Efficient, Fair, and Incentive-Compatible Healthcare Rationing

Haris Aziz[†] Florian Brandl[‡]

As the COVID-19 pandemic shows no clear signs of subsiding, fair and efficient rationing of healthcare resources has emerged as an important issue that has been discussed by medical experts, policy-makers, and the general public. We consider a healthcare rationing problem where medical units are to be allocated to patients. Each unit is reserved for one of several categories and the patients have different priorities for the categories. We present an allocation rule that respects the priorities, complies with the eligibility requirements, allocates the largest feasible number of units, and does not incentivize agents to hide that they qualify through a category. Moreover, the rule is polynomial-time computable. To the best of our knowledge, it is the first known rule with the aforementioned properties.

Keywords: Allocation under priorities, healthcare rationing, assignment maximization.

1 Introduction

The COVID-19 pandemic has emerged as one of the biggest challenges the world has faced. It has resulted in a frantic scientific race to produce the most effective and safe vaccine to stem the devastating effects of the pandemic. Whereas there is encouraging initial news on the creation of vaccines, there are still several scientific challenges on how to distribute, allocate, and administer them in an efficient and fair manner.

Since healthcare resources such as ventilators, antiviral treatments, and vaccines can be scarce or costly, a fundamental question that arises is who to prioritize when making allocation decisions. For example, three important priority groups that are highlighted by medical practitioners and policy-makers are (1) health care workers; (2) other essential workers and people in high-transmission settings; (3) and people with medical vulnerabilities associated with poorer

[†]UNSW Sydney and Data61 CSIRO, Australia, haris.aziz@unsw.edu.au
[‡]Princeton University, USA, brandl.ffx@gmail.com

COVID-19 outcomes (Persad et al., 2020; Truog et al., 2020). Other concerns that have been discussed include racial equity (Bruce and Tallman, 2021).

When healthcare resources need to be allocated, it is not enough to identify priority groups. There is also a need to algorithmically and transparently make these prioritized allocations decisions (Emanuel et al., 2020; WHO, 2020). In a New York Times article, the issue has been referred to as the one of the hardest decisions, health organizations need to make (Fink, 2020). Since the decisions need to be justified to the public, they must be aligned with the ethical guidelines such as respecting priorities of various categories. These decisions are not straightforward, especially when a patient is eligible for more than one category. When patients are eligible for multiple categories, the decision on which category is used can have compounding effects on what categories other agents can use. A fundamental question that arises is the following one:

> *How do we allocate scarce medical resources fairly and efficiently while taking into account various ethical principles and priority groups?*

The question is not just fundamental but the solution to the problem is time–critical as various states, city councils, and municipalities start to roll out vaccines using their particular ethical guidelines. The problem of health care rationing has recently been formally studied by market designers. Pathak et al. (2020) were among the first to frame the problem as a two-sided matching problem in which patients are on one side and the resources units are on the other side. By doing so, they linked the healthcare rationing problem with the rich field of two-sided matching (Roth and Sotomayor, 1990).

Pathak et al. suggested dividing the units into different reserve categories, each with its own priority ranking of the patients. The categories and the category-specific priorities represent the ethical principles and guidelines that a policy-maker may wish to implement.[1] For example, a category for senior people may have an age-specific priority ranking that puts the eldest citizens first. Having a holistic framework that considers different types of priorities has been termed important in healthcare rationing .[2] The approach of Pathak et al. has been recommended or adopted by various organizations including the NASEM (National Academies of Sciences, Engineering, and Medicine) Framework for Equitable Vaccine Allocation (NASEM-National Academies of Sciences, Engineering, and Medicine, 2020) and has been endorsed in medical circles (Persad et al., 2020; Sönmez et al., 2020). The approach has been covered widely in the media, including the New York Times and the Washington Post.[3]

For their two-sided matching formulation, Pathak et al. (2020) proposed a solution for the problem. One of their key insights was that running the Deferred Acceptance algorithm (Gale

[1] See for example, the book by Bognar and Hirose (2014) on the ethics of healthcare rationing that discusses many of these principles.

[2] In a report issued by the Deeble Institute for Health Policy Research, Martin (2015) writes that "To establish robust healthcare rationing in Australia, decision-makers need to acknowledge the various implicit and explicit priorities that influence the process and develop a decision-making tool that incorporates them."

[3] https://www.covid19reservesystem.org/media

and Shapley, 1962; Roth, 2008) on the underlying problem satisfies basic relevant axioms (eligibility compliance, respect of priorities, and non-wastefulness). They also showed when all the category priorities are consistent with a global baseline priority, then there is a smart reserves algorithm that computes a maximum size matching satisfying the basic axioms. The smart reserves approach makes careful decisions about which category should be availed by which patient. However, the problem of such a smart reserves approach for the general problem with *heterogeneous* priorities has not been addressed in the literature. In this paper, we set out to address this issue and answer the following research problem.

> *For the general healthcare rationing problem with heterogeneous priorities, how do we allocate resources in a fair, economically efficient, strategyproof, and computationally tractable way?*

Contribution For the general healthcare rationing problem, we first highlight that naively ascribing strict preferences over the categories to the agents can have adverse effects on the efficiency of the outcome when patients are eligible for multiple categories. If the eligibility requirements are treated as hard constraints, it leads to inefficient allocation of resources. If the eligibility requirements are treated as soft constraints, then the outcome does not allocate the resources optimally to the highest priority patients, thereby undermining important healthcare guidelines and ethical principles.

