format X and cost function c is the following worst-case ratio:

$$\mathrm{dist}_{\mathsf{X}}(f, \vec{\rho}, c) = \sup_{\substack{U, \vec{\gamma}: \\ \min_{i \in N} \gamma_i = \gamma_{\min}, \\ V_{\vec{\gamma}, U} \triangleright \vec{\rho}}} \frac{\max_{S \in \mathcal{F}} \mathrm{sw}(S, U)}{\mathbb{E}_{S' \sim f(\vec{\rho})} \mathrm{sw}(S', U)}.$$

The (overall) distortion of f is obtained by taking the worst case over all instances $(\vec{\rho}, c)$ and all n:

$$\mathrm{dist}_{\mathsf{X}}(f) = \sup_{n \geqslant 1} \sup_{\vec{\rho} \in \mathcal{L}_{\mathsf{X}}^n,\ c \in [0,1]^m} \mathrm{dist}_{\mathsf{X}}(f, \vec{\rho}, c).$$

The resulting distortion is a function of m and $\gamma_{\min}$; we fix arbitrary $m \geqslant 2$ and $\gamma_{\min} \in (0, 1]$ throughout the paper. We are interested in the lowest distortion enabled by each ballot format, across all aggregation rules for that ballot format. This is a measure of the usefulness of the information contained in the ballot format for social welfare maximization.

Supporting results. Let us state a lemma that we use throughout the paper. This is a simple generalization of Lemma 3.1 of Flanigan et al. [2023]; the proof is in Appendix C.1.

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

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

Finally, for comparison, we remark that for all ballot formats we consider, when there is no public spirit and the utilities are unrestricted, all deterministic voting rules have unbounded distortion and the randomized rules have at best m distortion (Appendix C.2).

3 SINGLE-WINNER VOTING

As mentioned before, single-winner voting can be seen as a special case of participatory budgeting problem in which all the alternatives have a cost equal to the budget, so only a single alternative can be selected. Flanigan et al. [2023] analyze the distortion of various deterministic voting rules for this single-winner case under public-spirited voting. In this section we give lower bounds on the distortion of any deterministic and randomized voting rule in this setting, and also design rules that match the lower bound. For the results in this section, we consider, as do Flanigan et al. [2023], the prominent ballot format of *rankings by value* (rbv). In this ballot format, each voter ranks the alternatives in a non-increasing order of her values for them. Formally, $\mathcal{L}_{\mathrm{rbv}}$ is the set of all rankings of the alternatives, and each voter i submits a ranking $\rho_i \in \mathcal{L}_{\mathrm{rbv}}$ such that for every $a, b \in A$ with $v_i(a) > v_i(b)$, we have $a \succ_{\rho_i} b$ (i.e., a appears above b in the ranking ρ_i); the voter can break ties among equal-PS-valued alternatives arbitrarily.

3.1 Lower bounds

We start by proving the lower bound for the *deterministic* rules.

Theorem 1 (Lower Bound - Deterministic). *Any deterministic single-winner voting rules f with ranked preferences has distortion*

$$\mathrm{dist}_{rbv}(f) \geqslant 1 + 2\frac{1 - \gamma_{\min}}{\gamma_{\min}} \cdot \frac{m^2}{2\gamma_{\min} + \gamma_{\min} m^2 + (2 - 3\gamma_{\min})m} \in \Omega\left(\frac{1}{\gamma_{\min}} \cdot \min\left\{ m, \frac{1}{\gamma_{\min}} \right\} \right).$$

Proof Sketch. 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,

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

so that N_i prefers alternative a_i most, and cycles through the rest. We use this instance to prove the lower bound. □

The full proof can be found in Appendix D.1. We include the instance that gives this lower bound here, because versions of it will be used to prove lower bounds throughout the paper. Using a similar instance, we can prove a lower bound on the distortion of any *randomized* voting rule. The full proof of this theorem is in Appendix D.2.

Theorem 2 (Lower Bound - Randomized). *Any randomized single-winner voting rules f with ranked preferences has distortion*

$$\text{dist}_{rbv}(f) \in \Omega\left(\min\left\{m, \frac{1}{\gamma_{\min}}\right\}\right).$$

3.2 Upper Bounds

In this section we focus on designing voting rules with distortion matching the lower bounds. First, in the deterministic case, we give a deterministic voting rule that directly combines upper bounds from Flanigan et al. [2023].

Corollary 1 (Upper Bound - Deterministic). *The deterministic single-winner rule f_{PC} that runs* Plurality *if $m \leqslant 1/\gamma_{\min}$ and* Copeland *otherwise, has distortion at most*

$$\text{dist}_{\text{rbv}}(f_{PC}) \leqslant \min\left\{\frac{m}{\gamma_{\min}} - m, \left(\frac{2}{\gamma_{\min}} - 1\right)^2\right\} \in O\left(\frac{1}{\gamma_{\min}} \cdot \min\left\{m, \frac{1}{\gamma_{\min}}\right\}\right).$$

Proof. Per Proposition 3.5 and Theorem 3.3 of Flanigan et al. [2023] respectively, $\text{dist}_{\text{rbv}}(f_{\text{Plurality}}) \leqslant m/\gamma_{\min} - m$ and that of $\text{dist}_{\text{rbv}}(f_{\text{Copeland}}) \leqslant (2/\gamma_{\min} - 1)^2$. Thus, by defining the rule that chooses the Plurality winner when $m \leqslant 1/\gamma_{\min}$ and the Copeland winner otherwise, we can guarantee achievement of the desired distortion. □

Now, we endeavor to find an optimal randomized voting rule. Since Flanigan et al. [2023] does not study randomized rules, we cannot apply their bounds. Here, we turn to *maximal lottery*, a randomized voting rule that was originally proposed by Kreweras [1965] and rediscovered numerous times in the social choice literature [Fishburn, 1984, Fisher and Ryan, 1995, Laffond et al., 1993, Rivest and Shen, 2010]. Curiously, Charikar et al. [2024] recently use this rule to derive a breakthrough result in the related setting of metric distortion. There are various alternative formulations of this rule, but the one most useful to us is the following.

