The Distortion of Public-Spirited Participatory Budgeting

Mark Bedaywi
University of Toronto

Bailey Flanigan
Carnegie Mellon University

Mohamad Latifian
University of Toronto

Nisarg Shah
University of Toronto

Abstract

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. Our results demonstrate that, within this model of voter behavior, several common ballot formats permit low distortion, often even outperforming existing bounds achieved under restricted utilities. These findings highlight 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.

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 then submit their preferences, which are used to select a set of projects with total cost at most B to be funded. PB is now used all over the world — even at the national level — to decide allocations of public funds [De Vries et al., 2022; Participedia, 2023; Wampler et al., 2021].

Given the growing deployment of PB, it is key to understand how much the budget allocations produced actually benefit society. 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 $\mathsf{sw}(S) = \sum_{i \in N} u_i(S)$.

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*. Benadè et al. [2021] study four ballot formats for PB: *rankings by value*, where each voter i ranks the alternatives in a non-increasing order of $u_i(a)$; *rankings by value-for-money*, where i ranks the alternatives in a non-increasing order of $u_i(a)/c_a$; *knapsack votes*, where i specifies her favorite budget-feasible set of projects $\text{argmax}_{S: \sum_{a \in S} c_a \leqslant B} u_i(S)$; and *threshold approval votes*, where the government specifies a threshold τ and i approves every project a with $u_i(a) \geqslant \tau$.

An *aggregation rule* takes the n ballots as input to find a budget-feasible set of projects. The quality of this outcome is measured by the *distortion*: the worst-case (over possible latent utilities) ratio of the best possible social welfare that of the outcome. By taking the worst case over inputs, we can measure the overall efficiency of a combination of ballot format and aggregation rule. Ideally, we would like to design ballot formats and aggregation rules that lead to low-distortion PB outcomes.

Ballot Format	Upper bounds		Lower bounds	
	Deterministic aggregation rules			
Rankings by value	$m\gamma_{\text{min}}^{-1} \cdot \min\{m, \gamma_{\text{min}}^{-1}\}$	(Thm. 2)	$(m-1)\gamma_{\text{min}}^{-1}$	(Thm. 1)
Rankings by value/money			∞	(Thm. 5)
Knapsack	$m + m^3\gamma_{\text{min}}^{-2}$	(Thm. 7)	$m(\gamma_{\text{min}}^{-1} - 1)$	(Thm. 8)
Threshold approvals	$m^2\gamma_{\text{min}}^{-1}$	(Thm. 10)	$m-1$	(Thm. 11)
	Randomized aggregation rules			
Rankings by value	$\log(m)\gamma_{\text{min}}^{-1}$	(Thm. 3)	$\log(m)$	(Thm. 4)
Rankings by value/money	$\log(m)\gamma_{\text{min}}^{-1}$	(Thm. 6)		
Knapsack	m	(Rem. 2)	$m - \gamma_{\text{min}}(m-1)$	(Thm. 9)

Table 1: Asymptotic (in m, γ_{min}) distortion bounds across ballot formats. Bounds are sometimes coarsened slightly for simplicity. We exclude randomized threshold approvals, as we consider two separate sources of randomness.

Unfortunately, we immediately encounter a stark impossibility: using *any* of the ballot formats listed above, *any* deterministic aggregation rule has unbounded distortion. If randomized aggregation rules are allowed, we can achieve distortion at most m by simply ignoring the ballots and funding *a single* project chosen uniformly at random; however, this is crude and unsatisfying, and we would like to do better when randomization is allowed.

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.

In this paper, our goal is to design ballot formats and aggregation rules that lead to low-distortion PB outcomes, regardless of voters' underlying utilities. Recent work by Flanigan et al. [2023] offers a promising approach: under unrestricted utilities, they achieve low distortion in single-winner elections by leveraging the established idea that voters may be *public-spirited*: when casting their ballots, voters consider others' interests in addition to their own. As they 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 that PB voters may *already* be public-spirited motivates our main research question:

> **Question:** *If voters are public-spirited, for which popular PB ballot formats — if any — do there exist aggregation rules with small distortion?*

An affirmative answer to this question would suggest a practicable approach — democratic deliberation — to achieving higher-welfare outcomes in PB elections.