We address the problem posed by Pathak et al. (2020), that of finding a maximum beneficiary assignment that respects priorities and other desirable axioms. Our main contribution is presenting an allocation rule that

(i) complies with the eligibility requirements

(ii) respects the priorities of the categories (for each category, patients of higher priority are served first)

(iii) is non-wasteful (there is no unit that is unused but could be used by some eligible patient)

(iv) yields a maximal beneficiary allocation (allocates the largest feasible number of units to patients who are eligible for the categories pertaining to the units)

(v) is strategyproof (does not incentivize agents to under-report the categories they qualify for)

(vi) is monotonic in the quotas (selected agents remain selected if the category quotas weakly increase) and

(vii) is strongly polynomial-time computable.

Our algorithm also immediately applies to the school choice problem in which students are only interested in being matched with one of their acceptable schools. It also applies to hiring settings in which applicants are interested in one of the positions and each of the departments has its own priorities. Finally, it applies to many other rationing scenarios such as allocation of limited slots at public events or visas to immigration applicants.

2 Related Work

The paper is related to an active area of research on matching with distributional constraints (see, e.g., Kojima, 2019). One general class of distributional constraints that have been examined in matching market design pertains to common quotas over institutions such as hospitals (Kamada and Kojima, 2015, 2017; Biró et al., 2010; Goto et al., 2016).

Within the umbrella of work on matching with distributional constraints, particularly relevant to healthcare rationing is the literature on school choice with diversity constraints (Ehlers et al., 2014; Echenique and Yenmez, 2015; Kurata et al., 2017; Aygün and Turhan, 2020; Aygün and Bó, 2020; Aziz et al., 2020; Gonczarowski et al., 2019). Categories in healthcare rationing correspond to affirmative action types in school choice. For a brief survey, we suggest the book chapter by Heo (2019). Except for the special case in which students have exactly one type (Ehlers et al., 2014), most of the approaches do not achieve diversity goals optimally whereas in the healthcare rationing problem that we focus on, our goal is to find matchings that are maximal in terms of beneficiary assignment. Ahmed et al. (2017), Dickerson et al. (2019), and Ahmadi et al. (2020) consider optimisation approaches for diverse matchings but their objective and models are different.

Pathak et al. (2020) were the first to frame a rationing problem with category priorities as a two-sided matching problem in which agents are simply interested in a unit of resource and the resources are reserved for different categories. They show that artificially enforcing strict preferences of the agents over the categories and running the deferred acceptance algorithm results in desirable outcomes for the rationing problem. They note, however, that this approach may lead to matchings that are not Pareto optimal. They then proposed to use the smart reserves approach of Sönmez and Yenmez (2020) for the restricted problem when all the categories have identical priorities. Our results can be viewed as simultaneously achieving the properties of the two approaches of Pathak et al.. Firstly, we propose a new algorithm that achieves the same properties for heterogeneous priorities as the smart reserves algorithm of Sönmez and Yenmez (2020) and Pathak et al. (2020) for homogeneous priorities. Secondly, our algorithm has an important advantage over the Deferred Acceptance formulation of Pathak et al. (2020) for the case of heterogeneous priorities as our approach additionally achieves the important property of maximality in beneficiary assignment. In followup work, Grigoryan (2020) considers optimisation approaches for variants of the problem but does not present any polynomial-time algorithm or consider incentive issues. In contrast to the papers on healthcare rationing discussed above, we also consider strategyproofness and monotonicity aspects and show that our rule complies with them.

In this paper, we attempt to compute what are essentially maximum size stable matchings. The problem of computing such matchings is NP-hard if both sides have strict preferences/priorities (Biró et al., 2010). In our problem, the agents essentially have dichotomous preferences (categories they are/are not eligible for) and, hence, we are able to obtain a polynomial-time algorithm for the problem.

Furthermore, our algorithm is strategyproof. In contrast, for other two-sided matching settings, it is known that maximizing the number of matched individuals results in incentive and fairness impossibilities (see, e.g., Afacan et al., 2020; Krysta et al., 2014). Computing outcomes that match as many agents as possible, has also been examined in related but different contexts (see, e.g., Aziz, 2018; Andersson and Ehlers, 2016; Abraham et al., 2007; Bogomolnaia and Moulin, 2015).

3 Model

We adopt the healthcare rationing model of Pathak et al. (2020) with one generalization: we allow the categories' priorities over agents to be weak rather than strict. There are q identical and indivisible units of some resource, which are to be allocated to the agents in a set N with $|N| = n$. Each category c has a quota $q_c \in \mathbb{N}$ with $\sum_{c \in C} q_c = q$ and a priority ranking $\succsim_c$, which is a preorder on $N \cup \{\emptyset\}$. An agent i is eligible for category c if $i \succ_c \emptyset$. We say that $I = (N, C, (\succsim_c), (q_c))$ is an instance (of the rationing problem). (For convenience, we will write $(\succsim_c)$ and (q_c) for the profile of priorities and quotas in the sequel.)