Definition 1 (Maximal Lottery). *Define the domination graph to be a directed graph G with alternatives in A as the vertices and an edge between every pair of vertices, oriented so that if a beats b in a pairwise election, then the edge goes from a to b. In the case of ties, we may pick orientation arbitrarily. The maximal lottery rule returns a distribution p over the vertices such that for any vertex $v \in A$, the probability of picking v or a vertex adjacent to v is at least $1/2$. The existence of such a distribution can be inferred from, e.g., Farkas' lemma (see Theorem 2.4 of Harutyunyan et al. [2017]).*

Theorem 3 (Upper Bound - Randomized). *There exists a randomized single-winner voting rule f with distortion at most*

$$\mathrm{dist}_{\mathrm{rbv}}(f) \leqslant \min\{m, 2(2/\gamma_{\min} - 1)\} \in O\left(\min\left\{m, \frac{1}{\gamma_{\min}}\right\}\right).$$

Proof. To match our piecewise lower bound, we must again decide between two voting rules: the voting rule which chooses an alternative uniformly at random (thereby achieving m distortion) and the maximal lottery rule, which we prove has distortion at most $2/\gamma_{\min} - 1$.

Indeed, 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:

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

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:

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

Importantly, because $\gamma_{\min}$ is unobservable to the voting rule, *implementing* these piecewise voting rules (for both the randomized and deterministic cases) is not quite practicable, ut the intuition — that for small m, Plurality is desirable, and for large m, Copeland is better — is.

4 RANKINGS BY VALUE

We now move on to the more general setting of participatory budgeting (PB). To begin with, we examine how powerful the same rankings by value ballot format is for PB. Note that while voters still rank individual alternatives by value, the fact that a (feasible) set of alternatives can be funded can significantly affect the power of this ballot format.

4.1 Deterministic Rules

First, we show that for rbv ballots, deterministic rules must incur a distortion at least $(m-1)\gamma_{\min}^{-1}$. The intuition for this bound is as follows: PB is easy when cheap alternatives are always ranked higher than costly ones, there is never any reason to pick the costly alternatives. So, to construct hard instances, have voters rank costly alternatives highly.

Theorem 4 (lower bound). *For rankings by value, every deterministic rule f has distortion*

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

Now, we show how to build directly on results from the single-winner case to give optimal rules for the much more general setting of PB. Specifically, to prove upper bounds, in both the deterministic case and the randomized case, we show how to construct a PB rule from any deterministic single-winner rule while losing an only a factor of m on the distortion.

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

Proof. Fix any instance and let f return the singleton set $\{a\}$. Let A^* be an optimal budget-feasible set. Then,

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

We now use this lemma to translate known results from the single-winner setting to PB. In single winner elections, Flanigan et al. [2023] show that Plurality has distortion at most $m(\gamma_{\min}^{-1} - 1) + 1$ and Copeland's rule has distortion at most $\left(2\gamma_{\min}^{-1} - 1\right)^2$. Plugging these bounds into Lemma 2, we conclude upper bounds for the PB setting:

Theorem 5 (upper bound). *For rankings by value,*

$$\mathrm{dist}_{\mathrm{rbv}}(f_{\mathrm{Plurality}}) \leqslant m^2(\gamma_{\min}^{-1} - 1) + m, \text{ and}$$
$$\mathrm{dist}_{\mathrm{rbv}}(f_{\mathrm{Copeland}}) \leqslant m\left(2\gamma_{\min}^{-1} - 1\right)^2.$$

Hence, there exists a deterministic rule f with distortion

$$\mathrm{dist}_{\mathrm{rbv}}(f) \in O\left(\frac{m}{\gamma_{\min}} \cdot \min\left\{m, \frac{1}{\gamma_{\min}}\right\}\right).$$

Remark 1. *Note that there remains a gap between our upper and lower bounds (in Theorem 5 and Theorem 4, respectively): Plurality achieves the optimal dependence on $\gamma_{\min}$, Copeland achieves the optimal dependence on m, but neither achieves both. Also, the "best" rule in Theorem 5 is again a piecewise rule that depends on $\gamma_{\min}$ to decide which of plurality and Copeland to execute. However, it is unclear if a $\gamma_{\min}$-agnostic rule can achieve the same (or even a better) distortion bound.*

4.2 Randomized Rules

Theorem 6 (upper bound). *For rankings by value, there exists a randomized rule f with distortion*

$$\mathrm{dist}_{\mathrm{rbv}}(f) \leqslant 4\left(\frac{2}{\gamma_{\min}} - 1\right) \cdot \left(\lceil \log_2(m) \rceil + 1\right) \in O\left(\frac{\log(m)}{\gamma_{\min}}\right).$$

To prove this bound, we will derive another general-purpose reduction — this time for randomized rules — from PB to k-committee selection (Lemma 3), and then from k-committee selection to single-winner selection (Lemma 4). The first will suffers $O(\log m)$ overhead; the latter suffers none (asymptotically). To apply this reduction, we want to plug in bounds on randomized single-winner rules; unfortunately, no such results exist in the public spirit model.

In response, we give in Theorem 3 a novel randomized single-winner rule with asymptotically optimal (in both m and $\gamma_{\min}$) distortion of at most $4\gamma_{\min}^{-1} - 2$. We now state and prove these results in succession, before applying them to prove Theorem 6.

Lemma 3 (Committee → PB - Randomized)**.** *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 for rankings by value with distortion at most $2d \cdot (\lceil \log_2(m) \rceil + 1)$.*

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 $j \in \left\{0, 1, \ldots, \lceil \log_2(m) \rceil\right\}$ 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 $\mathrm{sw}(A^* \cap A_j) \leqslant 2\mathrm{sw}(L_j^*)$ for any $j \neq 0$.