Results and contributions. We study distortion under the four PB ballot formats mentioned above. Following Flanigan et al. [2023], we depart from the standard model of how each voter i translates her utilities into a ballot: instead of evaluating each project a by just her own utility $u_i(a)$, she evaluates a according to her *public-spirited (PS) value*, given by 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.

Within this model, we prove upper and lower bounds describing what distortion is fundamentally achievable when using the *best* (deterministic and randomized) aggregation rules for each ballot format. This allows us to compare the ballot formats themselves. Our bounds, summarized in Table 1, are parameterized by $\gamma_{\text{min}} = \min_i \gamma_i$, the minimum public spirit level of any voter. These bounds reflect the following contributions:

Contribution 1. First, we find that public spirit can dramatically improve the distortion for multiple popular ballot formats. Using certain ballot designs, the distortion of some deterministic rules drops from unbounded to at most m, with only mild dependencies on γ_{min}. For randomized rules, certain ballot formats permit distortion of at most $\log(m)$. This suggests that the use of deliberation, provided it truly promotes public spirit, can lead to dramatic improvements in the welfare of PB outcomes. Despite requiring no assumptions on the utilities, our bounds often even outperform bounds under the unit-sum utilities assumption, which we provide for comparison throughout the paper. The most significant such improvement occurs for *knapsack* ballots with deterministic rules, where the distortion drops from exponential in m to order m^3.

Contribution 2. Toward our original goal of identifying the ballot format permitting the *lowest* distortion, we find that when voters are public-spirited, rankings-by-value permits the best possible distortion using either randomized and deterministic aggregation rules, for all but very small values of γ_{min}. This is convenient, as it is the simplest ballot format for voters to complete.

Contribution 3. Finally, our technical analysis introduces entails multiple ideas of independent interest. First, we show how to transform deterministic and randomized single-winner rule into corresponding PB rules while incurring only an additional factor of $O(m)$ and $O(\log(m))$, respectively (Lemma 2 and Lemmas 3 and 4). Our analysis of knapsack ballots also includes a novel approach of comparing entire subsets of alternatives. Finally, in analyzing PB we derive new bounds for two restricted settings: single-winner voting (Proposition 4, Lemma 5) and k-committee selection (Lemma 4, Lemma 6, Theorem 14).

1.1 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 (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 and Procaccia, 2011; Caragiannis et al., 2017] and $\Theta(\sqrt{m})$ [Boutilier et al., 2015; Ebadian et al., 2022], respectively. Optimal distortion bounds have also been identified for k-committee selection [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

In participatory budgeting (PB), 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, 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 $\mathcal{F} = \{S \subseteq A : c(S) \leqslant B\}$ be all *budget-feasible* sets of alternatives. The goal of PB is to select a such a budget-feasible set by eliciting and aggregating voter preferences.

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 sets of k alternatives. We use "k-committee rule" to refer to an aggregation 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 an aggregation rule designed 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 $\mathsf{sw}(a, U) = \sum_{i \in N} u_i(a)$; for a subset of alternatives $S \subseteq A$, define $\mathsf{sw}(S, U) = \sum_{a \in S} \mathsf{sw}(a, U)$. We use $\mathsf{sw}(a)$ or $\mathsf{sw}(S)$ when U is clear from context.

PS-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 will depend on the minimum public spirit level of the voters, denoted by $\gamma_{\min} = \min_{i \in N} \gamma_i$.

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, i's *PS-value* for a is given by

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

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 as $\mathsf{sw}(a) = \sum_{i \in N} u_i(a) = \sum_{i \in N} v_i(a)$ 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 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" 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.