A matching $\mu\colon N \to C \cup \{\emptyset\}$ is a function that maps each agent to a category or to $\emptyset$ and satisfies the capacity constraints: for each $c \in C$, $|\mu^{-1}(c)| \leq q_c$. For an agent $i \in N$, $\mu(i) = \emptyset$ means that i is unmatched (that is, does not receive any unit) and $\mu(i) = c$ means that i receives a unit reserved for category c. When convenient, we will identify a matching μ with the set of agent-category pairs $\{\{i, \mu(i)\}\colon \mu(i) \neq \emptyset\}$.[4]

We introduce four axioms that are natural in the context of allocating medical units and well-grounded in practice. For further motivation of these axioms, we recommend the detailed discussions by Pathak et al. (2020).

The first axiom we consider requires that matchings comply with eligibility requirements. It specifies that a patient should only take a unit of a category for which the patient is eligible. For example, a young person should not take a unit from the units reserved for elderly people.

Definition 1 (Compliance with eligibility requirements)**.** A matching μ *complies with eligibility requirements* if for any $i \in N$ and $c \in C$, $\mu(i) = c \implies i \succ_c \emptyset$.

The second axiom concerns the respect of priorities of categories. It rules out that a patient is matched to a unit of some category c while some other agent with a higher priority for c is unmatched.

Definition 2 (Respect of priorities)**.** A matching μ *respects priorities* if for any $i, j \in N$ and $c \in C$, $\mu(i) = c$ and $\mu(j) = \emptyset \implies j \not\succ_c i$.

[4] In graph theoretic terms, μ is a b-matching because multiple edges in μ can be adjacent to a category c.

An astute reader who is familiar with the theory of stable matchings will immediately realise that the axiom "respect of priorities" is equivalent to *justified envy-freeness* in the context of school-choice matchings (Abdulkadiroğlu and Sönmez, 2003).

Next, non-wastefulness requires that if an agent is unmatched despite being eligible for a category, then all units reserved for that category are matched to other agents.

Definition 3 (Non-wastefulness)**.** A matching μ is *non-wasteful* if for any $i \in N$ and $c \in C$, $i \succ_c \emptyset$ and $\mu(i) = \emptyset \implies |\mu^{-1}(c)| = q_c$.

Not all non-wasteful matchings allocate the same number of units. In particular, some may not allocate as many units as possible. A stronger efficiency notion prescribes that the number of allocated units is maximal subject to compliance with the eligibility requirements.

Definition 4 (Maximal beneficiary assignment)**.** A matching μ is a *maximal beneficiary assignment* if it has maximum size among all matchings complying with eligibility requirements.

These four axioms capture the first guideline put forth in the report by the National Academies of Sciences, Engineering, and Medicine: "ensure that allocation maximizes benefit to patients, mitigates inequities and disparities, and adheres to ethical principles" (NASEM-National Academies of Sciences, Engineering, and Medicine, 2020, page 69)

The following example illustrates the definitions above.

Example 1. Suppose there are three agents and two categories with one reserved unit each.

$$N = \{1, 2, 3\}, \quad C = \{c_1, c_2\}, \quad q_{c_1} = 1, q_{c_2} = 1.$$

The priority ranking of c_1 is $2 \succ_{c_1} 3 \succ_{c_1} \emptyset \succ_{c_1} 1$ and the priority ranking of c_2 is $2 \succ_{c_2} \emptyset \succ_{c_2} 1 \succ_{c_2} 3$. Figure 1 illustrates this instance of the rationing problem.

Note that agent 1 is not eligible for any category, agent 2 is eligible for c_1 and c_2, and agent 3 is eligible only for c_1. Thus, the following matchings comply with the eligibility requirements.

$$\mu_1 = \emptyset \qquad \mu_2 = \{\{2, c_1\}\} \qquad \mu_3 = \{\{2, c_2\}\}$$
$$\mu_4 = \{\{3, c_1\}\} \qquad \mu_5 = \{\{2, c_2\}, \{3, c_1\}\}$$

All of these matchings except for μ_4 respect priorities. Only μ_2 and μ_5 are non-wasteful. The only matching that is a maximal beneficiary assignment is μ_5.

We are interested in allocation rules, which, for each instance, return a matching.

Definition 5 (Allocation rule)**.** An allocation rule maps every instance I to a matching for I.

We say that an allocation rule f satisfies one of the axioms in Definitions 1 to 4 if $f(I)$ satisfies the axiom for all instances I. Moreover, we define a notion of strategyproofness for allocation rules. Note that all units are identical and agents have no preferences over the category of the

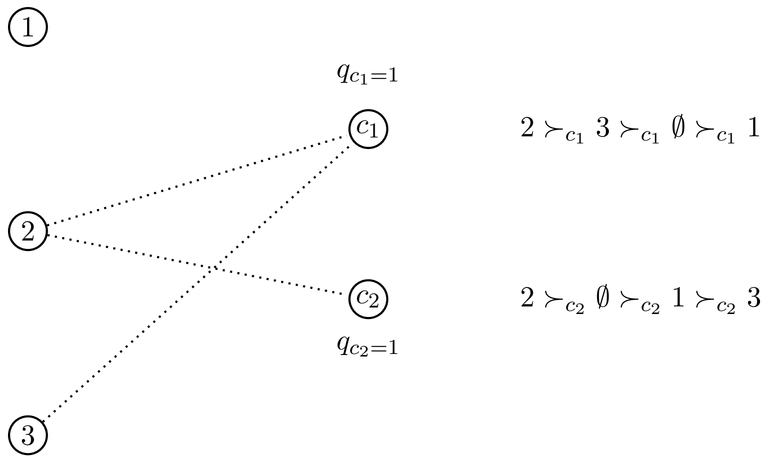

Figure 1: The problem instance described in Example 1. A dotted line between an agent and a category indicates that the agent is eligible for the category.

