



# The distortion of threshold approval matching

Mohamad Latifian<sup>1</sup> · Alexandros A. Voudouris<sup>2</sup>

Received: 16 December 2024 / Accepted: 7 October 2025  
© The Author(s) 2025

## Abstract

We study matching settings in which a set of agents have private utilities over a set of items. Each agent reports a partition of the items into approval sets of different threshold utility levels. Given this limited information on input, the goal is to compute an assignment of the items to the agents (subject to cardinality constraints depending on the application) that (approximately) maximizes the social welfare (the total utility of the agents for their assigned items). We first consider the well-known, simple one-sided matching problem in which each of  $n$  agents is to be assigned exactly one of  $n$  items. We show that with  $t$  threshold utility levels, the distortion of deterministic matching algorithms is  $\Theta(\sqrt[t]{n})$  while that of randomized algorithms is  $\Theta(\sqrt[t+1]{n})$ . We then show that our distortion bounds extend to a more general setting in which there are multiple copies of the items, each agent can be assigned a number of items (even copies of the same one) up to a capacity, and the utility of an agent for an item depends on the number of its copies that the agent is given.

**Keywords** Distortion · Matching problems · Approval · Social choice

## 1 Introduction

The assignment of papers to reviewers in conference management systems like CMT, Hot-CRP, EasyChair and OpenReview is computed using bidding information that classifies the papers into sets based on whether the reviewers are, for example, eager, willing, or not willing to handle them. In a sense, this process defines a collection of *threshold levels* that the reviewers (or, more generally, *agents*) can use to partition the papers (or, more generally, *items*) into associated *approval sets* based on their preferences (which can be dependent on their experience, their interests, and so on). Similar elicitation methods can be used in other

---

✉ Alexandros A. Voudouris  
alexandros.voudouris@essex.ac.uk

<sup>1</sup> School of Informatics, University of Edinburgh, Scotland, UK

<sup>2</sup> School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK

important *constrained resource allocation* applications, such as the classic *house allocation* problem (and generalizations of it), where interested buyers (playing the role of agents) can bid for houses (corresponding to the items) by classifying them into categories depending on monetary value or other criteria.

Eliciting only threshold approvals rather than more detailed information about the underlying utility preferences of the agents for the items inevitably leads to inefficiency in terms of natural, cardinal objectives such as the well-known *social welfare* (the total utility). Typically, the loss of efficiency of decision-making methods that have access only to incomplete information is captured by the notion of *distortion*, which is defined as the worst-case ratio of the maximum possible social welfare over that of the computed solution. The distortion was originally used for social choice settings (such as voting) where decisions are made only based on ordinal information (rankings) [13, 31], but has recently been studied for settings in which different types of information is available or can be elicited (e.g., see [3, 4, 28–30]).

In the matching setting we study in this work, the threshold approvals reported by the agents is a type of information that lies in-between fully cardinal and fully ordinal since it provides some information about the range of the cardinal utilities and there is an order between items in different sets. Hence, while we cannot hope to achieve full efficiency, we can hope to achieve distortion better than what is possible just with ordinal preferences, depending on how detailed the threshold approvals are. In particular, we are interested in the possible *tradeoffs* between the distortion and the number of threshold levels both for when allocations are computed deterministically (which is the most natural way of doing so in social choice problems), as well as when randomization can be exploited.

## 1.1 Our contribution

We start by considering the fundamental *one-sided matching problem* (also known as *house allocation*) to introduce the main ideas of our techniques, before turning to a more general setting. In one-sided matching, there is a set of  $n$  agents with utilities for a set of  $n$  items. The utilities are private and are not explicitly reported by the agents. Instead, for a number  $t$  of decreasing threshold values, each agent reports a collection of  $t$  *approval sets* consisting of items of different utility level; in particular, each approval set is associated with a threshold value and includes all items for which the agent has utility that is at least this threshold. Given the approval sets as input, our goal is to determine a one-to-one matching between agents and items so that the social welfare (total utility of the agents for their assigned items) is maximized.

We show tight bounds on the best possible distortion achieved by matching mechanisms for any number  $t$  of thresholds. In particular, we show a bound of  $\Theta(\sqrt[t]{n})$  for deterministic mechanisms, and a bound of  $\Theta(\sqrt[t+1]{n})$  for randomized ones under the standard assumption of unit-sum utilities [13, 17].<sup>1</sup> The lower bounds are presented in Section 3 and the upper bounds in Section 4. To put the bounds into perspective, we note that just one threshold is

---

<sup>1</sup>We remark that without any normalization assumption, the distortion of deterministic mechanism can be shown to be unbounded in cases where we cannot distinguish between agents that have very large and very small utilities. In addition, normalization is necessary for the elicitation of threshold approval sets to be meaningful and provide useful information. See the paper of [10] for further motivation and justification of normalization assumptions.

sufficient to obtain distortion  $\Theta(n)$  for deterministic algorithms, beating the best possible distortion of  $\Theta(n^2)$  that can be achieved using ordinal information [4]. Similarly, a distortion of  $\Theta(\sqrt{n})$  can be achieved with randomization, matching the best possible distortion achieved by ordinal randomized algorithms [23].

Our deterministic mechanism is based on the following simple idea: Given the approval sets of the agents (induced by appropriately chosen threshold values), we compute a maximum-weight matching in a bipartite graph with the agents on one side, the items on the other side, and edges between them with weights equal to the lower threshold values associated with the approval sets. The slightly improved asymptotically tight randomized upper bound is obtained by running the deterministic mechanism with probability 1/2, while with the remaining 1/2 probability we compute a random matching (chosen equiprobably among all possible ones).

In Section 5, we turn our attention to a more general setting where the agents have *capacities*, indicating the maximum number of items they can receive, and the items have *supplies*, indicating the number of copies of them that are available. We make a budget-balance assumption that the total capacity is asymptotically of the same order of the total supply; for example, in the paper assignment problem, all papers must receive a (constant) number of reviews, and so the total capacity must be sufficiently larger than the total supply, but of the same magnitude. We also assume that, when agents can receive multiple copies of an item, their utility depends on the number of these copies, and thus the copies are not treated as independent items. Our goal is to compute an allocation of items to agents, such that the capacity and the supply constraints are satisfied, and the social welfare is maximized.

As this setting is a generalization of the one-sided matching (in which the number of agents is equal to the number of items, and there are unit capacities and supplies), our lower bounds from Section 3 extend directly. For the upper bounds, we show that, for  $t$  thresholds, the best possible distortion achieved by deterministic mechanisms is  $O(c \cdot \sqrt[t]{T})$ , while the best distortion achieved by randomized mechanisms is  $O(c \cdot \sqrt[t+1]{T})$ , where  $T$  is the total available supply (or capacity) and  $c$  is a parameter that depends either on the maximum capacity or the ratio between the number of items and agents. From this, we get bounds  $\Theta(\sqrt[t]{n})$  and  $\Theta(\sqrt[t+1]{n})$  when the capacities and the supplies are constant.

## 1.2 Related work

The distortion was originally defined by [31] to measure the worst-case loss in social welfare when voting decisions are made using only ordinal information. Since then, the distortion has been studied for several different voting problems, including utilitarian voting [13, 16, 17, 22, 31], metric voting [7, 18–20, 24, 27], and combinations of the two [25]. It has also been studied for social choice problems beyond voting, such as one-sided matching that we also consider in this paper [4, 5, 23], as well as other clustering and graph problems [1, 6, 14]. See the survey of [8] for an introduction to the distortion framework and the various models that have been considered.

Our paper follows a relatively recent stream of papers within the distortion literature that have considered elicitation methods beyond ordinal information. In this direction, [29, 30] showed tradeoffs between the best possible distortion and a communication complexity measure (the number of bits the agents can use to report information) for utilitarian voting. Results of similar flavour for metric voting have also been shown, for example, by [26].

More related to our work, in a series of papers, [3–5] studied voting and matching settings in which the agents provide ordinal information and, on top of that, are capable of answering *value queries* about their utilities for specific alternatives. They showed lower and upper bounds on the distortion of deterministic mechanisms that are functions of the number of queries per agent that are of similar to ours in the sense that the distortion decreases with the number of queries; some of their lower bounds (related to the number of queries required to achieve constant distortion) were recently improved by [15]. Another related paper is that of [28] who considered the one-sided matching problem when the agents can answer binary threshold queries about whether their utility for specific alternatives is larger than appropriately chosen thresholds. Using an approach similar to that of [4], they showed bounds on the distortion of deterministic mechanisms that is a function of the number of queries in terms of the social welfare among matchings that satisfy properties such as Pareto or rank-maximality. Ignoring differences in the models, the elicitation methods in these papers are related to the one we consider since the threshold approval sets can be computed using a number of (value or binary threshold) queries. Hence, our elicitation method is in a sense a bit more demanding. However, for the setting we focus here, we are able to show asymptotically tight bounds not only for deterministic mechanisms, but also for randomized ones, which have not been studied before.

Threshold approvals have also been recently explored in various other works on the distortion of voting mechanisms, most notably by [21] who showed that a single, appropriately chosen threshold is sufficient to achieve a distortion of  $O(\sqrt{m})$  in utilitarian single-winner voting with  $m$  alternatives. In metric voting, [9] showed improved distortion bounds using an approval set per agent computed by a threshold value that is relative (rather than absolute) to the distance from the top-ranked alternative. Threshold approvals have also been considered in the context of participatory budgeting by [11], and voting under truthfulness constraints by [12]. All these works use just a single threshold, whereas we here explore the full potential of this elicitation method (for matching problems, rather than voting) using multiple thresholds and show tight bounds on the possible distortion.