Following Benadè et al. [2021], we study four ballot formats — Rankings by Value, Rankings by value for money, Knapsack Votes, and Threshold Approval Votes — which we formally define 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 of 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 rule f on input profile $\vec{\rho}$ in format X and cost function c is the following worst-case ratio:

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

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

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

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 then consider the lowest distortion enabled by each ballot format, across all voting rules taking in ballots of that format. This is roughly a measure of the usefulness of the information contained in the ballot format for social welfare maximization.

Supporting results. Before presenting our main results, we state a lemma we will use throughout the paper. This is a generalization of Lemma 3.1 due to Flanigan et al. [2023]; the proof is in Appendix A.1.

Lemma 1. *Let $A_1, A_2 \subseteq A$ be two arbitrary sets of alternatives. For any fixed constant $\alpha \geqslant 0$, let $N_{A_1 \succ A_2} = \{i \in N : \alpha v_i(A_1) \geqslant v_i(A_2)\}$. Then:*

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

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 and randomized rules have at best m distortion (Appendix A.2).

3 Rankings by Value

In the ballot format *rankings by value* (rbv), each voter ranks the alternatives in a non-increasing order of her values for them. Formally, $\mathcal{L}_{\mathsf{rbv}}$ is the set of all rankings of the alternatives, and each voter i submits a ranking $\rho_i \in \mathcal{L}_{\mathsf{rbv}}$ 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. Note that rbv is the canonical ballot format in single-winner and k-committee selection — both special cases of PB — so rules for these cases will be assumed to use rankings by value unless stated otherwise.

3.1 Deterministic Rules

First, we show that for rbv ballots, deterministic rules must incur a distortion at least $(m-1)\gamma_{\text{min}}^{-1}$:

Theorem 1 (lower bound)**.** *For rankings by value, every deterministic rule f has distortion*

$$\operatorname{dist}_{\mathsf{rbv}}(f) \geqslant (m-1)\gamma_{\mathit{min}}^{-1}.$$

The intuition for this bound is as follows: first, observe that even when $\gamma_i = 1$ for all i (i.e., when all voters rank alternatives by social welfare alone), any deterministic rule must still incur $\Omega(m)$ distortion. This is because rankings are insufficiently informative, even when voters agree, to recover the optimal budget-feasible set. When $\gamma_{\text{min}} < 1$ and voters can disagree, the lower bound gets worse. The full proof is in Appendix B.1.

Next, we prove an upper bound that matches our lower bound with respect to m by applying single-winner voting rules with low distortion under public spirit. To enable this approach, we first prove a general result: when used in the PB context, any single-winner rule will incur up to an additional factor of m distortion.

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

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

$$\begin{aligned} \mathsf{sw}(A^*)/\mathsf{sw}(a) &= \sum_{a^* \in A^*} \mathsf{sw}(a^*)/\mathsf{sw}(a) \\ &\leqslant m \cdot \max_{a^* \in A^*} \mathsf{sw}(a^*) \,/\, \mathsf{sw}(a) \leqslant m \cdot d. \square \end{aligned}$$

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 $m(\gamma_{\text{min}}^{-1} - 1) + 1$ and Copeland's rule has distortion at most $\left(2\gamma_{\text{min}}^{-1} - 1\right)^2$. Plugging these bounds into Lemma 2, we conclude upper bounds for the PB setting:

Theorem 2 (upper bound)**.** *For rankings by value,*

$$\begin{aligned} \operatorname{dist}_{\mathsf{rbv}}(\mathit{plurality}) &\leqslant m^2(\gamma_{\mathit{min}}^{-1} - 1) + m, \ \mathit{and} \\ \operatorname{dist}_{\mathsf{rbv}}(\mathit{copeland}) &\leqslant m\left(2\gamma_{\mathit{min}}^{-1} - 1\right)^2. \end{aligned}$$

Remark 1. Note that there remains a gap between our upper and lower bounds (in Theorem 2 and (Theorem 1, respectively): Plurality achieves the optimal dependence on γ_{min}, Copeland achieves the optimal dependence on m, but neither achieves both. This gap parallels the gap present in the single-winner case [Flanigan et al., 2023]. If a single-winner rule with distortion $O(\gamma_{\text{min}}^{-1})$ were identified, plugging that into Lemma 2 would close the gap in our bounds for PB as well. We leave this for future work.