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

$$\mathbb{E}_{L\sim\delta}[\mathrm{sw}(L)] = \frac{1}{\lceil \log_2(m) \rceil + 1} \sum_{j=0}^{\lceil \log_2(m) \rceil} \mathrm{sw}(L_j)$$

$$\geqslant \frac{1}{\lceil \log_2(m) \rceil + 1} \sum_{j=0}^{\lceil \log_2(m) \rceil} \frac{\mathrm{sw}(A^* \cap A_j)}{2d} \geqslant \frac{\mathrm{sw}(A^*)}{2d(\lceil \log_2(m) \rceil + 1)}. \qquad \square$$

Next, we reduce k-committee selection to single-winner selection without any asymptotic overhead. The idea is to simply add an alternative to the committee using the single-winner randomized rule, then remove the selected alternative, and repeat the procedure k times.

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

The deterministic case is proved in Theorem 8 of Goel et al. [2018]. Their key idea is to repeatedly pick alternatives using the single winner rule k times. We extend their result to the randomized case using the same argument. We include the proof in Appendix E.2.

Having reduced the PB problem to that of single-winner selection, we now use the novel randomized single-winner rule presented in Theorem 3 to prove the desired bound.

Proof of Theorem 6. Finally, we apply Lemmas 3 and 4 and theorem 3 to prove Theorem 6. By Lemma 3, there exists a randomized single-winner rule (for any m) that achieves distortion at most $4\gamma_{\min}^{-1} - 2$. Thus, by Lemma 4, we get a randomized k-committee selection rule (for any m and $k \in [m]$) that achieves distortion at most $4\gamma_{\min}^{-1} - 2$. Finally, by Lemma 3, we get a randomized PB rule with the desired distortion. $\square$

We prove that this is asymptotically optimal as a function of m in Theorem 7, thereby proving that our reduction is, in a sense, tight. Deriving the optimal dependence on $\gamma_{\min}$ is left as an open question.

Theorem 7 (Lower Bound). *For rankings by value, every randomized rule f has distortion*

$$\mathrm{dist}_{\mathrm{rbv}}(f) \geqslant \ln(m)/2 \in \Omega(\log(m)).$$

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 $\overline{p}_\ell = \frac{1}{\ell} \sum_{a \in A_\ell} p_a$ 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

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

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 $\text{sw}(A_t) = t$. 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 $\sum_{a \in \cup_{\ell \in [t]} A_\ell} p_a \cdot 1 = \sum_{\ell \in [t]} \ell \cdot \overline{p}_\ell$, which means the distortion is at least

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

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

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

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

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

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

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

Hence, the left hand side sums to at most $\sum_{\ell \in [k]} \overline{p}_\ell \leqslant 1$. 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 $H_{\lceil \sqrt{m} \rceil} \geqslant \ln(\lceil \sqrt{m} \rceil) \geqslant \ln(\sqrt{m}) = \frac{1}{2}\ln(m)$. □

Remark 2 (Rankings by value-for-money). *Another ranking-based ballot format considered in the PB literature is rankings by value-for-money, which force voters to consider the cost-benefit analysis of different alternatives, rather than just the benefits. In Appendix A, we give analogous upper and lower bounds for this ballot format, showing unbounded deterministic distortion in Theorem 13, and randomized distortion analogous to ranking by value $O\left((\log m)/\gamma_{\min}\right)$ in Theorem 14. We demote this ballot format to the appendix because it can be difficult for voters to compute, and in the deterministic case it is bad; in the randomized case, it behaves similarly to pure rankings-by-value.*

5 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.

5.1 k-approval ballots

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 8 (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 $\varepsilon > 0$ is arbitrarily small, giving 1 utility to a_1, ε utility for all of $a_2 \dots a_k$, and 0 for all $A \setminus \{a_1, \dots, a_k\}$. Then, everyone's public-spirited values are identical to their utilities. All agents approve $a_1, \dots, a_k$, and the deterministic rule must pick $k-1$ of these arbitrarily. Let the deterministic rule pick $a_2 \dots a_k$. The best possible welfare is n, achieved by any $k-1$-subset including a_1; the winner has welfare εn, making the distortion $\frac{1}{\varepsilon}$ (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*

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

Proof. Pick the most approved alternative a. This is in fact the plurality winner and by Theorem 5, the plurality rule achieves the claimed distortion. □

The following proposition shows that this is the best we can hope for. The full proof of Proposition 2 is available in Appendix F.1.

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

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

Proof Sketch. Consider an instance with $\frac{m}{2}$ alternatives of cost 1 where each of them are approved by $\frac{2}{m}$ voters. In addition the remaining $\frac{m}{2}$ alternatives have cost $\frac{m}{2}$, and are never approved by any voter, .

Any PB rule must pick one of the approved alternative, since otherwise we can take the underlying utility profile that gives the unapproved alternatives utility zero. In this case, we can make unapproved alternatives to appear in the second to the $m/2+1$-th position of every voter which gives us the claimed bound. □

Remark 3. *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.

5.2 Knapsack ballots

For the ballot format *knapsack* (knap), the set of possible ballots $\mathcal{L}_{\text{knap}} = \mathcal{F}$ is the set of all budget-feasible subsets of A. Each voter i submits the subset she values most: $\rho_i \in \mathrm{argmax}_{S\in\mathcal{F}} v_i(S)$. 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 2 (LB, knap, Deterministic)**.** *For knapsack ballot format, every deterministic rule* f *has distortion*

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

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 Appendix G.1.

Theorem 9 (LB, knap, Randomized). *For knapsack ballot format, every randomized rules* f *has distortion*

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

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 4 (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* $\mathrm{dist}_{\textit{knap}}(f) \leqslant m$*.*

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

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

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., $c_{a^{\mathrm{m}}} + \sum_{b \in A^+} c_b > 1$. We will show that there exists some $b \in A^+$ such that $\frac{n_{b,a^{\mathrm{m}}}}{n_b} \leqslant \frac{m-1}{m}$.