## 2 The one-sided matching problem

For any integer  $t$ , let  $[t] = \{1, \dots, t\}$ . We start with the simple *one-sided matching* setting to express the core idea; in Section 5, we show that our results extend to a more general setting that more accurately captures applications such as paper assignment in peer reviewing. Let  $\mathcal{N}$  be a set of  $n$  agents and  $\mathcal{M}$  be a set of  $n$  items. Agent  $i$  has a *utility function*  $u_i : \mathcal{M} \rightarrow [0, 1]$  over the items. We assume that these utility functions satisfy the unit-sum assumption, which means for each agent  $i \in \mathcal{N}$ ,  $\sum_{j \in \mathcal{M}} u_i(j) = 1$ . Together, these utility functions form the *utility profile*  $\vec{u}$ . A *matching* of the items to the agents is a bijection  $A : \mathcal{N} \rightarrow \mathcal{M}$ . With a slight abuse of notation, we use  $A_i = A(i)$  to refer to the item matched to agent  $i$ , and also  $A(a)$  to refer to the agent matched to item  $a$ . We define the *social welfare* of a matching  $A$  under utility profile  $\vec{u}$  to be the total utility of the agents for the items they are matched to, i.e.,

$$\text{sw}(A, \vec{u}) = \sum_{i \in \mathcal{N}} u_i(A_i) = \sum_{a \in \mathcal{M}} u_{A(a)}(a).$$

The goal is to compute a matching with high social welfare in the worst case. For ease of notation, we will drop  $\vec{u}$  from  $\text{sw}(A, \vec{u})$  whenever it is clear from context.

**Elicitation method** In this paper, we focus on eliciting threshold approval votes with a *threshold vector*  $\vec{\tau} = (\tau_1, \dots, \tau_t)$  such that  $\tau_1 > \dots > \tau_t$ . Given such a vector we ask each agent  $i$  to submit  $t$  disjoint *threshold approval subsets* of  $\mathcal{M}$ , denoted by  $S_{i,1}, \dots, S_{i,t}$ , where  $S_{i,k}$  includes the items for which the agent has utility in  $[\tau_{k-1}, \tau_k)$ , with  $\tau_0 := 1$ . In other words,  $S_{i,k} = \{j \in \mathcal{M} : \tau_{k-1} \geq u_i(j) > \tau_k\}$ . These sets do not necessarily partition  $\mathcal{M}$  as there might exist item  $j$  for which  $u_i(j) < \tau_t$ . All these  $n \times t$  threshold approval sets form the *input profile*  $S$ . Note that different utility profiles might induce the same input profile. We say that a utility profile  $\vec{u}$  is consistent with an input profile  $S$  (in which case we write  $\vec{u} \triangleright S$ ) if for each agent  $i \in \mathcal{N}$ ,  $k \in [t]$ , and  $j \in S_{i,k}$ ,  $\tau_{k-1} \geq u_i(j) > \tau_k$ .

**Mechanisms and distortion** A *mechanism*  $f$  defines a threshold vector  $\vec{\tau}$ , takes an input profile  $S$  based on  $\vec{\tau}$ , and then outputs a matching  $f(S)$  of the items to the agents. The distortion of a matching  $A$  on input profile  $S$  is defined as:

$$\text{dist}(A, S) = \sup_{\vec{u} \triangleright S} \frac{\text{sw}(A^*, \vec{u})}{\text{sw}(A, \vec{u})},$$

where  $A^*$  is the matching with the maximum social welfare with respect to  $\vec{u}$ . The distortion of a matching mechanism  $f$  is defined as the worst case distortion of  $f$  on any input profile:

$$\text{dist}(f) = \sup_S \text{dist}(f(S), S).$$

### 3 Lower bounds

In this section we show lower bounds on the best possible distortion achievable by deterministic and randomized mechanisms for the one-sided matching problem. In particular, for mechanisms that use  $t \geq 1$  thresholds, we show a lower bound of  $\Omega(\sqrt[t]{n})$  for deterministic mechanisms and a lower bound of  $\Omega(\sqrt[t+1]{n})$  for randomized mechanisms. We start by showing a technical lemma that holds for randomized mechanisms that will be useful in establishing the lower bounds in several cases. For a randomized mechanism  $f$ , denote by  $p(i, a)$  the probability that item  $a$  is assigned to agent  $i$  according to  $f$ .

**Lemma 3.1** *For any subset of items  $M \subseteq \mathcal{M}$ , let  $A_M$  be the matching of the items in  $M$  to the agents with minimum sum of probabilities over the items matched to them in  $A_M$  with respect to  $f$ . Then,  $\sum_{a \in M} p(A_M(a), a) \leq 1$ .*

**Proof** We will prove the claim by constructing a matching  $A$  with sum of probabilities at most 1; then, since  $A_M$  is the matching with minimum sum of probabilities, the same must hold for it as well. Consider the following greedy algorithm: Starting with  $T = M$  and  $N = \mathcal{N}$ , in each iteration, find the pair  $(i, a) \in N \times T$  with the minimum possible  $p(i, a)$ , assign  $a$  to  $i$ , and then remove  $a$  and  $i$  from  $T$  and  $N$ , respectively.

Let  $P_j$  be the sum of probabilities of the assigned items at the end of the  $j$ -th iteration. We will show by induction that  $P_j \leq j/|M|$ . For  $j = 1$ , by the pigeonhole principle, we can find a pair with  $p(i, a) \leq 1/|M|$ . So, we now assume that our hypothesis holds for any  $j < \ell$ , and want to prove that  $P_\ell \leq \ell/|M|$ . Let  $(i, a)$  be a pair with minimum  $p(i, a)$  in the  $\ell$ -th iteration. Since the algorithm is greedy, we know that for any item  $b$  that was previously assigned to agent  $A(b)$ ,  $p(A(b), b) \leq p(i, b)$ . This means that the sum of the probabilities of  $i$  receiving any item assigned in a previous iteration is at least  $P_{\ell-1}$ . Hence, by the pigeonhole principle again,  $p(i, a) \leq \frac{1-P_{\ell-1}}{|M|-\ell+1}$ , which implies that

$$\begin{aligned}
 P_\ell &= P_{\ell-1} + p_{i,a} \leq P_{\ell-1} + \frac{1 - P_{\ell-1}}{|M| - \ell + 1} \\
 &\leq \frac{\ell - 1}{|M|} + \frac{1 - \frac{\ell-1}{|M|}}{|M| - \ell + 1} = \frac{\ell - 1}{|M|} + \frac{\frac{|M|-\ell+1}{|M|}}{|M| - \ell + 1} = \frac{\ell}{|M|}.
 \end{aligned}$$

Consequently, we overall have that  $\sum_{a \in M} p(A_M(a), a) \leq P_{|M|} \leq 1$ . □

We are now ready to show the lower bounds via a sequence of lemmas capturing different cases. The first lower bound depends on the ratio of consecutive threshold levels and holds for any mechanism (randomized or deterministic).

**Lemma 3.2** *Consider a threshold vector  $\vec{\tau} = (\tau_1, \dots, \tau_t)$ , and let  $k \in [t]$  be such that  $\delta = \tau_{k-1}/\tau_k$  is the largest multiplicative gap between two consecutive thresholds (assuming  $\tau_0 = 1$ ). Then, the distortion of any matching mechanism  $f$  that uses  $\vec{\tau}$  is  $\Omega(\delta)$ .*

**Proof** For constant  $\delta$  the bound is trivial, so consider  $\delta > 2$ . For an infinitesimal  $\varepsilon$ , define  $\tau_k^+ := \tau_k + \varepsilon$ , and find integer  $m$  such that  $\tau_{k-1}/2 + m\tau_k^+ \leq 1 < \tau_{k-1}/2 + (m + 1)\tau_k^+$ . We partition the items into three sets:  $\mathcal{M}_1$  with  $m + 1$  items,  $\mathcal{M}_2$  with 1 item, and  $\mathcal{M}_3$  with  $n - m - 2$  items. Note that since  $\varepsilon$  is an infinitesimal, we have

$$\tau_{k-1}/2 + (m + 1)\tau_k^+ > 1 \implies (m + 1)\tau_k^+ > 1/2 \implies m + 1 > \frac{\delta}{2}. \tag{1}$$

Now consider an instance with input profile such that

- $S_{i,k} = \mathcal{M}_1$ ,
- $S_{i,k'} = \mathcal{M}_2$  for some  $k' > k$ , and
- $S_{i,j} = \emptyset$  for  $j \in [t] \setminus \{k, k'\}$