3.2 Randomized Rules

Theorem 3 (upper bound)**.** *For rankings by value, there exists a randomized rule f with distortion*

$$\mathrm{dist}_{\mathsf{rbv}}(f) \leqslant 4(\lceil \log_2(m) \rceil + 1)\left(2\gamma_{min}^{-1} - 1\right).$$

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 a novel randomized single-winner rule with asymptotically optimal (in both m and γ_{min}) distortion of at most $4\gamma_{\text{min}}^{-1} - 2$ (Lemma 5). We now state and prove these results in succession, before applying them to prove Theorem 3.

Lemma 3. *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 for $i \neq 0$,

$$A_i = \left\{a \in A : 2^{i-1}/m < c_a \leqslant 2^i/m\right\}, \text{ and}$$
$$A_0 = \{a \in A : c_a \leqslant 1/m\}.$$

The randomized PB rule f is as follows:

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

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

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

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

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. *Fix any $k \in [m]$, and $d \geqslant 1$. If there exists a randomized single-winner rule with distortion at most d for each $m' \leqslant m$, then there exists a randomized k-committee selection rule with distortion at most d.*

We defer the full proof to Appendix B.2. The main idea is to repeatedly pick alternatives using the rule k times.

Having reduced the PB problem to that of single-winner selection, we now present a novel randomized single-winner rule with the following distortion. The proof is located in Appendix B.3.

Lemma 5. *There exists a randomized single-winner voting rule with distortion at most* $4\gamma_{min}^{-1} - 2$.

Proof of Theorem 3. Finally, we apply Lemmas 3, 4 and 5 to prove Theorem 3. By Lemma 5, there exists a randomized single-winner rule (for any m) that achieves distortion at most $4\gamma_{\text{min}}^{-1} - 2$. Thus, by Lemma 4, we get a randomized k-committee selection rule (for any m and $k \in [m]$) that achieves distortion at most $4\gamma_{\text{min}}^{-1} - 2$. Finally, by Lemma 3, we get a randomized PB rule with the desired distortion. □

We prove that this is asymptotically optimal as a function of m in Appendix B.4, thereby proving that our reduction is, in a sense, tight. Deriving the optimal dependence on γ_{min} is left as an open question.

Theorem 4 (lower bound)**.** *For all randomized f,*

$$\text{dist}_{\mathsf{rbv}}(f) \geqslant \ln(m)/2.$$

4 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.

4.1 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 Appendix C.1.

Theorem 5 (lower bound)**.** *For rankings by value for money, every deterministic rule f has unbounded distortion:* $\text{dist}_{\mathsf{vfm}}(f) = \infty$.

4.2 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. The full proof is in Appendix C.2.

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

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

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

5 Knapsack Votes

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

5.1 Deterministic Rules

Benadè et al. [2021] prove that, under the unit-sum assumption, any deterministic rule with knapsack votes has distortion in $\Omega(2^m/\sqrt{m})$. This exponential lower bound might suggest that knapsack votes carry little information useful for welfare maximization. However, we show that under public-spirited voting, knapsack votes permit deterministic rules to achieve distortion at most polynomial in m.