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

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

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

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

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 $\frac{\mathrm{sw}(a)}{\mathrm{sw}(A^+)} \leqslant 2m^2 \frac{1-\gamma_{\min}}{\gamma_{\min}} + 1$. 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 $\frac{\mathrm{sw}(a)}{\mathrm{sw}(A^+)} \leqslant 2m^2 \frac{1-\gamma_{\min}}{\gamma_{\min}} + 1$, as needed.

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

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

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:

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

It remains to bound $\frac{\mathrm{sw}(A^* \setminus A_0)}{\mathrm{sw}(A_0)}$. 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,

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

Thus,

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

□

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.

Remark 5. *For the special case of committee selection, we show in Appendix G.2 that this bound can be improved to* $m^2(\gamma_{\min}^{-1} - 1) + m \in O\left({}^{m^2}/_{\gamma_{\min}}\right)$.

Remark 6 (Threshold approvals). *Another approval-based ballot format considered in the literature is threshold approvals, which are categorically different than knapsack and k-approvals: instead of approving a limited set of alternatives, voters approve any alternative for which their utility exceeds a certain threshold. In Appendix B, we give analogous upper and lower bounds for this ballot format. For deterministic rules, we show unbounded deterministic distortion for a fixed choice of threshold in Proposition 3 and $\Omega(m)$ and $O\left({}^{m^2}/_{\gamma_{\min}}\right)$ distortion when the threshold is variable in Theorems 16 and 15. For randomized rules, we show $\Omega(\sqrt{m})$ with fixed thresholds and $\Omega(\log m)$ with variable thresholds in Theorems 17 and 18 using the ideas in Benadè et al. [2021]. We demote this ballot format to the appendix due to its limited practicability: even if people can assign internally-consistent numeric values to their utilities, they may not consider their utilities on the same scale, making it hard for people to reliably approve alternatives according a given threshold.*

6 A THRIFTY ORDINAL BALLOT GETS SUBLINEAR DISTORTION

Let us revisit the story so far for deterministic aggregation rules, which is the more practical case. Rankings by value allowed us to achieve $O({}^{m}/_{\gamma_{\min}^2})$ distortion, and approval-based ballots, which could outperform rankings by value in the unit-sum model [Benadè et al., 2021], fail to do so in the public spirit model, leaving our quest of achieving distortion sublinear in m (via a practical ballot format) unfulfilled.

In this section, we introduce a new (family of) ballot format(s), *ranking of predefined bundles (rpb)*, which meets both these desiderata. Not only does it allow achieving sublinear distortion via a deterministic aggregation rule, it is also extremely practical in participatory budgeting due to four reasons:

- *Explainable:* It simply asks voters to rank bundles of projects by value instead of individual projects.
- *Ordinal*: It asks voters to only ordinally compare bundles of projects.
- *Thrifty*: The number of bundles that voters rank is at most m, making the number of bits of information elicited from each voter polynomial in m.
- *Reduction to single-winner voting:* The bundles we create below are budget-feasible (so voters can realistically imagine them being implemented) and pairwise disjoint (so voters can easily compare them). Further, the subset of projects funded in the end is precisely one of the bundles on the ballot. This creates a reduction to single-winner voting, where voters understand that they are effectively expressing preferences over possible final outcomes. This also opens up the possibility of using well-known aggregation rules from single-winner voting (such as our use of Copeland's rule below), which voters may already be familiar with.

Specifically, an rpb ballot is characterized by a set $\mathcal{P} = \{P_1, \ldots, P_\ell\}$ of ℓ feasible subsets of A. We suggest that ℓ should be at most polynomial in m. Thus, $\mathcal{L}_{\text{rpb}(\mathcal{P})}$ is the set of all rankings over $\mathcal{P}$. Each voter i submits a ranking $\rho_i \in \mathcal{L}_{\text{rpb}(\mathcal{P})}$ such that for all bundles $P, P' \in \mathcal{P}$ with $v_i(P) > v_i(P')$, we have $P \succ_{\rho_i} P'$. An aggregation rule f for this format gets $\vec{\rho} \in \mathcal{L}^n_{\text{rpb}(\mathcal{P})}$ as input.

We show how to use the rpb ballot to achieve $O({}^{\sqrt{m}}/_{\gamma_{\min}^2})$ distortion in a one-round voting system, and an even better $O({}^{(\log m)}/_{\gamma_{\min}^4})$ distortion in a two-round voting system.

6.1 Sublinear Distortion in One Round

Let us describe our proposed voting system, which comprises of an rpb ballot we term *high-low bundling* (HLB) along with a deterministic aggregation rule (Copeland's rule).

Ballot: rpb with high-low bundling (HLB). We initialize an rpb ballot with the set $\mathcal{P}^{\text{HLB}}$ constructed as follows. Let $L = \{a \in A : c(a) \leqslant 1/\sqrt{m}\}$ be the set of *low-cost* alternatives, and $H = \{a \in A : c(a) > 1/\sqrt{m}\}$ be the set of *high-cost* alternatives. $\mathcal{P}^{\text{HLB}}$ consists of an arbitrary partition of L into at most $\sqrt{m}$ feasible bundles[2] and an arbitrary partition of H into feasible bundles.[3] Note that $|\mathcal{P}| \leqslant |H| + |L| = m$.[4] The voters are asked to rank the bundles in $\mathcal{P}^{\text{HLB}}$, which generates an input profile $\vec{\rho}$.

Aggregation rule. We simply run Copeland's rule on $\vec{\rho}$, treating each bundle as an alternative in single-winner voting, to select one of the feasible bundles as the final output.