We define the following consistent utility profile (that is, the utilities of each agent sum up to 1 and the utilities induce the aforementioned input profile). Let  $A_{\mathcal{M}_1}$  be the matching of the items in  $\mathcal{M}_1$  to the agents with minimum sum of probabilities, and let  $\mathcal{N}_1$  be the set of agents that are assigned an item from  $\mathcal{M}_1$  according to the matching  $A_{\mathcal{M}_1}$ , that is,  $\mathcal{N}_1 := \{i \in \mathcal{N} : \exists a \in \mathcal{M}_1, A_{\mathcal{M}_1}(a) = i\}$ . For fixed item  $a^* \in \mathcal{M}_1$ , we define the utility function

$$u_i(a) = \begin{cases} \tau_{k-1}/2 & a \in \mathcal{M}_1, A_{\mathcal{M}_1}(a) = i \\ \tau_{k-1}/2 & a = a^*, i \notin \mathcal{N}_1 \\ 1 - \tau_{k-1}/2 - m\tau_k^+ & a \in \mathcal{M}_2 \\ 0 & a \in \mathcal{M}_3 \\ \tau_k^+ & o.w. \end{cases}$$

Observe that, for any agent  $i$ , the utilities sum up to 1. For any agent  $i$  there is exactly one item in  $\mathcal{M}_1$  for which  $i$  has value  $\tau_{k-1}/2$  (either  $i \in \mathcal{N}_1$  and thus there is a single item  $a \in \mathcal{M}_1$  such that  $A_{\mathcal{M}_1}(a) = i$  for which  $i$  has value  $\tau_{k-1}/2$ , or  $i \notin \mathcal{N}_1$  and thus  $i$  has value  $\tau_{k-1}/2$  for  $a^*$ ), and  $|\mathcal{M}_1| - 1 = m$  items for which  $i$  has value  $\tau_k^+$ . Consequently,

$$\sum_{a \in \mathcal{M}} u_i(a) = \frac{\tau_{k-1}}{2} + m\tau_k^+ + 1 - \frac{\tau_{k-1}}{2} - m\tau_k^+ = 1.$$

Also, observe that the utility function is consistent to the input profile: for any item  $a \in \mathcal{M}_1$ , the utility of agent  $i$  is either  $\tau_{k-1}/2$  or  $\tau_k^+$ , i.e., in the interval  $[\tau_{k-1}, \tau_k)$ ; for the single item  $a \in \mathcal{M}_2$ , by the choice of  $m$ , the utility of agent  $i$  is

$$1 - \frac{\tau_{k-1}}{2} - m\tau_k^+ < \frac{\tau_{k-1}}{2} + (m + 1)\tau_k^+ - \frac{\tau_{k-1}}{2} - m\tau_k^+ = \tau_k^+ = \tau_k + \varepsilon.$$

Since the inequality is strict, the utility is at most  $\tau_k$  for infinitesimal  $\varepsilon$ .

Now consider the expected social welfare of  $f(S)$  when  $\varepsilon \rightarrow 0$  according to the above utility profile. The maximum achieved utility is  $\tau_{k-1}/2$  from  $a^*$ ,

$$p(A_{\mathcal{M}_1}(a), a) \cdot \tau_{k-1}/2 + (1 - p(A_{\mathcal{M}_1}(a), a)) \cdot \tau_k$$

from each item  $a \in \mathcal{M}_1 \setminus \{a^*\}$ , and at most  $\tau_k$  from the item in  $\mathcal{M}_2$ . By Lemma 3.1, we have that  $\sum_{a \in \mathcal{M}_1} p(A_{\mathcal{M}_1}(a), a) \leq 1$ , and thus

$$\begin{aligned} \mathbb{E}_{A \sim f(S)}[\text{sw}(A, \vec{u})] &\leq \frac{\tau_{k-1}}{2} + \sum_{a \in \mathcal{M}_1} \left( p(A_{\mathcal{M}_1}(a), a) \cdot \left( \frac{\tau_{k-1}}{2} - \tau_k \right) + \tau_k \right) + \tau_k \\ &\leq \tau_{k-1} + (m + 1)\tau_k. \end{aligned}$$

In the optimal matching, we can assign each item  $a \in \mathcal{M}_1$  to agent  $A_{\mathcal{M}_1}(a)$  for a social welfare of  $(m + 1)\tau_{k-1}/2$ . Due to (1), the distortion is

$$\frac{1}{2} \cdot \frac{(m + 1)\tau_{k-1}}{\tau_{k-1} + (m + 1)\tau_k} \geq \min \left( \frac{m + 1}{2}, \frac{\tau_{k-1}}{2\tau_k} \right) \in \Omega(\delta).$$

and the proof is complete. □

Our next two lemmas provide lower bounds for deterministic and randomized mechanisms, respectively, for when the last threshold level is sufficiently large.

**Lemma 3.3** *Consider a threshold vector  $\vec{\tau} = (\tau_1, \dots, \tau_t)$  such that  $\tau_t \geq 1/(n - 1)$ . Then, the distortion of any deterministic matching mechanism  $f$  that uses  $\vec{\tau}$  is unbounded.*

**Proof** Consider the input profile  $S$  where the threshold approval sets of any agent are empty, and thus the utility of any agent for any item is at most  $\tau_t$ . Let  $A = f(S)$  be the matching computed by the deterministic matching mechanism  $f$ , and let  $B$  be another matching such that  $A(a) \neq B(a)$  for every item  $a \in \mathcal{M}$ . Consider the utility profile  $\vec{u}$  where agents have utility 0 for their matched item in  $A$ , utility  $\tau_t$  for their matched item in  $B$ , and  $(1 - \tau_t)/(n - 2)$  for each of the remaining  $n - 2$  items. Note that  $\tau_t \geq 1/(n - 1) \implies (1 - \tau_t)/(n - 2) \leq \tau_t$ , and hence  $\vec{u} \triangleright S$ . Since  $\text{sw}(A, \vec{u}) = 0$  and  $\text{sw}(B, \vec{u}) = n \cdot \tau_t > 0$ , the distortion is unbounded.  $\square$

**Lemma 3.4** Consider a threshold vector  $\vec{\tau} = (\tau_1, \dots, \tau_t)$  such that  $\tau_t > 1/n$ . Then, the distortion of any randomized matching mechanism  $f$  that uses  $\vec{\tau}$  is  $\Omega(n \cdot \tau_t)$ .

**Proof** Consider the input profile  $S$  where the threshold approval sets of any agent are empty, and thus the utility of any agent for any item is at most  $\tau_t$ . Let  $A_{\mathcal{M}}$  be the matching over  $\mathcal{M}$  with minimum sum of probabilities; by Lemma 3.1,  $\sum_{a \in \mathcal{M}} p(A_{\mathcal{M}}(a), a) \leq 1$ . Now, consider the utility profile  $\vec{u}$  where each agent has utility  $\tau_t$  for the item she is matched to according to  $A_{\mathcal{M}}$  and utility  $(1 - \tau_t)/(n - 1)$  for each of the remaining items. Note that  $\tau_t \geq 1/n \implies (1 - \tau_t)/(n - 1) \leq \tau_t$ , and hence  $\vec{u} \triangleright S$ . The expected social welfare of the mechanism is

$$\begin{aligned} \mathbb{E}_{A \sim f(S)} \left[ \sum_{a \in \mathcal{M}} u_{A(a)}(a) \right] &= \sum_{a \in \mathcal{M}} \left( p(A_{\mathcal{M}}(a), a) \cdot \tau_t + (1 - p(A_{\mathcal{M}}(a), a)) \cdot \frac{1 - \tau_t}{n - 1} \right) \\ &= \left( \tau_t - \frac{1 - \tau_t}{n - 1} \right) \sum_{a \in \mathcal{M}} p(A_{\mathcal{M}}(a), a) + n \cdot \frac{(1 - \tau_t)}{n - 1} \\ &\leq \tau_t - \frac{1 - \tau_t}{n - 1} + n \cdot \frac{(1 - \tau_t)}{n - 1} = 1. \end{aligned}$$

Since  $\text{sw}(A_{\mathcal{M}}, \vec{u}) = n \cdot \tau_t$ , the distortion is at least this much.  $\square$

By appropriately combining Lemmas 3.2, 3.3, and 3.4, we can establish the desired lower bounds on the distortion of the different types of mechanisms.

**Theorem 3.5** The distortion of any deterministic matching mechanism  $f$  that uses a threshold vector  $\vec{\tau}$  of length  $t$  is  $\Omega(\sqrt[t]{n})$ .

**Proof** If  $\tau_t \geq 1/(n - 1)$ , by Lemma 3.3, the distortion is unbounded. Otherwise, if  $\tau_t \leq 1/(n - 1)$ , let  $k \in \arg \max_{j \in [t]} \tau_{j-1}/\tau_j$  with  $\tau_0 = 1$ . Clearly,

$$\left( \frac{\tau_{k-1}}{\tau_k} \right)^t \geq \prod_{j \in [t]} \frac{\tau_{j-1}}{\tau_j} = \frac{1}{\tau_t} \implies \delta = \frac{\tau_{k-1}}{\tau_k} \geq \tau_t^{-1/t} \geq \sqrt[t]{n},$$

and thus, by Lemma 3.2, the distortion is  $\Omega(\sqrt[t]{n})$ .  $\square$

**Theorem 3.6** The distortion of any matching mechanism  $f$  that uses a threshold vector  $\vec{\tau}$  of length  $t$  is  $\Omega(\sqrt[t+1]{n})$ .