unit they receive However, they may have an incentive to hide being eligible for a category, or, more generally, to aim for a lower priority for some category.[5]

Formalizing strategyproofness requires the following definition. Let $(\succsim_c)$ and $(\succsim'_c)$ be priority profiles and $i \in N$. We say agent i's priority decreases from $(\succsim_c)$ to $(\succsim'_c)$ if for all $j, k \neq i$ and $c \in C$,

$$j \succsim_c k \longleftrightarrow j \succsim'_c k$$
$$j \succsim_c i \longrightarrow j \succsim'_c i \text{ and } j \succ_c i \longrightarrow j \succ'_c i$$

That is, the priority rankings over agents other than i are the same in both profiles and i can only move down in the priority rankings from $(\succsim_c)$ to $(\succsim'_c)$. We also say that i's priority decreases from $I = (N, C, (\succsim_c), (q_c))$ to $I' = (N, C, (\succsim'_c), (q_c))$. Strategyproofness requires that if i is unmatched for I, then i is also unmatched for I'.

Definition 6 (Strategyproofness)**.** An allocation rule f is strategyproof if $f(I)(i) = \emptyset$ implies $f(I')(i) = \emptyset$ whenever i's priority decreases from I to I'.

In particular, with a strategyproof allocation rule, agents cannot benefit from hiding that they are eligible for a category.

Finally, we define a monotonicity axiom that requires that matched agents remain matched if the categories' quotas weakly increase.

Definition 7 (Monotonicity in quotas)**.** An allocation rule f is monotonic in quotas if for any two instances I and I' with quotas (q_c) and (q'_c) so that $q_c \leq q'_c$ for all $c \in C$, every agent who is matched in $f(I)$ is also matched in $f(I')$.

[5]In the context of school choice, lowering oneself in the priority ranking of a school is akin to students deliberately underperformining in an entrance exam.

4 Issues with Breaking Ties and Applying the General Solution of Pathak et al. (2020)

The approach of Pathak et al. (2020) is to frame the healthcare rationing problem as a two-sided matching problem. They showed that if one artificially introduces (strict) preferences for the agents over the categories they are eligible for and applies the Deferred Acceptance algorithm, the resulting matching complies with the eligibility requirements, respects priorities, and is non-wasteful (Pathak et al., 2020, Theorem 2). They state the algorithmic implications of their result.

> *"Not only is this result a second characterization of matchings that satisfy our three basic axioms, it also provides a concrete procedure to calculate all such matchings."*

Although considering all possible artificial preferences and running Deferred Acceptance gives us all the matchings satisfying the three axioms, it is computationally expensive to consider $|C|!^{|N|}$ different preference profiles.

Not every preference profile leads to a compelling outcome even if the categories have strict priorities. For example, many preference profiles lead to matchings that are not maximal beneficiary assignments. Our next example highlights this issue.

Example 2. Consider the instance in Example 1. Suppose we run the Deferred Acceptance algorithm assuming all agents prefer c_1 to c_2 to c_3. Assuming agents can only be matched to categories they are eligible for (compliance with eligibility requirements), the resulting matching is $\mu_2 = \{\{2, c_1\}\}$. This matching is however not the most efficient use of the resources because it is possible to allocate all units while still satisfying the axioms in Definitions 1 to 3: $\mu_5 = \{\{3, c_1\}, \{2, c_2\}\}$.

Hence, artificially inducing preferences of agents and running Deferred Acceptance can lead to inefficient allocations. Even if we ignore computational concerns and can assign preferences to agents so that the matching selected by Deferred Acceptance is of maximum size and respects priorities, it is not clear whether such a rule satisfies strategyproofness and monotonicity properties like the ones we introduced above. We propose a rule that circumvents both issues.

5 The Allocation Rule

Our allocation rule is based on an arbitrary permutation π of the agents called a baseline ordering. For an agent i, $\pi(i)$ is interpreted as the position of i in the baseline ordering. Given an instance $I = (N, C, (\succsim_c), (q_c))$, we construct the corresponding *reservation graph* $G_I^\pi = (N \cup C, E, w, (q_c))$, which is a bipartite, weighted graph. The two independent vertex sets are N and C and the edge set E consists of all agent-category pairs so that the agent is eligible for the category.

$$E = \{\{i, c\} : i \in N, c \in C, i \succ_c \emptyset\}$$

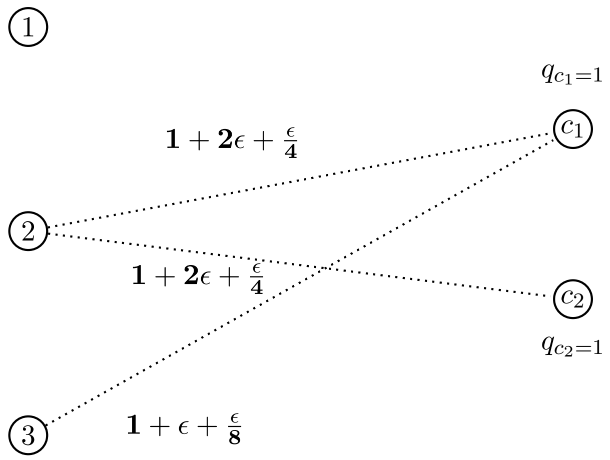

Figure 2: The reservation graph G_I^π for the instance I in Example 1. The dotted lines correspond to edges in E. The numbers close to the edges are their weights according to w.