Theorem 11 (Upper Bound). *The distortion of (deterministic) Copeland's aggregation rule $f_{Copeland}$ applied to the HLB ballot is*

$$\text{dist}_{\text{rpb}(\mathcal{P}^{HLB})}(f_{Copeland}) \leqslant \frac{2\sqrt{m}}{\gamma_{\min}^2} \in O\left(\frac{\sqrt{m}}{\gamma_{\min}^2}\right).$$

Proof. Let A^* be an optimal budget-feasible subset of alternatives. The elements of A^* are distributed among L and H, so $\text{sw}(L \cap A^*) + \text{sw}(H \cap A^*) = \text{sw}(A^*)$, implying that either $\text{sw}(L \cap A^*) \geqslant \frac{1}{2}\text{sw}(A^*)$ or $\text{sw}(H \cap A^*) \geqslant \frac{1}{2}\text{sw}(A^*)$. We claim that there exists a bundle $P^* \in \mathcal{P}^{\text{HLB}}$ for which $\text{sw}(P^*) \geqslant \frac{\text{sw}(A^*)}{2\sqrt{m}}$.

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

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

Note that Copeland's rule receives rankings over bundles in $\mathcal{P}^{\text{HLB}}$ as input to pick a bundle P. Using its distortion bound (from single-winner voting), we know that

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

yielding distortion at most $\frac{2\sqrt{m}}{\gamma_{\min}^2} \in O\left(\frac{\sqrt{m}}{\gamma_{\min}^2}\right)$. □

Remark 7. *In the unit-sum model of Benadè et al. [2021] (without public spirit), the distortion of any deterministic aggregation rule on any rpb ballot remains $\Omega(m^2)$ due to single-winner instances (as a special case of PB). When each bundle is budget-feasible, this creates precisely a single-winner instance. And it is easy to see that grouping any two alternatives together can lead to infinite distortion if the voters unanimously find that bundle the most preferable but we may pick the bad alternative in that bundle which the voters have zero value for.*

6.2 Logarithmic Distortion in Two Rounds

Next, we describe a two-round voting system, which beats even the sublinear distortion achieved above and yields a logarithmic distortion.

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

[3] One can use this flexibility of partitioning L and H arbitrarily to make the resulting bundles meet practical desiderata, e.g., including a diverse set of projects. Alternatively, one can also create partitions of L and H into the fewest feasible bundles to reduce the size of the ballot.

[4] In practice, with many low-cost projects, we expect $|\mathcal{P}|$ to be much smaller.

First ballot: rankings by value. Simply use the rankings by value ballot, where voters are asked to rank the alternatives in A.

Second ballot: rpb with tiered-cost bundling (TCB). For $r \in \{0, 1, \ldots, \lceil \log_2 m \rceil\}$, define tiers of costs as

$$T_r = \begin{cases} \{a \in A : c(a) \leqslant 1/m\} & \text{if } r = 0, \\ \{a \in A : 2^{r-1}/m < c(a) \leqslant 2^r/m\} & \text{if } r > 0. \end{cases}$$

For each $r \in \{0, 1, \ldots, \lceil \log_2 m \rceil\}$, use the committee selection rule from Lemma 4 to pick $P_r \subseteq T_r$ of size $t_r = \lfloor \min(|T_r|, \max(1, m/2^r)) \rfloor$. Note that each P_r is budget-feasible. Our rpb ballot in the second stage is now defined by $\mathcal{P}^{\text{TCB}} = (P_0, \ldots, P_{\lceil \log_2 m \rceil})$. Each voter submits a ranking ρ_i over $\mathcal{P}^{\text{TCB}}$.

Aggregation rule. Run (deterministic) Copeland's rule on the input $\vec{\rho}$ and return the bundle $P \in \mathcal{P}^{\text{TCB}}$ that it picks.

Theorem 12. *The distortion of the two-round voting system that uses rankings by value, then the rpb ballot with tiered-cost bundling, and then Copeland's rule is at most* $2(\lceil \log_2 m \rceil + 1) \cdot (2\gamma_{\min}^{-1} - 1)^4$.

Proof. Let A^* be an optimal budget-feasible subset of the alternatives. Fix any $r \in \{0, 1, \ldots, \lceil \log_2 m \rceil\}$. Let P_r^* be the optimal t_r-sized subset of T_r (note that this is feasible by the definition of t_r). Using the distortion bound of the committee selection rule from Lemma 4, we have $\text{sw}(P_r^*) \leqslant (2\gamma_{\min}^{-1} - 1)^2 \cdot \text{sw}(P_r)$. Since A^* is feasible, $|A^* \cap T_r| \leqslant 2t_r$, so $A^* \cap T_r$ can be partitioned into two feasible subsets of T_r of size at most t_r each, yielding $\text{sw}(A^* \cap T_r) \leqslant 2 \cdot \text{sw}(P_r^*) \leqslant 2\left(2\gamma_{\min}^{-1} - 1\right)^2 \cdot \text{sw}(P_r)$.

Since $T_0, \ldots, T_{\lceil \log_2 m \rceil}$ partitions the set of alternatives A, we have

$$\text{sw}(A^*) = \textstyle\sum_{r \in \{0,1,\ldots,\lceil \log_2 m \rceil\}} \text{sw}(A^* \cap T_r) \leqslant 2(\lceil \log_2 m \rceil + 1)\left(2\gamma_{\min}^{-1} - 1\right)^2 \cdot \max_{r \in \{0,1,\ldots,\lceil \log_2 m \rceil\}} \text{sw}(P_r).$$

Using the distortion bound of Copeland's rule, we have that for the bundle P picked by the rule,

$$\text{sw}(P) \geqslant \frac{\max_{r \in \{0,1,\ldots,\lceil \log_2 m \rceil\}} \text{sw}(P_r)}{\left(2\gamma_{\min}^{-1} - 1\right)^2} \geqslant \frac{\text{sw}(A^*)}{2(\lceil \log_2 m \rceil + 1) \cdot \left(2\gamma_{\min}^{-1} - 1\right)^4}. \qquad \square$$

We remark that there are no known lower bounds that prohibit one from achieving even constant distortion using a one-round voting system that uses an rpb (or some other fully ordinal) ballot format with only polynomially many comparisons. We leave this as a major open question that can have implications for PB ballot design in practice.