**Proof** Suppose that the threshold vector  $\vec{\tau}$  is such that  $\tau_t > n^{-t/(t+1)}$ . Since  $n^{-t/(t+1)} \geq n^{-1}$ , by Lemma 3.4, the distortion of  $f$  is  $\Omega(n \cdot \tau_t) = \Omega(n^{t+1/\sqrt{t+1}})$ . So, we can now assume that  $\tau_t \leq n^{-t/(t+1)}$ . As in the proof of Theorem 3.5, we have that  $\delta = \tau_{k-1}/\tau_k \geq \tau_t^{-1/t} \geq n^{t+1/\sqrt{t}}$ , and thus, by Lemma 3.2, the distortion of  $f$  is  $\Omega(\delta) = \Omega(n^{t+1/\sqrt{t}})$ .  $\square$

### 4 Upper bounds

In this section we present asymptotically tight upper bounds for deterministic and randomized matching mechanisms. Our deterministic mechanism (described below) computes a maximum-weight matching by assuming that each agent has the minimum possible utility (according to the thresholds) for all the items in the different approval set given as input.

**Definition 4.1** For  $\delta > 1$  and  $t \in [n]$ , we define the threshold vector  $\vec{\tau} = (\delta^{-1}, \delta^{-2}, \dots, \delta^{-t})$ . The *deterministic matching mechanism*  $f_t$  uses this threshold vector  $\vec{\tau}$  and, given an input profile  $S$ , constructs the following weighted bipartite graph  $G_S$ : There are  $2n$  nodes in total, consisting of a node  $v_i$  for each agent  $i \in \mathcal{N}$  on the left side and a node  $z_a$  for each item  $a \in \mathcal{M}$  on the right side. For  $i \in \mathcal{N}, k \in [t]$  and  $a \in S_{i,k}$ , there is an edge from  $v_i$  to  $z_a$  with weight  $w(v_i, z_a) = \tau_k$ . The mechanism  $f_t$  finds the maximum weighted matching in  $G_S$  and, for each matched pair  $(v_i, z_a)$ , assigns item  $a$  to agent  $i$ . If there are unmatched pairs remaining,  $f_t$  completes the allocation arbitrarily.

**Example 4.2** Let  $t = 2$  and  $\vec{\tau} = (\tau_1, \tau_2)$ . Suppose that  $S_{1,1} = \{a, c\}$ ,  $S_{2,1} = \{d\}$ ,  $S_{2,2} = \{c\}$ ,  $S_{3,2} = \{a, c, d\}$ , while the remaining approval sets are empty. Mechanism  $f_t$  constructs the graph  $G_S$  shown in Fig. 1, computes a maximum-weight matching, and then assigns any unmatched items arbitrarily.

Before we bound the distortion of the mechanism, we prove two very useful technical lemmas. The first one provides us with a lower bound on the weight of the maximum-weight

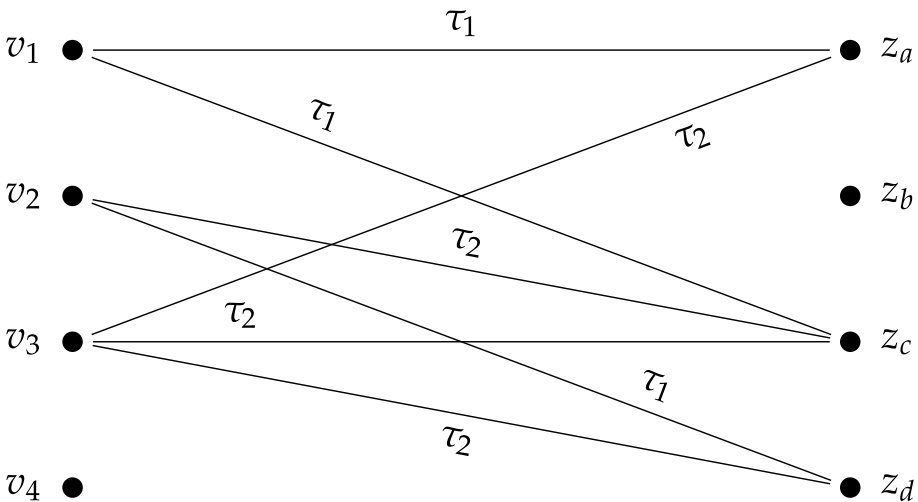


Fig. 1 The graph  $G_S$  that is used by  $f_t$  in Example 4.2

matching in a bipartite graph whose nodes satisfy certain properties; this will be used extensively to lower bound the social welfare of the matching computed by  $f_t$ , and also by the deterministic mechanism in Section 5.

**Lemma 4.3** *Consider a weighted bipartite graph  $G$  with  $n$  nodes on the left side  $\{v_1, \dots, v_n\}$  and  $m$  nodes  $\{z_1, \dots, z_m\}$  on the right side. If  $\sum_{a \in [m]} w(v_i, z_a) \geq W$  for each  $v_i$ , and  $w(v_i, z_a) \geq L$  for each edge  $(v_i, z_a)$ , then there is matching in  $G$  with weight at least  $\min\{W, nL\}$ .*

**Proof** We construct a matching  $B$  in  $G$  as follows: For every  $a \in [m]$ , we match node  $z_a$  to a node  $v_i$  such that the edge  $(v_i, z_a)$  exists in  $G$  and  $w(v_i, z_a)$  is the maximum possible among all available (not already matched) nodes in  $\{v_1, \dots, v_n\}$ . If all nodes in  $\{v_1, \dots, v_n\}$  are matched to a node in  $\{z_1, \dots, z_m\}$ , then since the weight of any edge is at least  $L$ , the total weight of  $B$  is at least  $nL$ . Otherwise, if there exists node  $v_i$  that is not matched to any node in  $\{z_1, \dots, z_m\}$  according to  $B$ , then by the definition of  $B$ ,  $w(v_{B(z_a)}, z_a) \geq w(v_i, z_a)$  for any  $z_a$ , and hence the weight of  $B$  is

$$\sum_{a \in [m]} w(v_{B(z_a)}, z_a) \geq \sum_{a \in [m]} w(v_i, z_a) \geq W.$$

Consequently, overall, the weight of  $B$  is at least  $\min\{W, nL\}$ . □

The second lemma shows that  $G_S$  admits a matching with weight that is relatively close to the social welfare of the optimal matching allocation for any utility profile consistent to the input profile  $S$ . We will use this relation in the analysis of the distortion of  $f_t$ , as well as in the analysis of our randomized matching mechanism later on.

**Lemma 4.4** *Let  $s^*$  be the social welfare of the optimal matching allocation. There exists a matching of  $G_S$  with weight at least  $\delta^{-1}(s^* - n\tau_t)$ .*

**Proof** For any agent  $i$ , let  $a_i^*$  be the item that  $i$  is given in the optimal matching allocation  $A^*$ . Clearly, either there exists  $j \in [t]$  such that  $a_i^* \in S_{i,j}$ , or  $a_i^* \notin \bigcup_{j \in [t]} S_{i,j}$ . The total utility accumulated by the agents of the second type is at most  $n\tau_t$ . For the agents of the first type, since  $a_i^* \in S_{i,j}$ , there is an edge between  $v_i$  and  $z_{a_i^*}$  in  $G_S$  of weight

$$\tau_j = \delta^{-j} = \delta^{-1} \cdot \delta^{-j+1} = \delta^{-1} \cdot \tau_{j-1} \geq \delta^{-1} \cdot u_i(a_i^*).$$

Hence, the intersection of  $A^*$  and  $G_S$  gives us a matching of weight at least  $\delta^{-1}(s^* - n\tau_t)$ . □

**Theorem 4.5** *For  $t \in [n]$  and  $\delta = \sqrt[t]{2n}$ , the distortion of the deterministic matching mechanism  $f_t$  is  $O(\sqrt[t]{n})$ .*

**Proof** Let  $A = f_t(S)$  be the matching computed by the mechanism  $f_t$  when given as input an arbitrary input profile  $S$  that is induced by some consistent utility profile  $\vec{u}$ . First, observe that if  $G_S$  admits a matching of weight  $W$  then  $sw(A, \vec{u}) \geq W$ . This follows by the fact that if node  $v_i$  is matched to node  $z_a$  in  $G_S$ , then agent  $i$  has utility at least  $w(v_i, z_a)$  for item  $a$ .