Theorem 7 (upper bound)**.** *For knapsack votes, there exists a deterministic rule f with distortion*

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

Proof sketch. To build an efficient voting rule, we need information on whether $v_i(a) \geqslant v_i(b)$ for voters $i \in N$ and alternatives $a, b \in A$. Knapsack votes give this to us in a peculiar way: For subsets of alternatives $A_1, A_2 \subseteq A$, if $c(A_1) \geqslant c(A_2)$ and yet voter i includes all of A_1 but none of A_2 in their knapsack vote, we know that $v_i(A_1) \geqslant v_i(A_2)$. Otherwise, the voter could have swapped A_1 with A_2 to get another feasible subset of higher utility. Thus, costly alternatives with many votes are extremely valuable. As such, the voting rule f that gives this bound goes out of its way to pick an expensive alternative a that sufficiently many voters included in their knapsacks. Then, using Lemma 1, it prunes all alternatives b such that $v_i(a) \geqslant v_i(b)$ for sufficiently many voters. This process continues until the budget is exhausted. The formal proof is in Appendix D.1. $\square$

For the special case of committee selection, we can improve this bound to $m^2(\gamma_{\min}^{-1} - 1) + m$, as shown in Theorem 14, Appendix D.3.

Finally, we lower-bound the distortion of deterministic rules with knapsack votes. In the instance, $c(a) = 1$ for all $a \in A$, thereby reducing the problem to single-winner selection with only plurality votes. The bound then follows using the same construction as in Proposition 3.15 of Flanigan et al. [2023].

Theorem 8 (lower bound)**.** *For knapsack votes, every deterministic rule f has distortion*

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

5.2 Randomized Rules

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 D.2.

Theorem 9 (lower bound)**.** *For all randomized rules f,*

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

Remark 2 (upper bound)**.** Note that ignoring all the ballots and simply picking a single alternative uniformly at random trivially yields an upper bound of m. Hence, the lower bound in Theorem 9 is tight at least in m. It is worth noting that the optimal distortion of randomized rules for knapsack votes is also $\Theta(m)$ under the unit-sum assumption [Benadè et al., 2021].

6 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, $\mathcal{L}_{\tau\text{-}\mathsf{th}} = 2^A$, as with knapsack votes. Benadè et al. [2021] introduce this ballot format for unit-sum utilities and our definition extends it to arbitrary utilities.[1]

[1] 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 Appendix E.1, we show that this leads to the worst possible distortion: unbounded for deterministic rules and m for randomized rules.

6.1 Deterministic Rules

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

Theorem 10 (upper bound)**.** *For threshold approval votes with threshold $\tau = 1/m$, there exists a deterministic rule f with distortion*

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

Proof sketch. With $\tau = 1/m$, each voter must approve at least one alternative. Thus, the most approved alternative must be approved by at least n/m voters. Lemma 1 implies that picking such an alternative achieves the desired bound. The proof is in Appendix E.2. □

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

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

Proof sketch. Consider the input profile where every voter only approves alternative a, and a cost function where the rule can either select a or all other alternatives. No matter which decision the rule makes we define a utility matrix that gives the desired distortion. The full proof is found in Appendix E.3. □

6.2 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 12 (lower bound)**.** *For threshold approval votes with any threshold $\tau > 0$, every randomized rule f has distortion*

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

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). The formal definition is given in Section 6.

Theorem 13 (lower bound)**.** *For randomized threshold approval votes with the threshold sampled from any distribution D, every randomized rule f has distortion*

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

Theorems 12 and 13 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 in Appendices E.4 and E.5 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.

7 Discussion

Our work lays out several interesting open questions as in some cases, our upper and lower bounds do not asymptotically match (see Table 1) either in m, $\gamma_{\min}$ or both. In addition to PB, some gaps also remain for the special cases of single-winner and committee selection — most notably, whether a deterministic single-winner rule with $O(\gamma_{\min}^{-1})$ distortion exists.

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. Approximating optimal social choice under metric preferences. *Artificial Intelligence*, 264:27–51, 2018.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Appendix

A Proofs from Section 2 (Preliminaries)

A.1 Proof of Lemma 1

Lemma 1. *Let $A_1, A_2 \subseteq A$ be two arbitrary sets of alternatives. For any fixed constant $\alpha \geqslant 0$, let $N_{A_1 \succ A_2} = \{i \in N : \alpha v_i(A_1) \geqslant v_i(A_2)\}$. Then:*

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

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:

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

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

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

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

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

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

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

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

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

□

A.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 1. *For any ballot format $\mathsf{X} \in \{\mathsf{rbv}, \mathsf{vfm}, \mathsf{knap}, \tau\text{-}\mathsf{th}, D\text{-}\mathsf{rth}\}$ (with any threshold τ and threshold distribution D), every deterministic rule has unbounded distortion when $\gamma_i = 0$ for all $i \in N$.*

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

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

- Under $\mathsf{X} \in \{\mathsf{rbv}, \mathsf{vfm}\}$, 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 $\epsilon \in (0, 1)$, consider the following consistent utility matrix U.