7 DISCUSSION

Our work lays out several interesting open questions as in some cases, our upper and lower bounds do not asymptotically match (see Tables 1 and 2) in either m, $\gamma_{\min}$ or both.

Our work posits, based on prior research, that democratic deliberation in real-world PB may cause voters to be public-spirited. However, modeling the exact level of public spirit achieved and using this to in turn optimize the design of the deliberation process itself would be an important direction for future research. More broadly, distortion has been studied in models beyond voting, such as matching [Filos-Ratsikas et al., 2014] and fair division [Halpern and Shah, 2021], to which the public-spirit model can also be applied. Finally, under the public-spirit model, participants take the utilitarian welfare into account when submitting their preferences, which works well since the goal is to optimize the utilitarian welfare as well. But the idea of distortion has been extended to other objectives such as the Nash welfare or proportional fairness [Ebadian et al., 2022], which raises the question: what form of public-spirit can be helpful in optimizing such objectives and how can it be cultivated?

REFERENCES

E. Anshelevich, O. Bhardwaj, E. Elkind, J. Postl, and P. Skowron. 2018. Approximating Optimal Social Choice under Metric Preferences. *Artificial Intelligence* 264 (2018), 27–51.

Elliot Anshelevich, Aris Filos-Ratsikas, Nisarg Shah, and Alexandros A Voudouris. 2021. Distortion in Social Choice Problems: The First 15 Years and Beyond. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI)*. 4294–4301. Survey Track.

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

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

G. Benadè, S. Nath, A. D. Procaccia, and N. Shah. 2021. Preference Elicitation For Participatory Budgeting. *Management Science* 65, 5 (2021), 2813–2827.

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

C. Boutilier, I. Caragiannis, S. Haber, T. Lu, A. D. Procaccia, and O. Sheffet. 2015. Optimal Social Choice Functions: A Utilitarian View. *Artificial Intelligence* 227 (2015), 190–213.

Markus Brill, Stefan Forster, Martin Lackner, Jan Maly, and Jannik Peters. 2023. Proportionality in Approval-Based Participatory Budgeting. In *Thirty-Seventh AAAI Conference on Artificial Intelligence*. 5524–5531.

Ioannis Caragiannis, Swaprava Nath, Ariel D. Procaccia, and Nisarg Shah. 2017. Subset Selection Via Implicit Utilitarian Voting. *Journal of Artificial Intelligence Research* 58 (2017), 123–152.

I. Caragiannis and A. D. Procaccia. 2011. Voting Almost Maximizes Social Welfare Despite Limited Communication. *Artificial Intelligence* 175, 9–10 (2011), 1655–1671.

Moses Charikar, Prasanna Ramakrishnan, Kangning Wang, and Hongxun Wu. 2024. Breaking the metric voting distortion barrier. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 1621–1640.

Michiel S De Vries, Juraj Nemec, and David Špaček. 2022. International Trends in Participatory Budgeting. *Cham: Palgrave Macmillan* (2022).

Soroush Ebadian, Anson Kahng, Dominik Peters, and Nisarg Shah. 2022. Optimized Distortion and Proportional Fairness in Voting. In *Proceedings of the 23rd ACM Conference on Economics and Computation (EC)*. 563–600.

Brandon Fain, Kamesh Munagala, and Nisarg Shah. 2018. Fair Allocation of Indivisible Public Goods. In *Proceedings of the 19th ACM Conference on Economics and Computation (EC)*. 575–592.

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

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

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

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

Bailey Flanigan, Ariel D Procaccia, and Sven Wang. 2023. Distortion Under Public-Spirited Voting. In *Proceedings of the 24th ACM Conference on Economics and Computation*. 700.

Ashish Goel, Reyna Hulett, and Anilesh K. Krishnaswamy. 2018. Relating Metric Distortion and Fairness of Social Choice Rules. In *Proceedings of the 13th Workshop on Economics of Networks, Systems and Computation* (Irvine, CA, USA) *(NetEcon '18)*. Association for Computing Machinery, New York, NY, USA, Article 4, 1 pages. https://doi.org/10.1145/3230654.3230658

Daniel Halpern and Nisarg Shah. 2021. Fair and Efficient Resource Allocation with Partial Information. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI)*. 224–230.

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

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

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

Participedia. 2023. https://participedia.net/search?selectedCategory=all&query=participatory%20budgeting.

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

A. D. Procaccia and J. S. Rosenschein. 2006. The Distortion of Cardinal Preferences in Voting. In *Proceedings of the 10th International Workshop on Cooperative Information Agents (CIA)*. 317–331.

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

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

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

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

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

APPENDIX

A 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.

A.1 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 13.

Theorem 13 (lower bound). *For rankings by value for money, every deterministic rule f has unbounded distortion:* $\mathrm{dist}_{\mathsf{vfm}}(f) = \infty$.

Proof. We use the exact same construction used by Benadè et al. [2021]. Fix $a, b \in A$, and let $c_a = \varepsilon > 0$ 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 $u_i(a) = \varepsilon$, $u_i(b) = 1$ and $u_i(x) = 0$ for $x \in A \setminus \{a, b\}$. Then,

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

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

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

which as $\varepsilon \longrightarrow 0$ tends to infinity. □

A.2 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 14 (upper bound). *For rankings by value for money, there exists a randomized rule f with distortion*

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

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 3. 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:

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

Then, the distortion of our rule is, by the same analysis in Theorem 3:

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

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 14. 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 $\left\lfloor \frac{m}{2^j} \right\rfloor$ 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 $j \in \left\{0, 1, \ldots, \lceil \log_2(m) \rceil\right\}$ uniformly at random.
(2) Consider the restricted instance with only the alternatives in A_j.
(3) With $m' = |A_j|$ and $k = \min(m', \left\lfloor \frac{m}{2^j} \right\rfloor)$, 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 $\left\lfloor \frac{m}{2^j} \right\rfloor$-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 $\mathrm{sw}(A^* \cap A_j) \leqslant 2\mathrm{sw}(L_j^*)$ for any $j \neq 0$.