Second, we argue that the weight of the maximum weight matching of  $G_S$  is at least  $\delta^{-1}/2$ . The total utility of an agent for the items in  $\bigcup_{j \in [t]} S_{i,j}$  is at least  $1/2$  since the utility for each of the remaining items is at most  $\tau_t = \delta^{-t} = 1/(2n)$ . Since  $\tau_{j-1} = \tau_j \cdot \delta$  with  $\tau_0 = 1$ , and  $v_i(a) \leq \tau_{j-1}$  and  $w(v_i, z_a) = \tau_j$  for every  $a \in S_{i,j}$ , we have

$$\begin{aligned} \frac{1}{2} &\leq \sum_{j \in [t]} \sum_{a \in S_{i,j}} v_i(a) \leq \sum_{j \in [t]} |S_{i,j}| \tau_{j-1} = \delta \cdot \sum_{j \in [t]} |S_{i,j}| \tau_j \\ &\implies \sum_{j \in [t]} \sum_{a \in S_{i,j}} w(v_i, z_a) \geq \frac{\delta^{-1}}{2}. \end{aligned}$$

Since the edge  $(v_i, z_a)$  exists in  $G_S$  only if  $a \in \bigcup_{j \in [t]} S_{i,j}$ , we have that  $\sum_{a \in \mathcal{M}} w(v_i, z_a) \geq \delta^{-1}/2$  and  $w(v_i, z_a) \geq \tau_t = 1/(2n)$ , and thus, by Lemma 4.3, we have that  $G_S$  admits a matching of weight at least  $\min\{\delta^{-1}/2, 1/2\}$ , which is equal to  $\delta^{-1}/2$  since  $\delta > 1$ . By Lemma 4.4, since  $n\tau_t = 1/2$ , we also have that  $G_S$  admits a matching of weight at least  $\delta^{-1}(s^* - 1/2)$ , where  $s^*$  is the optimal social welfare. Consequently, we overall have established that the social welfare of the allocation  $A$  computed by the mechanism is

$$\text{sw}(A) \geq \max \left\{ \frac{1}{2} \cdot \delta^{-1}, \delta^{-1}(s^* - 1/2) \right\}.$$

Hence, the distortion is at most

$$\frac{\delta \cdot s^*}{\max\{1/2, s^* - 1/2\}}$$

If  $s^* \leq 1$ , then the distortion is at most  $2\delta s^* \leq 2\delta$ . Otherwise, if  $s^* \geq 1$ , then  $s^* - 1/2 \geq s^*/2$  and the distortion is at most  $\delta \cdot \frac{s^*}{s^*/2} = 2\delta$ . In any case, the distortion is at most  $2\delta \in O(\sqrt[4]{n})$ .  $\square$

Due to Theorem 4.5, we can achieve linear distortion using a single threshold level (which is a significant improvement compared to the  $O(n^2)$  distortion that can be achieved with ordinal information) and constant distortion using a logarithmic number of threshold levels.

**Corollary 4.6** *We can achieve distortion  $O(n)$  by using one threshold level and distortion  $O(1)$  by using  $t = O(\log n)$  threshold levels.*

We now turn our attention to randomization. We design a mechanism that is a convex combination of the naive rule which chooses a random matching equiprobably among all possible ones, and the deterministic matching mechanism  $f_t$  that was analyzed above.

**Definition 4.7** The *randomized matching mechanism*  $R_t$  with probability  $1/2$  chooses a matching uniformly at random, and with probability  $1/2$  runs the deterministic mechanism  $f_t$  with threshold vector  $\vec{\tau} = (\delta^{-1}, \delta^{-2}, \dots, \delta^{-t})$  for  $t \in [n]$  and some  $\delta > 1$ .

**Theorem 4.8** For  $t \in [n]$  and  $\delta = \sqrt[t+1]{n}$ , the distortion of the randomized matching mechanism  $R_t$  is  $O(\sqrt[t+1]{n})$ .

**Proof** Let  $A^*$  be an optimal matching with social welfare  $s^*$ , and  $A^2$  the matching computed by the second part of the mechanism (the outcome of  $f_t$ ). In the first part of the mechanism (where a random matching is chosen with probability  $1/2$ ), since each possible matching has probability at least  $1/(n!)$ , each agent is matched to each item with probability at least  $1/n$ . Since the sum of the utilities of each agent for all items is 1, the expected social welfare from the first part is at least

$$\frac{1}{2} \sum_{i \in \mathcal{N}} \sum_{a \in \mathcal{M}} \frac{1}{n} \cdot u_i(a) = \frac{1}{2} \cdot n \cdot \frac{1}{n} = \frac{1}{2}.$$

For the second part of the mechanism (where the deterministic mechanism  $f_t$  using  $\vec{r}$  is employed), since  $s := n\tau_t = n\delta^{-t} = \sqrt[t+1]{n}$ , by Lemma 4.4, there is a matching in  $G_S$  of weight at least  $\delta^{-1}(s^* - s)$ , and thus the expected social welfare of the mechanism from the second part is at least  $\frac{1}{2} \cdot \delta^{-1}(s^* - s)$ . Overall, we have established that

$$\begin{aligned} \mathbb{E}_{A \sim R_t(S)}[\text{sw}(A)] &\geq \frac{1}{2} + \frac{1}{2} \cdot \delta^{-1}(s^* - s) \\ &\geq \frac{1}{2} \cdot \max\left\{1, \delta^{-1}(s^* - s)\right\}, \end{aligned}$$

and thus the distortion is at most

$$\frac{2 \cdot s^*}{\max\{1, \delta^{-1}(s^* - s)\}}.$$

If  $s^* \geq 2s$ , then  $s^* - s \geq s^*/2$  and the distortion is at most  $\frac{2 \cdot s^*}{\delta^{-1}s^*/2} = 4\delta = 4 \sqrt[t+1]{n}$ . Otherwise, if  $s^* < 2s$ , the distortion is at most  $2s^* \leq 4s = 4 \sqrt[t+1]{n}$ . In any case, the distortion is  $O(\sqrt[t+1]{n})$ .  $\square$

### 5 Generalized setting

In this section we consider a generalized setting. Similarly to before,  $\mathcal{N}$  represents a set of  $n \geq 1$  agents. However, here it is not necessarily the case that we have an equal number of items; we define  $\mathcal{M}$  to be a set of  $m \geq 1$  items. Each item  $a \in \mathcal{M}$  has a supply  $m_a \geq 1$ , and each agent  $i \in \mathcal{N}$  has a capacity  $c_i \geq 1$ . For simplicity, we assume that the total supply is equal to the total capacity, that is,  $T := \sum_{i \in \mathcal{N}} c_i = \sum_{a \in \mathcal{M}} m_a$ .<sup>2</sup> While our results hold for any capacity, to make our results interpretable, we will pay particular focus on the case of constant capacities at the end of the proof.

<sup>2</sup>Our results hold even when  $\sum_{i \in \mathcal{N}} c_i = \Theta(\sum_{a \in \mathcal{M}} m_a)$ , in which case we would need to define  $T$  as the maximum between these two quantities.

Agents are allowed to receive copies of the same item, in which case their utility depends on the number of copies they receive; in other words, copies of an item are not considered independent. For each agent  $i$ , item  $a$  and  $j \in [\min\{c_i, m_a\}]$ , we denote by  $u_i(a, j)$  the *marginal utility* that agent  $i$  gets when receiving his  $j$ -th copy of item  $a \in \mathcal{M}$ , and by  $u_i^+(a, j)$  his *total utility* when receiving  $j \leq \min\{c_i, m_a\}$  copies of item  $a \in \mathcal{M}$ , i.e.,  $u_i^+(a, j) = \sum_{k \in [j]} u_i(a, k)$ .

An *allocation*  $X = (x_i(a))_{i \in \mathcal{N}, a \in \mathcal{M}}$  determines the number  $x_i(a)$  of copies of item  $a$  that agent  $i$  is assigned to, such that  $\sum_{a \in \mathcal{M}} x_i(a) \leq c_i$  for every  $i \in \mathcal{N}$  and  $\sum_{i \in \mathcal{N}} x_i(a) \leq m_a$  for every  $a \in \mathcal{M}$ . Given an allocation  $X$ , the utility of  $i$  for  $X$  is  $u_i(X) = \sum_{a \in \mathcal{M}} u_i^+(a, x_i(a))$ . We assume that the utility function of each agent  $i$  satisfies the unit-sum assumption, that is,  $\sum_{a \in \mathcal{M}} u_i^+(a, \min\{c_i, m_a\}) = 1$ . The definition of the social welfare of an allocation is the same as before, that is, it is the total utility of the agents for the allocation. We aim to compute allocations with high social welfare that maximally assign the items to the agents; observe that the maximum possible number of items that can be allocated is  $T$  (or  $\min\{\sum_{i \in \mathcal{N}} c_i, \sum_{a \in \mathcal{M}} m_a\}$  in the more general case), and that any allocation  $Y$  that assigns less than  $T$  items is dominated in terms of social welfare by any allocation  $X$  that assigns the items allocated by  $Y$  in the same way, but also somehow assigns the remaining items. So, in the following, we focus on such maximal allocations only.

Since the utility functions depend on the number of item copies that the agents receive, we need to appropriately redefine the *elicitation method*. For a threshold vector  $\vec{\tau} = (\tau_1, \dots, \tau_t)$ , each agent  $i$  reports  $t$  disjoint threshold approval sets  $S_{i,1}, \dots, S_{i,t}$ , where  $S_{i,k}$  includes pairs of items and indices for which  $i$  has marginal utility in  $[\tau_{k-1}, \tau_k)$ , where  $\tau_0 := 1$ . In other words,  $S_{i,k} = \{a \in \mathcal{M}, j \in [\min\{c_i, m_a\}] : \tau_{k-1} \geq u_i(a, j) > \tau_k\}$ . The input profile  $S$  now consists of these threshold approval sets reported by all agents. The definition of distortion can also be appropriately refined by taking the worst case over utility profiles and input profiles consistent with them.

We now define a parametric min-cost flow instance that will be used by our algorithms later.