The weight of an edge $\{i, c\} \in E$ is

$$w(\{i, c\}) = 1 + w_c(\{i, c\}) + w_\pi(\{i, c\})$$

where

(i) $w_c(\{i, c\}) = (n - \ell)\epsilon$ if i is in the ℓth most-preferred equivalence class in c's priority ranking, and

(ii) $w_\pi(\{i, c\}) = \frac{\epsilon}{2^{\pi(i)}}$,

for some $\epsilon < \frac{1}{n^2}$. The reservation graph for the instance in Example 1 is depicted in Figure 2 where the baseline ordering is chosen as 123.

We say that a subset of edges M is a *matching* of G_I^π if each $i \in N$ is adjacent to at most one edge in M and each $c \in C$ is adjacent to at most q_c edges in M. Hence, each matching μ of I induces a matching $M_\mu = \{\{i, c\} \in E \colon \mu(i) = c\}$ of G_I^π and vice versa. We will not distinguish between μ and M_μ in the sequel.

The intuition behind the edge weights is as follows. We want to find a maximum size matching of G_I^π. Among the maximum size matchings, we want to restrict attention to those that respect priorities. Finally, among the latter set of matchings, we want to prioritize the agents who are ranked early in the baseline ordering π (to achieve strategyproofness and monotonicity). This can be done as follows. We first care about whether a category finds an agent acceptable, in which case there is an edge with weight at least 1. Secondly, we give additional weight if an agent is high in the priority ranking of the category. Finally, matching agents who come early in the baseline ordering gives more weight than matching agents who come late. By the choice of ϵ, the weight a matching can obtain through w_c and w_π is always less than 1. Hence, ordering matchings by weight refines the order induced by size. Similarly, w_π is chosen so that the sum of weights a matching can get through w_π is less than ϵ (which is smallest possible non-zero difference in the weight of any two matchings according to w_c). Thus, matching agents to categories for which they have high priority is prioritized over matching agents who come early in the baseline ordering.

Consider an allocation rule that chooses a maximal weight matching of G_I^π for every instance I (with π fixed). The following three lemmas show that any such allocation rule satisfies *all* of the axioms defined above. By construction of the reservation graph, each maximal weight matching complies with the eligibility requirements, is a maximum size matching, and respects priorities. We verify these facts in Lemma 1. Lemma 2 and Lemma 3 show that the above allocation rules are strategyproof and monotonic, with the former taking up the bulk of the work.

Lemma 1. *Every maximal weight matching of G_I^π*

(i) complies with acceptability requirements,

(ii) is of maximum size among feasible matchings,

(iii) respects priorities, and

(iv) matches the same set of agents.

Proof. Let μ be a maximal weight matching of G_I^π.

(i) If $\{i, c\}$ is an edge in G_I^π, then $i \succ_c \emptyset$. Hence, any matching of G_I^π complies with the acceptability requirements. In particular, μ does.

(ii) For any edge $e \in E$,

$$w(e) \leq 1 + (n-1)\epsilon + \frac{\epsilon}{2} < 1 + \frac{1}{n}$$

since $\epsilon < \frac{1}{n^2}$ (recall the definition of w_c and w_π). Hence, $w(e) \in [1, 1 + \frac{1}{n})$. Moreover, every matching μ' of G_I^π has cardinality at most n as $|\mu'| \leq n = |N|$ since every agent can be matched to at most one category. Hence, $w(\mu') = \sum_{e \in \mu'} w(e) \in [|\mu'|, |\mu'| + 1)$. So $w(\mu) \geq w(\mu')$ implies $|\mu| \geq |\mu'|$. Hence, μ is a maximal size matching (and, thus, also non-wasteful).

(iii) If μ does not respect priorities, there exist $i, j \in N$ and $c \in C$ so that $\mu(i) = c$, $\mu(j) = \emptyset$, and $j \succ_c i$. In particular, $l_j \leq l_i - 1$, where l_j is the rank of j's equivalence class in the priority ranking of c (and similarly for i). Then

$$\begin{aligned} w(\{j, c\}) &= 1 + (n - l_j)\epsilon + \frac{\epsilon}{2^{\pi(j)}} \\ &> 1 + (n - l_j - 1)\epsilon + \epsilon \\ &\geq 1 + (n - l_i)\epsilon + \frac{\epsilon}{2^{\pi(i)}} = w(\{i, c\}) \end{aligned}$$

It follows that $\mu' = \mu - (\{\{i, c\}\} \cup \{\{j, c\}\}$ is a matching of G_I^π with $w(\mu') > w(\mu)$. Since this contradicts that μ has maximal weight, μ has to respect priorities.