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

$2d\text{sw}(L_j)$. Let δ be the distribution of the output of the mechanism, we have:

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

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.

B 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.[5]

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. □

B.1 Deterministic Rules

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

Theorem 15 (upper bound). *For threshold approval votes with threshold $\tau = 1/m$, there exists a deterministic rule f with distortion*

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

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:

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

[5]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.

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,

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

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

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

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 $t+\varepsilon$ and the utility of b_j be $t-\varepsilon$ for all voters, and any small $\varepsilon > 0$.

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

B.2 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 17 (lower bound). *For threshold approval votes with any threshold $\tau > 0$, every randomized rule f has distortion*

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

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 $\text{sw}(a^*) = n/2$ and for $a \in A \setminus \{a^*\}$, $\text{sw}(a) \leqslant n/\sqrt{m}$. That gives us

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

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 $\text{sw}(a^*) = n/\lfloor\sqrt{m}\rfloor$, and $sw(a) = n(1 - 1/\lfloor\sqrt{m}\rfloor)/(m-1)$ for $a \in A \setminus \{a^*\}$. That gives us:

$$\text{dist}_\tau\text{-th}(f) \geqslant \frac{\text{sw}(a^*)}{\mathbb{E}_{a\sim f(\vec{\rho},c)}[\text{sw}(a)]}$$

$$\geqslant \frac{\frac{n}{\lfloor\sqrt{m}\rfloor}}{\frac{1}{m}\frac{n}{\lfloor\sqrt{m}\rfloor} + \frac{m-1}{m}\frac{n\left(1-\frac{1}{\lfloor\sqrt{m}\rfloor}\right)}{m-1}} \geqslant \frac{m}{\lfloor\sqrt{m}\rfloor} \geqslant \lfloor\sqrt{m}\rfloor,$$

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 $\mathcal{L}_{D\text{-rth}} = \mathcal{L}_{\tau\text{-th}} = 2^A$. 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 18 (lower bound). *For randomized threshold approval votes with the threshold sampled from any distribution D, every randomized rule f has distortion*

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

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:

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

Theorems 17 and 18 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.

C PROOFS FROM SECTION 2 (PRELIMINARIES)

C.1 Proof of Lemma 1

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

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

Proof. The proof is the same as the proof 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:

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

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

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

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

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

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

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

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

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

C.2 Distortion Without 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* $\mathsf{X} \in \{\mathsf{rbv}, \mathsf{vfm}, \mathit{knap}, \tau\text{-}\mathsf{th}, D\text{-}\mathsf{rth}\}$ *(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 $\mathsf{X} \in \{\mathsf{rbv}, \mathsf{vfm}\}$, 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 $\varepsilon \in (0, 1)$, 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

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

Because $\varepsilon \in (0, 1)$ was chosen arbitrarily, we can take the worst case by letting $\varepsilon \to 0$, 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* $\mathsf{X} \in \{\mathsf{rbv}, \mathsf{vfm}, \mathit{knap}, \tau\text{-}\mathsf{th}, D\text{-}\mathsf{rth}\}$ *(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 $\mathsf{X} \in \{\mathsf{rbv}, \mathsf{vfm}\}$, 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 $\varepsilon \in (0, 1)$. 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 $p_{\mathsf{X}} \cdot 1 + (1 - p_{\mathsf{X}}) \cdot \varepsilon \leqslant 1/m + (1 - 1/m) \cdot \varepsilon$. Thus, the distortion of f_{X} is at least $\frac{1}{1/m + (1-1/m)\cdot\varepsilon}$. Since $\varepsilon \in (0, 1)$ was chosen arbitrarily, we can take the worst case by letting $\varepsilon \to 0$, 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. □

D PROOFS FROM SECTION 3 (SINGLE WINNER)

D.1 Proof of Theorem 1

Theorem 1 (Lower Bound - Deterministic). *Any deterministic single-winner voting rules f with ranked preferences has distortion*

$$\text{dist}_{rbv}(f) \geqslant 1 + 2\frac{1-\gamma_{\min}}{\gamma_{\min}} \cdot \frac{m^2}{2\gamma_{\min} + \gamma_{\min}m^2 + (2-3\gamma_{\min})m} \in \Omega\left(\frac{1}{\gamma_{\min}} \cdot \min\left\{m, \frac{1}{\gamma_{\min}}\right\}\right).$$

Proof. Suppose we have m alternatives $a_1, \dots, a_m$ 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,

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

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 $\text{sw}(a_m) > \text{sw}(a_{m-1}) > \dots > \text{sw}(a_2) > \text{sw}(a_1)$. 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 $i \in N_1$,

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

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

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

and when $k \geqslant j$:

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

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

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

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

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

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

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

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

Via the same reasoning,

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

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

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

Now, we can finally solve for S:

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

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

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

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

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

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

This gives distortion, after simplifying:

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

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 $\Omega\left(\frac{\min\{1/\gamma, m\}}{\gamma}\right)$. □

D.2 Proof of Theorem 2

Theorem 2 (Lower Bound - Randomized). *Any randomized single-winner voting rules f with ranked preferences has distortion*

$$\mathrm{dist}_{rbv}(f) \in \Omega\left(\min\left\{m, \frac{1}{\gamma_{\min}}\right\}\right).$$

Proof. Use the same input profile $\vec{\rho}$ as in the proof of Theorem 1. Let $p(a_i)$ be the probability that a_i is picked by rule f and without the loss of generality, suppose that $a_{\min} = \mathrm{argmin}_a p(a)$.

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 1, $\mathrm{sw}(a_1) \leqslant \mathrm{sw}(a_2) \leqslant \cdots \leqslant \mathrm{sw}(a_m)$, and so we can maximize social welfare by picking a_m.