**Definition 5.1** (Min-cost Flow Instance) Let  $C = (c_i)_{i \in \mathcal{N}}$  be a capacity vector of the  $n$  agents with maximum value  $c$ ,  $M = (M_a)_{a \in \mathcal{M}}$  be a supply vector of the  $m$  items, and  $V_{n \times m \times c}$  a value matrix in which  $V(i, a, j)$  is the value of agent  $i$  when receiving  $j$  copies of item  $a$ . Let  $G(C, M, V)$  be the following min-cost flow instance:  $G$  has a source node  $s$  and a destination node  $t$ . Furthermore, there is a node  $v_i$  for each agent  $i$  and a node  $z_a$  for each item  $a$ . For every  $i$ , there is an edge  $(s, v_i)$  with capacity  $c_i$  and cost 0. For every  $a$ , there is an edge  $(z_a, t)$  with with capacity  $M_a$  and cost 0. For each agent  $i$  and item  $a$ , we add a component to the graph as shown in Fig. 2. Nodes  $s$  and  $t$  have supply and demand equal to  $\min\left(\sum_{i \in [n]} c_i, \sum_{a \in [m]} M_a\right)$ , respectively. The goal is to find the minimum cost for satisfying this flow.

In the described network, each unit of flow represents one item. The capacity of the edge  $(s, v_i)$  ensures that agent  $i$  gets at most  $c_i$  items. This unit of flow should go towards  $t$  through one of the gadgets that are designed for agent  $i$  and each of the items. Furthermore, if  $w$  units of flow move through the gadget between  $v_i$  and  $v_a$ , the min-cost solution for this flow is formed by using the edges  $(v_{i,a,k}, v_a)$  for  $k \in [w]$ , and the cost of such a solution

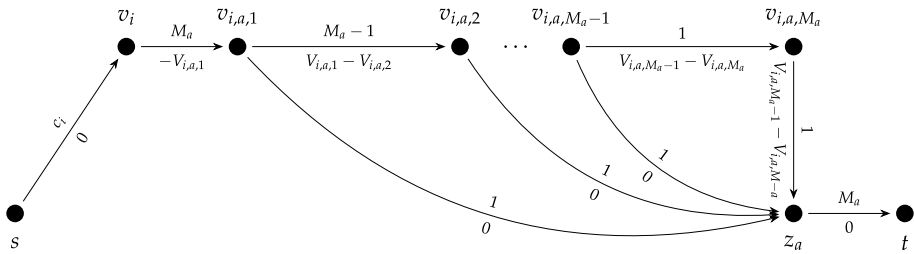


Fig. 2 Graph component between each pair  $(i, a) \in \mathcal{N} \times \mathcal{M}$

is equal to minus the utility that agent  $i$  gets from  $w$  copies of item  $a$ . This helps us find the optimal matching with the min-cost flow solution. The details of this reduction are presented in the rest of this section. Since the min-cost flow problem can be solved in polynomial time via linear programming (and also using various other algorithms [2]), we have the following property.

**Lemma 5.2** *For capacity vector  $C$ , supply vector  $M$  and value matrix  $V$ , the min-cost flow instance defined in Definition 5.1 has an integral solution which we can find in polynomial time.*

Our next lemma provides a connection between the solution of the min-cost flow instance of Definition 5.1 and the social welfare of the corresponding allocation for the instance of our problem.

**Lemma 5.3** *The absolute value of the minimum-cost flow in  $G(C, M, V)$  is equal to the maximum social welfare of an allocation of items to agents with respect to values in  $V$ .*

**Proof** Let  $X^*$  be an allocation with the maximum social welfare w.r.t.  $V$ , and  $F$  be a min-cost flow in  $G(C, M, V)$ . For  $v, v' \in G$ , let  $F(v, v')$  be the flow from node  $v$  to node  $v'$ , and  $\text{Cost}(F)$  the cost of  $F$ . We will show that  $\text{sw}(X^*, V) = -\text{Cost}(F)$  by bounding  $\text{Cost}(F)$  from above and below by  $-\text{sw}(X^*, V)$ .

To prove  $\text{Cost}(F) \leq -\text{sw}(X^*, V)$ , we use  $X^*$  to construct a flow  $F^*$  as follows:

- $F^*(v_i, v_{i,a,1}) = x_i^*(a)$  for  $i \in \mathcal{N}$  and  $a \in \mathcal{M}$ ;
- $F^*(v_{i,a,j}, v_{i,a,j+1}) = x_i^*(a) - j$  for  $j \in [x_i^*(a) - 1]$ ;
- $F^*(v_{i,a,j}, v_a) = 1$  for  $j \in [x_i^*(a)]$ ;
- $F^*(s, v_i) = \sum_{a \in \mathcal{M}} x_i^*(a)$  for  $i \in \mathcal{N}$ ;
- $F^*(v_a, t) = \sum_{i \in \mathcal{N}} x_i^*(a)$  for each  $a \in \mathcal{M}$ .

Observe that these flows satisfy the capacities of the edges, and thus  $F^*$  is valid. Since  $F$  is the cost-minimizing flow, we have

$$\begin{aligned}
 \text{Cost}(F) &\leq \text{Cost}(F^*) \\
 &= \sum_{i \in \mathcal{N}} \sum_{a \in \mathcal{M}} -V_{i,a,1} \cdot F^*(v_i, v_{i,a,1}) + \sum_{i \in \mathcal{N}} \sum_{a \in \mathcal{M}} \sum_{j=1}^{x_i^*(a)-1} (V_{i,a,j} - V_{i,a,j+1}) \cdot F^*(v_i, v_{i,a,1}) \\
 &= \sum_{i \in \mathcal{N}} \sum_{a \in \mathcal{M}} -V_{i,a,1} \cdot x_i^*(a) + \sum_{i \in \mathcal{N}} \sum_{a \in \mathcal{M}} \sum_{j=1}^{x_i^*(a)-1} (V_{i,a,j} - V_{i,a,j+1}) \cdot (x_i^*(a) - j) \\
 &= \sum_{i \in \mathcal{N}} \sum_{a \in \mathcal{M}} -V_{i,a,x_i^*(a)} + \sum_{i \in \mathcal{N}} \sum_{a \in \mathcal{M}} \sum_{j=1}^{x_i^*(a)-1} V_{i,a,j} ((x_i^*(a) - j) - (x_i^*(a) - (j - 1))) \\
 &= \sum_{i \in \mathcal{N}} \sum_{a \in \mathcal{M}} -V_{i,a,x_i^*(a)} + \sum_{i \in \mathcal{N}} \sum_{a \in \mathcal{M}} \sum_{j=1}^{x_i^*(a)-1} -V_{i,a,j} \\
 &= \sum_{i \in \mathcal{N}} \sum_{a \in \mathcal{M}} \sum_{j=1}^{x_i^*(a)} -V_{i,a,j} = -\text{sw}(X^*, V).
 \end{aligned}$$

Next, to prove  $\text{Cost}(F) \geq -\text{sw}(X^*, V)$ , we use  $F$  to construct an allocation  $X$  with  $x_i(a) = F(v_i, v_{i,a,1})$ . Doing similar calculations as above, and since  $X^*$  is the welfare-maximizing allocation, we now have that

$$\text{Cost}(F) = \sum_{i \in \mathcal{N}} \sum_{a \in \mathcal{M}} \sum_{j=1}^{x_i(a)} -V_{i,a,j} = -\text{sw}(X, V) \geq -\text{sw}(X^*, V).$$

This completes the proof. □

Note that if we had complete access to the utilities (rather than access to threshold approvals), we could use the same methods to find the optimal solution. The size of the graph depends on sum of the capacities of the agents and number of items, which means with general capacities the problem is pseudo-polynomial. However, with constant capacities, the graph will have  $O(n^2)$  nodes and the problem is solvable in polynomial time. We are now ready to present our deterministic mechanism  $g_t$ , which is a generalization of the deterministic mechanism  $f_t$  that we used for the one-sided matching setting.

**Definition 5.4** For  $\delta > 1$  and  $t \in [n]$ , we define the threshold vector  $\vec{\tau} = (\delta^{-1}, \delta^{-2}, \dots, \delta^{-t})$ . The *deterministic generalized matching mechanism*  $g_t$  uses this threshold vector  $\vec{\tau}$  and gets as input a profile  $S$ , constant agent capacities  $\{c_1, \dots, c_n\}$ , and constant item supplies  $\{m_1, \dots, m_m\}$ . The mechanism defines the vector  $C = \langle c_1, \dots, c_n \rangle$ , the vector  $M = \langle m_1, \dots, m_m \rangle$ , and the matrix  $V$  as follows: For every  $i \in \mathcal{N}$ ,  $a \in \mathcal{M}$  and  $j \in [\min\{c_i, m_a\}]$ , if  $(a, j) \in S_{i,k}$  for some  $k \in [t]$ , then the mechanism defines  $V_{i,a,j} = \tau_k$ ; otherwise, if  $(a, j) \notin \bigcup_{k=1}^t S_{i,k}$ , then it defines  $V_{i,a,j} = 0$ . The mechanism computes the solution of the min-cost flow instance defined in Definition 5.1 with input  $C, M$  and  $V$ . For each agent  $i \in \mathcal{N}$  and item  $a \in \mathcal{M}$  the flow from  $v_i$  to  $v_{i,a,1}$  in the computed solution is the number of copies of item  $a$  that agent  $i$  receives.

Before we bound the distortion of the mechanism, we prove a technical lemma similar to Lemma 4.4 which provides us with a lower bound on the social welfare of the allocation computed by the mechanism in relation to the optimal social welfare.