(iv) We show that, more generally, any two matchings μ and μ' of G_I^π with $w(\mu) = w(\mu')$ match the same set of agents. In the proof of (ii), we have observed that $w(\mu) \in [|\mu|, |\mu| + 1)$

and $w(\mu') \in [|\mu'|, |\mu'| + 1)$. So the fact that $w(\mu) = w(\mu')$ implies $|\mu| = |\mu'|$. Recalling that $w(\mu) = |\mu| + w_c(\mu) + w_\pi(\mu)$ (and similarly for μ') gives

$$w_c(\mu) + w_\pi(\mu) = w_c(\mu') + w_\pi(\mu') \tag{1}$$

Now $w_c(\mu)$ and $w_c(\mu')$ are integer multiples of ϵ and $w_\pi(\mu) < \epsilon$ and $w_\pi(\mu') < \epsilon$. It follows from (1) that

$$w_\pi(\mu) = w_\pi(\mu') \tag{2}$$

Inserting the definition of w_π gives

$$\sum_{\{i,c\}\in\mu} \frac{\epsilon}{2^{\pi(i)}} = \sum_{\{i,c\}\in\mu'} \frac{\epsilon}{2^{\pi(i)}}$$

This equality holds if and only if $\{i \in N : \mu(i) \in C\} = \{i \in N : \mu'(i) \in C\}$. That is, if μ and μ' match the same set of agents. □

The proof of Lemma 2 will use some graph-theoretical terminology. Let $a_0, \ldots, a_m$ be a sequence of vertices in G_I^π. If $P = \{\{a_{l-1}, a_l\} : l \in \{1, \ldots, m\}\} \subset E$, we call P a path in G_I^π. The length of P is $|P|$ and a_0 and a_m are the initial and terminal vertex of P. If μ is a matching of G_I^π, P is an *alternating path* for μ if $\{a_{i-1}, a_i\} \in \mu$ if and only if i is even and either $a_0 \in N$ and a_0 is unmatched in μ or $a_0 \in C$ and $|\mu^{-1}(a_0)| < q_c$. If, moreover, $\sum_{e\in P\setminus\mu} w(e) > \sum_{e\in P\cap\mu} w(e)$, P is an *augmenting path* for μ in G_I^π.

Lemma 2. *Let $i \in N$ and I and I' be instances so that i's priority decreases from I to I'. If i is unmatched in* some *maximal weight matching of G_I^π, then i is unmatched in* all *maximal weight matchings of $G_{I'}^\pi$.*

Proof. Let $G_I^\pi = (N \cup C, E, w, (q_c))$ and $G_{I'}^\pi = (N \cup C, E', w', (q_c))$. It suffices to consider the case that i's priority for only one category decreases from I to I'. That is, there is $d \in C$ so that $\succsim_c = \succsim'_c$ for all $c \neq d$. Repeated application of this case yields the general statement.

Let μ and μ' be maximal weight matchings of G_I^π and $G_{I'}^\pi$, respectively. Assume for contradiction that $\mu(i) = \emptyset$ and $\mu'(i) \neq \emptyset$.

Claim 1. μ and μ' match the same number of agents ($|\mu| = |\mu'|$).

Proof. It follows from the definition of the reservation graph that either $E = E'$ (if $i \succ'_d \emptyset$) or $E = E' \cup \{\{i, d\}\}$ (if $\emptyset \succ_d i$). In particular, $E' \subset E$ so that μ' is also a matching of G_I^π. Since μ is a maximal weight matching of G_I^π, it follows from Lemma 1(ii) that μ is a maximal size matching of G_I^π. Hence, $|\mu| \geq |\mu'|$. Similarly, since $\mu(i) = \emptyset$, μ is also a matching of $G_{I'}^\pi$. Since μ' is a maximal weight matching of $G_{I'}^\pi$, we conclude $|\mu'| \geq |\mu|$. Together, $|\mu| = |\mu'|$. □

Figure 3 depicts the situation we reason about in the remainder in the proof.

Claim 2. There is an augmenting path for μ in $G^{\pi}_{I'}$ that alternates between edges in μ' and edges in μ with initial vertex i and an agent who is unmatched in μ' as terminal vertex. That is, there is a path $P = P_{\mu'} \cup P_{\mu}$ with

$$
\begin{aligned}
&P_{\mu'} = \{\{j_{s-1}, c_s\} \colon s \in \{1, \dots, n\}\} \subset \mu' - \mu \\
&P_{\mu} = \{\{j_s, c_s\} \colon s \in \{1, \dots, n\}\} \subset \mu - \mu' \\
&j_0 = i \text{ and } \mu'(j_n) = \emptyset \\
&\text{the } j_s \text{ are pairwise distinct} \\
&w'(P_{\mu'}) > w'(P_{\mu})
\end{aligned}
$$