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

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

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

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

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

For randomized threshold approval votes with any threshold distribution D, we cannot fix the input profile upfront as it depends on the threshold τ sampled from D. However, we can assume that for each τ the rule sees the profile described above for $\tau\text{-}\mathsf{th}$. 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). $\square$

Proposition 2. *For any ballot format $\mathsf{X} \in \{\mathsf{rbv}, \mathsf{vfm}, \mathsf{knap}, \tau\text{-}\mathsf{th}, D\text{-}\mathsf{rth}\}$ (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 1. 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 $\epsilon \in (0, 1)$. We define a consistent utility matrix U that works for all of the above ballot formats:

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

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

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

B Proofs from Section 3 (Rankings by Value)

B.1 Proof of Theorem 1

Theorem 1 (lower bound)**.** *For rankings by value, every deterministic rule f has distortion*

$$\mathrm{dist}_{\mathsf{rbv}}(f) \geqslant (m-1)\gamma_{\mathit{min}}^{-1}.$$

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

Now consider two cases.

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

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

and $v_i(b_j) = (1-\gamma_{\min}) \times 0 + \gamma_{\min}p = \gamma_{\min}p$. 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 $\{b_1, \ldots, b_{m-1}\}$ we can get a social welfare of $n(m-1)p$, while instead we got $n(1-p)\frac{\gamma_{\min} p}{1-p\gamma_{\min}}$ by choosing a. This leaves us with a distortion of

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

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

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

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

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

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

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

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

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

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

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

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

B.2 Proof of Lemma 4

Lemma 4. *Fix any $k \in [m]$, and $d \geqslant 1$. If there exists a randomized single-winner rule with distortion at most d for each $m' \leqslant m$, then there exists a randomized k-committee selection rule with distortion at most d.*

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

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

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

□

B.3 Proof of Lemma 5

Lemma 5. *There exists a randomized single-winner voting rule with distortion at most* $4\gamma_{min}^{-1} - 2$*.*

Proof. Define the domination graph to be a directed graph G with vertices A and an edge between every two 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.

Due to a known implication of Farkas' lemma (see Theorem 2.4 of Harutyunyan et al. [2017]), there exists a probability distribution p over the vertices such that for any vertex $v \in A$, the probability of picking v or a vertex with an edge to v is at least $1/2$.

The randomized voting rule that picks alternatives using this distribution then achieves the required bound. Indeed, let a^* be the optimal alternative. If we pick a^* or an alternative that beats a^* in a pairwise election, b, we get distortion by Lemma 1:

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

Let the set of such alternatives be A'. Then, the distortion of our rule is:

$$\begin{aligned}
\frac{\mathsf{sw}(a^*)}{\sum_{a \in A} p(a)\mathsf{sw}(a)} &\geqslant \frac{\mathsf{sw}(a^*)}{\mathsf{sw}_{a \in A'} p(a)\mathsf{sw}(a)} \\
&\geqslant \frac{\mathsf{sw}(a^*)}{\min_{a \in A'} \mathsf{sw}(a) \sum_{a \in A} p(a)} \\
&\geqslant 2\frac{\mathsf{sw}(a^*)}{\min_{a \in A'} \mathsf{sw}(a)} \\
&\geqslant 4\frac{1-\gamma_{\text{min}}}{\gamma_{\text{min}}} + 2 = \frac{4}{\gamma_{\text{min}}} - 2,
\end{aligned}$$

as claimed. □

B.4 Proof of Theorem 4

Theorem 4 (lower bound)**.** *For all randomized f,*

$$\text{dist}_{\mathsf{rbv}}(f) \geqslant \ln(m)/2.$$

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

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

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

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

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

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

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

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

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

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

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

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

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

which completes the proof after observing that $H_{\lceil \sqrt{m} \rceil} \geqslant \ln(\lceil \sqrt{m} \rceil) \geqslant \ln(\sqrt{m}) = \frac{1}{2} \ln(m)$.