**Lemma 5.5** *If there is an allocation with social welfare  $s^*$ , then  $g_t$  outputs an allocation with social welfare at least  $\delta^{-1}(s^* - T \cdot \tau_t)$ .*

**Proof** Let  $X^*$  be an optimal allocation. We have

$$sw(X^*) \leq \sum_{i \in \mathcal{N}} \sum_{k \in [t]} \sum_{(a,j) \in S_{i,k}} u_i(a, j) + T \cdot \tau_t.$$

Recall that, if  $(a, j) \in S_{i,k}$  for some  $k$ , then  $u_i(a, j) \leq \delta \cdot V_{i,a,j}$ . This implies

$$sw(X^*) \leq \delta \cdot \sum_{i \in \mathcal{N}} \sum_{k \in [t]} \sum_{(a,j) \in S_{i,k}} V_{i,a,j} + T \cdot \tau_t.$$

Consequently, with respect to  $V$ ,  $X^*$  has a social welfare of at least  $\delta^{-1}(sw(X^*) - T \cdot \tau_t)$ . This is a lower bound on the social welfare of the allocation computed by  $g_t$ , since, by Lemma 5.3, this is at least the social welfare of the allocation with maximum social welfare with respect to  $V$ . □

We are now ready to show the upper bound on the distortion of  $g_t$ .

**Theorem 5.6** *For  $t \in [T]$  and  $\delta = \sqrt[t]{2T}$ , the distortion of the deterministic generalized matching mechanism  $g_t$  is  $O(c \cdot \sqrt[t]{T})$ , where  $c = \max_{i \in \mathcal{N}} \{c_i\}$ .*

**Proof** The structure of the proof is very similar to that of Theorem 4.5. Let  $X = g_t(S)$  be the allocation computed by  $g_t$  when given as input an arbitrary input profile  $S$  that is induced by some consistent utility profile  $\vec{u}$ . Clearly, for any  $i \in \mathcal{N}$ ,  $a \in \mathcal{M}$ , and  $j \in [\min\{c_i, m_a\}]$ ,  $u_i(a, j) \geq V_{i,a,j}$  since  $V_{i,a,j}$  is defined as a lower bound on this utility. By this and Lemma 5.3, if the graph  $G(C, M, V)$  admits a flow of cost  $W$ , then the social welfare of the corresponding allocation is at least  $|W|$ .

We now argue that the mechanism outputs an allocation with social welfare at least  $\delta^{-1}/(2c)$ . To show this, consider any agent  $i$ . Since  $\tau_t = \delta^{-t} = 1/(2T)$ , the sum of the utilities of agent  $i$  for the pairs  $(a, j)$  that he does not include in any approval set is at most  $T \times 1/(2T) = 1/2$ . Moreover, by the definition of  $\bar{\tau}$ ,  $u_i(a, j) \leq \delta \cdot V_{i,a,j}$  for every  $(a, j) \in \bigcup_{k=1}^t S_{i,k}$ , and thus, for each agent  $i$ , we have

$$\sum_{k \in [t]} \sum_{(a,j) \in S_{i,k}} u_i(a, j) \geq 1/2 \implies \sum_{k \in [t]} \sum_{(a,j) \in S_{i,k}} V_{i,a,j} \geq \delta^{-1}/2. \tag{2}$$

Next, we find a partial allocation  $B^*$  with social welfare at least  $\delta^{-1}/(2c)$ . To do so, we bundle all copies of each item together and assign them all to a single agent (up to the capacity of the agent). In particular, construct a bipartite graph  $G'$  with a node  $v_i$  for each agent  $i \in \mathcal{N}$

and a node  $z_a$  for each item  $a \in \mathcal{M}$ . There is an edge of weight  $\max_{j \in [\min(c_i, m_a)]} V_{i,a,j}$  between  $v_i$  and  $z_a$ . Equation 2 implies that

$$\sum_{k \in [t]} \sum_{(a,j) \in S_{i,k}} w(v_i, z_a) \geq \frac{\delta^{-1}}{2c}.$$

Furthermore, for each edge  $(v_i, z_a)$  in this bipartite graph,  $w(v_i, z_a) > \tau_t = 1/(2T)$ . Hence, by Lemma 4.3,  $G'$  has a matching of weight  $\min\{\delta^{-1}/(2c), n/(2T)\}$ . Since

$$\frac{n}{2T} = \frac{n}{2 \sum_{i \in \mathcal{N}} c_i} \geq \frac{n}{2nc} = \frac{1}{2c} \geq \frac{\delta^{-1}}{2c},$$

we finally conclude that assigning all copies of an item to its matched agent in  $G'$  gives us a social welfare of at least  $\delta^{-1}/(2c)$ .

By Lemma 5.5, since  $T \cdot \tau_t = 1/2$ , we also have a lower bound of  $\delta^{-1}(s^* - 1/2)$  on the social welfare of the allocation  $X$  computed by the mechanism, where  $s^*$  is the optimal social welfare. So, overall, we have established that

$$\text{sw}(X) \geq \max \left\{ \delta^{-1}/(2c), \delta^{-1}(s^* - 1/2) \right\},$$

and the distortion is at most

$$\frac{2\delta \cdot s^*}{\max \{1/c, 2s^* - 1\}}.$$

If  $s^* \geq 1$ , then  $2s^* - 1 \geq s^*$ , and hence the distortion is at most  $2\delta \in O(\sqrt[t]{T})$ . Otherwise, if  $s^* < 1$ , then the distortion is at most  $2\delta c \in O(c \cdot \sqrt[t]{T})$ .  $\square$

In many applications, the capacities and the supplies are constants. In this case, we have that  $T = \Theta(n) = \Theta(m)$  and the following result, which is tight given the corresponding lower bound in Section 3.

**Corollary 5.7** *When the capacities and supplies are constant, for any  $t \in [T]$ , there is a deterministic mechanism with distortion  $O(\sqrt[t]{n})$ .*

We also generalize our randomized mechanism to achieve a slightly better distortion bound.

**Definition 5.8** The *generalized randomized matching mechanism*  $GR_t$  works as follows: With probability  $1/2$ , it bundles the copies of each item together and selects a matching of the items to the agents uniformly at random; once the matching has been chosen, it assigns all copies of an item to its matched agent, subject to capacity and supply constraints. With the remaining  $1/2$  probability, it runs the deterministic mechanism  $g_t$  with threshold vector  $\vec{\tau} = (\delta^{-1}, \delta^{-2}, \dots, \delta^{-t})$  for  $t \in [n]$  and some  $\delta > 1$ .

The proof of the next theorem follows along the lines of the proof of Theorem 4.8.

**Theorem 5.9** For  $t \in [T]$  and  $\delta = \sqrt[t+1]{2T}$ , the distortion of the generalized randomized matching mechanism  $GR_t$  is  $O(c \cdot \sqrt[t+1]{T})$ , where  $c = \max\{n, m\}/n$ .

**Proof** The structure of this proof is similar to that of Theorem 4.8. Let  $X^*$  be an optimal allocation with social welfare  $s^*$ , and  $X^2$  the allocation computed by the first part of the mechanism (the outcome of  $g_t$ ).

In the first part of the mechanism (where a random matching between agents and items is chosen with probability 1/2), each agent is matched to each item with probability at least  $1/\max\{n, m\}$ . Indeed, fix an agent  $i$  and an item  $a$ . If  $n \geq m$ , there are  $\binom{n}{m} \cdot m!$  different matchings of items to agents, and  $i$  gets  $a$  in  $\binom{n-1}{m-1} \cdot (m-1)!$  of them, leading to a probability of

$$\frac{\binom{n-1}{m-1} \cdot (m-1)!}{\binom{n}{m} \cdot m!} = \frac{\binom{n-1}{m-1}}{\frac{n}{m} \cdot \binom{n-1}{m-1} \cdot m} = \frac{1}{n}.$$

Otherwise, if  $m > n$ , there are  $\binom{m}{n} \cdot n!$  different matchings, and  $i$  gets  $a$  in  $\binom{m-1}{n-1} \cdot (n-1)!$  of them, leading to a probability of

$$\frac{\binom{m-1}{n-1} \cdot (n-1)!}{\binom{m}{n} \cdot n!} = \frac{\binom{m-1}{n-1}}{\frac{m}{n} \cdot \binom{m-1}{n-1} \cdot n} = \frac{1}{m}.$$

Since the sum of the utilities of each agent when given all items up to capacity is 1, the expected social welfare from the first part is at least

$$\frac{1}{2} \sum_{i \in \mathcal{N}} \sum_{a \in \mathcal{M}} \frac{1}{\max\{n, m\}} \cdot u_i^+(a, \min\{c_i, m_a\}) = \frac{n}{2 \cdot \max\{n, m\}}.$$

For the second part (where the deterministic mechanism  $g_t$  using  $\vec{\tau}$  is employed), since  $s := T \cdot \tau_t = T \cdot \delta^{-t} = \sqrt[t+1]{T}$ , by Lemma 5.5,  $g_t$  outputs an allocation with social welfare at least  $\delta^{-1}(s^* - s)$ , and thus the expected social welfare of the mechanism from the second part is at least  $\frac{1}{2} \cdot \delta^{-1}(s^* - s)$ .