Proof. Since i is matched in μ' and unmatched in μ ($\mu'(i) \neq \emptyset$ and $\mu(i) = \emptyset$), there is a path $P = P_{\mu'} \cup P_{\mu}$ in $G^{\pi}_{I'}$ that alternates between edges in μ' and μ and starts at i. ($P_{\mu'} = \{i, \mu'(i)\}$ and $P_{\mu} = \emptyset$ is one such path.) Assume that P is of maximal length among all such path starting at i. By Claim 1 and Lemma 1(ii), μ is a maximal size matching of $G^{\pi}_{I'}$. Hence, P has even length (otherwise alternating along P would yield a matching of larger size than μ, namely the matching $\mu - P_{\mu} \cup P_{\mu'}$). We conclude that P is induced by a sequence $i = j_0, c_1, j_1, \dots, c_n, j_n$ with $\{j_{s-1}, c_s\} \in \mu'$ and $\{j_s, c_s\} \in \mu$ for all $s \in \{1, \dots, n\}$ for some $j_1, \dots, j_n \in N$ and $c_1, \dots, c_n \in C$. Since P is of maximal length, it follows that $\mu'(j_n) = \emptyset$.

Now P_{μ} and $P_{\mu'}$ are matchings of $G^{\pi}_{I'}$. Since i is unmatched in P_{μ} but is matched in $P_{\mu'}$, P_{μ} and $P_{\mu'}$ do not match the same set of agents. By the proof of Lemma 1(iv), $w'(P_{\mu}) \neq w'(P_{\mu'})$. If $w'(P_{\mu}) > w'(P_{\mu'})$, then P is an augmenting path for μ' in $G^{\pi}_{I'}$, so that $\mu' - P_{\mu'} \cup P_{\mu}$ is a matching with larger weight than μ'. This contradicts that μ' is a maximal weight matching of $G^{\pi}_{I'}$. Hence, $w'(P_{\mu'}) > w'(P_{\mu})$. □

Claim 3. $c_s = d$ for some $s \in \{1, \dots, n\}$.

Proof. We have $w(\{j, c\}) = w'(\{j, c\})$ for all $j \in N$ and $c \neq d$. So if $c_s \neq d$ for all $s \in \{1, \dots, n\}$, then, by Claim 2, P is an augmenting path for μ in G^{π}_{I}. Since this would contradict that μ is a maximal weight matching of G^{π}_{I}, the claim follows. □

If $i \sim_d j$ for some agent $j \neq i$, then $w(e) = w'(e)$ for possibly all edges $e \in E'$ except for $e = \{i, d\}$. Hence, since i is unmatched in μ, $w(P_{\mu}) = w'(P_{\mu})$ and $w(P_{\mu'}) \geq w'(P_{\mu'})$. Thus, by Claim 2, $\mu - P_{\mu} \cup P_{\mu'}$ is a matching of G^{π}_{I} with larger weight than μ, which is a contradiction. Hence, we assume that the equivalence class of i in the priority ranking of d contains only i itself.

Let $S = \{s_1, \dots, s_{m-1}\}$ be the indices of the agents who are matched to category d in μ' labeled in increasing order ($\mu'(j_{s_l}) = d$, $s_{l-1} < s_l$). By Claim 3, S is non-empty. Let $s_0 = 0$ and $s_m = n$. For $l \in \{1, \dots, m\}$, denote by P^l the segment of P from $j_{s_{l-1}}$ to j_{s_l} (for example, if $s_1 = 1$, $P^1 = \{\{j_0, c_1\}, \{j_1, c_1\}\}$ and $\mu'(j_1) = d$). Lastly, let $P^l_{\mu} = P_{\mu} \cap P^l$ and $P^l_{\mu'} = P_{\mu'} \cap P^l$. Note that

$$|P^u_{\mu}| = |P^u_{\mu'}| \tag{3}$$

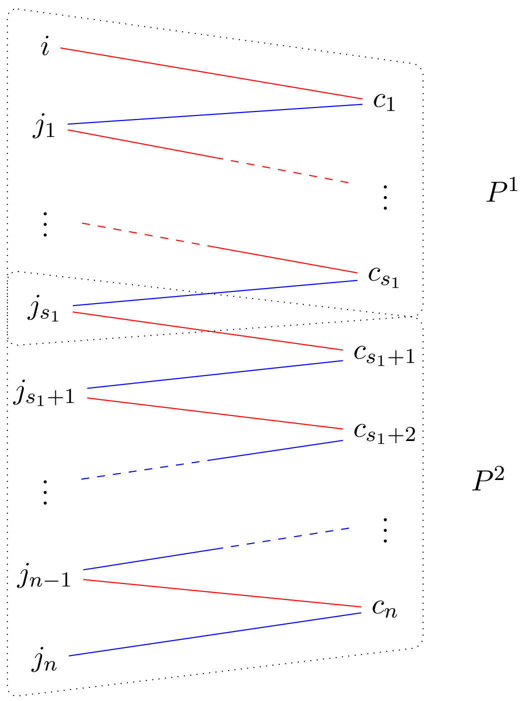

Figure 3: Augmenting path for μ in $G_{I'}$ alternating between edges in μ' (in red) and edges in μ (in blue). The dotted regions indicate the paths P^1 and P^2 in the proof of Claim 4 (when $m = 2$).

for all $l \in \{1, \ldots, m\}$ since each P^l has even length (the initial and terminal vertex of each P^l is an agent) and P^l alternates between edges in μ' and μ.