The expected social welfare of f is at most:

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

So, the distortion is:

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

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

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

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

E PROOFS FROM SECTION 4 (RANKINGS BY VALUE)

E.1 Proof of Theorem 4

THEOREM 4 (LOWER BOUND). *For rankings by value, every deterministic rule f has distortion*

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

PROOF. Consider an instance with $A = \{a, b_1, \dots b_{m-1}\}$, 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 $(a \succ b_1 \succ \dots \succ b_{m-1})$ and members of N_2 vote $(b_1 \succ \dots \succ b_{m-1} \succ a)$.

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 $\mathrm{sw}(a) = n(1-p)\frac{\gamma_{\min} p}{1-\gamma_{\min} p}$, and $\mathrm{sw}(b) = np$ 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 $v_i(b_j) = (1-\gamma_{\min}) \times 0 + \gamma_{\min} p = \gamma_{\min} p$. 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:

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

This gives us:

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

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

By picking budget-feasible set $\{b_1, \dots, b_{m-1}\}$ 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 $\gamma_{\min} \leqslant 1$, $p \geqslant p\gamma_{\min}$, 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 $\mathrm{sw}(a) = n(1-p)$, and $\mathrm{sw}(b) = np\frac{\gamma_{\min}(1-p)}{1-\gamma_{\min}(1-p)}$ 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:

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

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

E.2 Proof of Lemma 4

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

Proof. Let $A^* = \{a_1^*, \ldots, a_k^*\}$ be the optimal budget-feasible set, sorted from highest social welfare to the lowest so that $i < j \implies \mathrm{sw}(a_i^*) \geqslant \mathrm{sw}(a_j^*)$. 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 $\mathrm{sw}(a_i^+) \geqslant \mathrm{sw}(a_i^*)$ and since the single-winner rule has expected distortion of d, we have $\mathbb{E}[\mathrm{sw}(a_i)] \geqslant \frac{\mathrm{sw}(a_i^+)}{d}$ which implies $\mathbb{E}[\mathrm{sw}(a_i)] \geqslant \frac{\mathrm{sw}(a_i^*)}{d}$. Summing this over all iterations and using linearity of expectation, we get that

$$\begin{aligned}
&\sum_{i=0}^{k} \mathbb{E}[\mathrm{sw}(a_i)] \geqslant \sum_{i=0}^{k} \mathrm{sw}(a_i^*) \,/\, d \\
&\implies \mathrm{sw}(A^*) \,/\, \mathbb{E}[\mathrm{sw}(S)] \leqslant d.
\end{aligned}$$ □

F PROOFS FROM SECTION 5.1 (k-APPROVALS)

F.1 Proof of Proposition 2

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

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

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 $b_1, \ldots, b_{m/2}$, 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 $\gamma\frac{m-2}{m-2\gamma_{\min}}$, 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, $\mathrm{sw}(b_i) = \frac{m-\gamma(m-2)}{m-2\gamma} \cdot \frac{m-2}{m} \cdot n$ for all i from 1 to $\frac{m}{2}$, and $\mathrm{sw}(a_i) = \frac{2n}{m}$ for $i \neq \frac{m}{2}$ with $\mathrm{sw}(a_{m/2}) = \frac{2n}{m} \cdot \gamma\frac{m-2}{m-2\gamma}$.

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

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

for all b_j, where the last inequality holds because $m \geqslant m - 2\gamma$. 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 $\mathrm{sw}(a_k) \leqslant \mathrm{sw}(a_j)$ and voter i gives a_k zero utility. So, a_j is indeed the highest ranking alternative.

But, the distortion we get is:

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

as claimed. □

G PROOFS FROM SECTION 5.2 (KNAPSACK)

G.1 Proof of Theorem 9

THEOREM 9 (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:

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

□

G.2 Knapsack for Committee Selection

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

Theorem 19. *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 10. 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:

$$\begin{aligned}\frac{\mathrm{sw}(A^*,U)}{\mathrm{sw}(A,U)} &= \frac{\sum_{a^*\in A^*}\mathrm{sw}(a^*,U)}{\sum_{a\in A}\mathrm{sw}(a,U)}\\ &= \frac{\sum_{a^*\in A^*\cap A}\mathrm{sw}(a^*,U)}{\sum_{a\in A}\mathrm{sw}(a,U)} + \frac{\sum_{a^*\in A^*\setminus A}\mathrm{sw}(a^*,U)}{\sum_{a\in A}\mathrm{sw}(a,U)}\\ &\leqslant 1 + \sum_{a^*\in A^*\setminus A}\frac{\mathrm{sw}(a^*,U)}{\sum_{a\in A}\mathrm{sw}(a,U)}. \end{aligned} \tag{2}$$

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

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

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:

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

(2) Suppose that for some $a^* \in A^* \setminus A$, and for all $a \in A$, $n_{a,a^*}/n_a > 1/2$. Let $a_{\max} = \mathrm{argmax}_{a\in A} n_a$ and $a_{\min} = \mathrm{argmin}_{a\in A} n_a$. Then, in particular,

$$\begin{aligned} n_{a_{\max}} &< 2n_{a_{\max},a^*}\\ &\leqslant 2n_{a^*}\\ &\leqslant 2n_{a_{\min}}, \end{aligned}$$

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]:

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

From here we can prove an m^2 bound easily by taking $a^*_{\max} = \mathrm{argmax}_{a^*\in A^*}\mathrm{sw}(a^*,U)$. Then, continuing off of (2), and using the fact that there exists some $\widehat{a} \in A$ such that $\frac{\mathrm{sw}(a^*_{\max},U)}{\mathrm{sw}(\widehat{a},U)} \leqslant$

$2\frac{1-\gamma_{\min}}{\gamma_{\min}}m+1$:

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

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 $a_{\min} = \mathrm{argmin}_{a\in A} N(a,a^*)$. 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! □