□

C Proofs from Section 4 (Rankings by Value for Money)

C.1 Proof of Theorem 5

Theorem 5 (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 $c_a = \epsilon > 0$ and $c_x = 1$ for all $x \in A \setminus \{a\}$. Construct an input profile $\vec{\rho}$ where each voter has alternatives a and b in positions 1 and 2, and let f be some deterministic aggregation rule.

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

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

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

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

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

which as $\epsilon \to 0$ tends to infinity. ☐

C.2 Proof of Lemma 6

Lemma 6. *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 Lemma 5. Indeed, everything is the same, except that when b beats a^* in a pairwise election (i.e. at least $n/2$ times), we get the following distortion by Lemma 1:

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

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

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

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

C.3 Proof of Theorem 6

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

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

Proof. Let g be the rule in Lemma 6, and let the distortion it achieves, $\left(4\frac{1-\gamma_{\min}}{\gamma_{\min}} + 2\right)$, be d. By the same mechanism in Lemma 3, we will convert g to a ranking by value per cost rule.

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

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

and

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

Recall the mechanism used:

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

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

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

and

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

The randomized PB rule f is as follows:

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

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

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

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

which gives us the desired distortion bound. $\square$

D Proofs from Section 5 (Knapsack)

Let us first introduce additional notation. Given any n knapsack sets and 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.

D.1 Proof of Theorem 7

Theorem 7 (upper bound)**.** *For knapsack votes, there exists a deterministic rule f with distortion*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Thus,

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

D.2 Proof of Theorem 9

Theorem 9 (lower bound)**.** *For all randomized rules f,*

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

Proof. Formally, consider a case where n is divisible by m, all the voters have the same PS-value of $\gamma = \gamma_{\text{min}}$, 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:

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

□

D.3 Proof of Theorem 14

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

Theorem 14. *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. 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.

Let $N(a)$ be the number of times alternative a appears, and let $N(a,b)$ be the number of times some voter voted for both a and b.

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

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

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

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

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

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

by considering two cases:

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

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

2. Suppose that for some $a^* \in A^* \setminus A$, and for all $a \in A$, $N(a,a^*)/N(a) > 1/2$. Let $a_{\max} = \text{argmax}_{a\in A} N(a)$ and $a_{\min} = \text{argmin}_{a\in A} N(a)$. 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 7 below, we know that for all $a^* \in A^* \setminus A$, there exists some $a \in A$ such that $N(a,a^*)/N(a) \leqslant (m-2)/m$. Therefore, $N(a) - N(a,a^*) \geqslant 2N(a)/m \geqslant n/2m$. Thus, by Lemma 1 in [Flanigan et al., 2023]:

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

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

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

as claimed! □

Lemma 7. *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 $a_{\text{min}} = \text{argmin}_{a \in A} N(a, a^*)$. Then, substituting this into the inequality above, and using that $|A| = \frac{m}{2}$:

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

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

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

and so finally

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

as desired! □

E Proofs from Section 6

E.1 Proof of Proposition 3

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

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

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

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

E.2 Proof of Theorem 10

Theorem 10 (upper bound)**.** *For threshold approval votes with threshold $\tau = 1/m$, there exists a deterministic rule f with distortion*

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

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

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

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

and so:

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

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

By Lemma 1,

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

as claimed. □

E.3 Proof of Theorem 11

Theorem 11 (lower bound)**.** *For all deterministic f and all threshold values $\tau > 0$,*

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

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

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

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

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

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

E.4 Proof of Theorem 12

Theorem 12 (lower bound)**.** *For threshold approval votes with any threshold $\tau > 0$, every randomized rule f has distortion*

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

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

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

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

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

which gives us the desired bound. □

E.5 Proof of Theorem 13

Theorem 13 (lower bound)**.** *For randomized threshold approval votes with the threshold sampled from any distribution D, every randomized rule f has distortion*

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

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

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

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

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

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

Indeed, let k be a value such that

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

which must exist by the pigeonhole principle.