Overall, we have established that

$$\begin{aligned} \mathbb{E}_{X \sim GR_t(S)}[\text{sw}(X)] &\geq \frac{n}{2 \cdot \max\{n, m\}} + \frac{1}{2} \cdot \delta^{-1}(s^* - s) \\ &\geq \frac{1}{2} \cdot \max \left\{ \frac{n}{\max\{n, m\}}, \delta^{-1}(s^* - s) \right\}, \end{aligned}$$

and thus the distortion is at most

$$\frac{2 \cdot s^*}{\max \left\{ \frac{n}{\max\{n, m\}}, \delta^{-1}(s^* - s) \right\}}.$$

If  $s^* \geq 2s$ , then  $s^* - s \geq s^*/2$  and the distortion is at most

$$\frac{2 \cdot s^*}{\delta^{-1}s^*/2} = 4\delta = 4 \cdot {}^{t+1}\sqrt{T}.$$

Otherwise, if  $s^* < 2s$ , the distortion is at most

$$2 \cdot \frac{\max\{n, m\}}{n} \cdot s^* \leq \frac{4 \max\{n, m\}}{n} \cdot {}^{t+1}\sqrt{T}.$$

In any case, the distortion is  $O(c \cdot {}^{t+1}\sqrt{T})$ , where  $c = \max\{n, m\}/n$ .  $\square$

Again, when the capacities and the supplies are constant, since the total capacity is of the same magnitude as the total supply, it follows that  $n$  and  $m$  are also of the same magnitude. Hence, the parameter  $c = \max\{n, m\}/n$  is a constant and  $T = \Theta(n) = \Theta(m)$ , giving us the following result, which is again tight due to the corresponding lower bound from Section 3.

**Corollary 5.10** *When the capacities and supplies are constant, for any  $t \in [T]$ , there is a randomized mechanism with distortion  $O({}^{t+1}\sqrt{n})$ .*

**Remark 5.11** In the generalized setting that we considered in this section, the agents are allowed to receive potentially all available copies of the items, up to their capacity. However, in several applications, we might want to disallow this and set a limit  $\ell_{i,a}$  on the number of copies of  $a \in \mathcal{M}$  that agent  $i \in \mathcal{N}$  can get. For example, in the paper assignment problem, each agent must be given at most one copy of each item since it does not make sense for someone to review a paper more than once. Such constraints can be handled in several ways. One of them is via the utility functions of the agents which, for scenarios like these, would simply assign a marginal value of 0 for any extra copy that exceeds the limit, that is,  $u_i(a, j) = 0$  for every  $j > \ell_{i,a}$ . If the utility function is not naturally defined this way, we can modify the min-cost flow instance by setting  $V_{i,a,j} = -\infty$  for  $j > \ell_{i,a}$ , or by removing the corresponding edges in the graph.

## 6 Conclusion and open problems

In this paper, we showed tight bounds on the best possible distortion of (both deterministic and randomized) mechanisms for matching settings (that capture important applications, including the one-sided matching problem and the paper assignment problem) when the given information about the preferences of the agents is of the form of threshold approvals.

Going forward, it would be interesting to explore whether improved tradeoffs can be achieved by using randomization not only during the decision phase of the mechanism but also for the definition of the threshold values, similarly to the works of [11] and [12]. Furthermore, one could explore other settings in which the same type of elicitation method can be applied, including various voting settings where the full potential of using multiple (as opposed to just one) threshold approvals has not been considered before, as well as other resource allocation settings, such as metric matching (which would correspond to the case

where the items are chores rather than goods). These would also be interesting to study in combination with other constraints, such as truthfulness or fairness.

**Acknowledgements** Mohamad Latifian was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grant EP/Y003624/1.

**Author Contributions** All authors contributed equally.

**Data Availability** No datasets were generated or analysed during the current study.

## Declarations

**Competing Interests** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Abramowitz, B., Anshelevich, E. (2018). Utilitarians without utilities: Maximizing social welfare for graph problems using only ordinal preferences. In: *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, pp. 894–901.
2. Ahuja, R K., Magnanti, T L, Orlin, J B. (1988). Network flows.
3. Amanatidis, G., Birmpas, G., Filos-Ratsikas, A., & Voudouris, A. A. (2021). Peeking behind the ordinal curtain: improving distortion via cardinal queries. *Artificial Intelligence*, 296, 103488
4. Amanatidis, G., Birmpas, G., Filos-Ratsikas, A., Voudouris, A A. (2022). A few queries go a long way: Information-distortion tradeoffs in matching. *Journal of Artificial Intelligence Research*, 74
5. The Two-Query Distortion of Matching Problems and Beyond. (2024). Georgios amanatidis, georgios birmpas, aris filos-ratsikas, and alexandros a voudouris. don't roll the dice, ask twice. *SIAM Journal on Discrete Mathematics*, 38, 1007–1029.
6. Anshelevich, E., Sekar, S. (2016). Blind, greedy, and random: Algorithms for matching and clustering using only ordinal information. In: *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*, pp. 390–396
7. Anshelevich, E., Bhardwaj, O., Elkind, E., Postl, J., & Skowron, P. (2018). Approximating optimal social choice under metric preferences. *Artificial Intelligence*, 264, 27–51.
8. Anshelevich, E., Filos-Ratsikas, A., Shah, N., Voudouris, A A. (2021). Distortion in social choice problems: The first 15 years and beyond. In: *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 4294–4301.
9. Anshelevich, E., Filos-Ratsikas, A., Jerrett, C., & Voudouris, A. A. (2025). Improved metric distortion via threshold approvals. *Artificial Intelligence*, 341, Article 104295.
10. Aziz, H. (2019). Justifications of welfare guarantees under normalized utilities. *SIGecom Exchanges*, 17(2), 71–75.
11. Benadè, G., Nath, S., Procaccia, A. D., & Shah, N. (2021). Preference elicitation for participatory budgeting. *Management Science*, 67(5), 2813–2827.
12. Bhaskar, U., Dani, V., Ghosh, A. (2018). Truthful and near-optimal mechanisms for welfare maximization in multi-winner elections. In: *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, pp. 925–932
13. Boutilier, C., Caragiannis, I., Haber, S., Tyler, L., Procaccia, A. D., & Sheffet, O. (2015). Optimal social choice functions: a utilitarian view. *Artificial Intelligence*, 227, 190–213.

14. Burkhardt, J., Caragiannis, I., Fehrs, K., Russo, M., Schwiegelshohn, C., Shyam, S. (2024). Low-distortion clustering with ordinal and limited cardinal information. In: *Proceedings of the 38th AAAI Conference on Artificial Intelligence (AAAI)*, pp. 9555–9563
15. Caragiannis, I., Fehrs, K. (2023). Beyond the worst case: Distortion in impartial culture electorate. *CoRR*, arXiv:2307.07350
16. Caragiannis, I., & Procaccia, A. D. (2011). Voting almost maximizes social welfare despite limited communication. *Artificial Intelligence*, 175(9–10), 1655–1671.
17. Caragiannis, I., Nath, S., Procaccia, A. D., & Shah, N. (2017). Subset selection via implicit utilitarian voting. *Journal of Artificial Intelligence Research*, 58, 123–152.
18. Caragiannis, I., Shah, N., & Voudouris, A. A. (2022). The metric distortion of multiwinner voting. *Artificial Intelligence*, 313, Article 103802.
19. Charikar, M., Ramakrishnan, P. (2022). Metric distortion bounds for randomized social choice. In: *Proceedings of the 2022 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 2986–3004
20. Charikar, M., Ramakrishnan, P., Wang, K., Wu, H. (2024). Breaking the metric voting distortion barrier. *Journal of the ACM*, 71 (6), 42:1–42:33,
21. Ebadian, S., Latifian, M., Shah, N. (2023). The distortion of approval voting with runoff. In: *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pp. 1752–1760
22. Ebadian, S., Kahng, A., Peters, D., Shah, N. (2024) Optimized distortion and proportional fairness in voting. *ACM Transactions on Economics and Computation*, 12 (1): 3:1–3:39.
23. Filos-Ratsikas, A., Frederiksen, S. K. S., Zhang, J. (2014) Social welfare in one-sided matchings: Random priority and beyond. In: *Proceedings of the 7th Symposium of Algorithmic Game Theory (SAGT)*, pp. 1–12.
24. Gkatzelis, V., Halpern, D., Shah, N. (2020). Resolving the optimal metric distortion conjecture. In: *Proceedings of the 61st IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 1427–1438.
25. Gkatzelis, V., Latifian, M., Shah, N. (2023). Best of both distortion worlds. In: *Proceedings of the 24th ACM Conference on Economics and Computation, (EC)*, pp. 738–758.
26. Kempe, D. (2020). Communication, distortion, and randomness in metric voting. In: *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI)*, pp. 2087–2094.
27. Kizilkaya, F E., Kempe, D. (2022). Plurality veto: A simple voting rule achieving optimal metric distortion. In: *Proceedings of the 31st International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 349–355.
28. Ma, T., Menon, V., Larson, K. (2021). Improving welfare in one-sided matchings using simple threshold queries. In: *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 321–327.
29. Mandal, D., Procaccia, A. D., Shah, N., Woodruff, D P. (2019). Efficient and thrifty voting by any means necessary. In: *Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS)*, pp. 7178–7189.
30. Mandal, D., Shah, N., Woodruff, D P. (2020). Optimal communication-distortion tradeoff in voting. In: *Proceedings of the 21st ACM Conference on Economics and Computation (EC)*, pp. 795–813.
31. Procaccia, A. D., Rosenschein, J. S. (2006). The distortion of cardinal preferences in voting. In: *International Workshop on Cooperative Information Agents (CIA)*, pp. 317–331.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.