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

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

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

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

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

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

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

Indeed, the expected social welfare of f is:

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

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

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

F Proofs for Single-winner rules

F.1 Proof of Proposition 4

Proposition 4. *Every deterministic single-winner rule f has distortion at least $2\frac{1-\gamma_{min}}{\gamma_{min}} + 1$.*

Proof. The key idea is that because we want to show distortion independent of m, we should be able to ignore all but two alternatives. Consider the following input profile with two types of voters:

$$A : a_1 \succ a_2 \succ a_3 \succ \cdots \succ a_m$$
$$B : a_2 \succ a_1 \succ a_3 \succ \cdots \succ a_m$$

Then, suppose that $n/2$ voters are of type A and the rest of type B, each of whom has a PS-level of $\gamma_{\min}$.

Suppose a deterministic voting rule f picks an alternative a_k with $k \notin \{1, 2\}$. Then, we can make the utility all voters have for a_1 and a_2 1, and the utility all voters have for $a_3, \ldots, a_m$ zero. Because all utilities are the same, this is consistent under PS-voting, and we have infinite distortion.

Without the loss of generality, suppose the voting rule picks a_1 instead. Then, suppose all voters of type A have a utility of $\frac{\gamma_{\min}}{2-\gamma_{\min}}$ for a_1 and 0 for everything else, and all voters of type B have utility 1 for a_2 and zero for everything else. Then, $\frac{\mathsf{sw}(a_1,U)}{n} = \frac{\gamma_{\min}}{2(2-\gamma_{\min})}$ and $\frac{\mathsf{sw}(a_2,U)}{n} = \frac{1}{2}$.

This is consistent. Indeed, because every voter has zero utility for them, the PS value is zero, and so any ranking involving them is consistent. Moreover, $a_1 \succ a_2$ for voters of type A as:

$$\begin{aligned}(1-\gamma_{\min})\frac{\gamma_{\min}}{2-\gamma_{\min}} + \gamma_{\min}\frac{\gamma_{\min}}{2(2-\gamma_{\min})} &= \frac{2\gamma_{\min} - 2\gamma_{\min}^2 + \gamma_{\min}^2}{2(2-\gamma_{\min})} \\ &= \frac{\gamma_{\min}(2-\gamma_{\min})}{2(2-\gamma_{\min})} \\ &= (1-\gamma_{\min})\cdot 0 + \gamma_{\min}\cdot\frac{1}{2}.\end{aligned}$$

Voters of type B have public spirited values of $\frac{\gamma_{\min}^2}{2(2-\gamma_{\min})}$ for a_1 and $(1-\gamma_{\min})+\gamma_{\min}\frac{1}{2} = \frac{2-\gamma_{\min}}{2}$. Since $(2-\gamma_{\min})^2 \geqslant \gamma_{\min}^2$ as $2-\gamma_{\min} \geqslant \gamma_{\min} \geqslant 0$, we know that

$$\frac{(2-\gamma_{\min})^2}{2(2-\gamma_{\min})} \geqslant \frac{\gamma_{\min}^2}{2(2-\gamma_{\min})},$$

and so the ranking is consistent for voters of type B as well.

Finally, we get that the distortion is at least:

$$\frac{\mathsf{sw}(a_2, U)}{\mathsf{sw}(a_1, U)} = \frac{2-\gamma_{\min}}{\gamma_{\min}} = 2\frac{1-\gamma_{\min}}{\gamma_{\min}} + 1. \qquad \square$$