We will in the following freely and without mentioning use the fact that

$$\sum_{e \in \mu} w_\pi(e) < \epsilon = w'_c(\{j, d\}) - w_c(\{j, d\}) = w'(\{j, d\}) - w(\{j, d\})$$

if $i \succ_d j$ and $j \succsim'_d i$ (and similarly for μ').

Claim 4. $w'(P^l_\mu) \geq w'(P^l_{\mu'})$ for $l \in \{1, \ldots, m\}$.

Proof.

Case 1. Let $l = 1$. We show that alternating μ along the path P^1 cannot increase the weight of μ in $G^\pi_{I'}$. By construction, no edge on the path P^1 is adjacent to the category d. Hence, $w(e) = w'(e)$ for all $e \in P^1$ so that

$$w'(P^1_\mu) = w(P^1_\mu) \text{ and } w'(P^1_{\mu'}) = w_c(P^1_{\mu'})$$

Moreover, $\mu - P^1_\mu \cup P^1_{\mu'}$ is a matching of G^π_I since i is unmatched in μ. Since μ is a maximal weight matching of G^π_I, it follows that

$$0 \geq w(\mu - P^1_\mu \cup P^1_{\mu'}) - w(\mu) = w(P^1_{\mu'}) - w(P^1_\mu) = w'(P^1_{\mu'}) - w'(P^1_\mu)$$

Hence, $w'(P^1_\mu) \geq w'(P^1_{\mu'})$.

Case 2. Let $l \in \{2, \dots, m\}$ and assume that $j_{s_l} \succsim_d i$. Then $w(e) = w'(e)$ for all $e \in P^l$. The same argument as in Case 1 shows that $w'(P^l_\mu) \geq w'(P^l_{\mu'})$.

Case 3. Let $l \in \{2, \dots, m\}$ and assume that $i \succ_d j_{s_l}$. Assume for contradiction that $w'(P^l_{\mu'}) > w'(P^l_\mu)$. Observe that

$$w'(\{j_{s_l}, d\}) \leq w(\{j_{s_l}, d\}) + \epsilon \leq w(\{i, d\}) \tag{4}$$

where the first inequality is an equality if and only if $j_{s_l} \succsim'_d i$. Now $\bar{\mu} = \mu - P^l_\mu \cup P^l_{\mu'} - \{\{j_{s_l}, d\}\} \cup \{\{i, d\}\}$ is a matching of G^π_I since i is unmatched in μ (note that $c_{s_l+1} = d$). We have

$$\begin{aligned} w(P^l_{\mu'} - \{\{j_{s_l}, d\}\}) + w(\{i, d\}) &= w'(P^l_{\mu'} - \{\{j_{s_l}, d\}\}) + w(\{i, d\}) \\ &\geq w'(P^l_{\mu'} - \{\{j_{s_l}, d\}\}) + w'(\{j_{s_l}, d\}) = w'(P^l_{\mu'}) \end{aligned} \tag{5}$$

where the first equality uses that $\{j_{s_l}, d\}$ is the only edge in $P^l_{\mu'}$ adjacent to d and the inequality uses (4). Hence,

$$w(\bar{\mu}) - w(\mu) = w(P^l_{\mu'} - \{\{j_{s_l}, d\}\} \cup \{\{i, d\}\}) - w(P^l_\mu) \geq w'(P^l_{\mu'}) - w'(P^l_\mu) > 0$$

where we used (5) and $w(P^l_\mu) \leq w'(P^l_\mu)$ for the first inequality and the assumption we seek to contradict for the last. This contradicts that μ is a maximal weight matching of G^π_I. Hence, $w'(P^l_\mu) \geq w'(P^l_{\mu'})$.

□

Finally, we observe that Claim 4 gives

$$w'(P_\mu) = \sum_{l=1}^{m} w'(P^l_\mu) \geq \sum_{l=1}^{m} w'(P^l_{\mu'}) = w'(P_{\mu'})$$

But this contradicts Claim 2. Hence, it cannot be that $\mu(i) = \emptyset$ while $\mu'(i) \neq \emptyset$.

□

Lemma 3. *Let $I = (N, C, (\succsim_c), (q_c))$ and $I' = (N, C, (\succsim_c), (q'_c))$ be two instances so that $q_c \leq q'_c$ for all $c \in C$. If $i \in N$ is matched in some maximal weight matching of G^π_I, then i is matched in any maximal weight matching of $G^\pi_{I'}$.*

Proof. Let μ and μ' be maximal weight matchings of G^π_I and $G^\pi_{I'}$, respectively. Assume that $\mu(i) = c \neq \emptyset$ and $\mu'(i) = \emptyset$. We derive a contradiction by constructing an augmenting path for either μ in G^π_I or μ' in $G^\pi_{I'}$. We will use without further mention that G^π_I and $G^\pi_{I'}$ have the same set of vertices and edges and the same weight function. Moreover, since $q_c \leq q'_c$ for all c, every matching of G^π_I is also a matching of $G^\pi_{I'}$.

Let $\mathcal{P}$ be the set of all paths in G^π_I that alternate between edges in μ and μ' and contain the edge $\{i, \mu(i)\} \in \mu - \mu'$. Note that $\mathcal{P}$ is non-empty since $\{\{i, \mu(i)\}\} \in \mathcal{P}$. Let $P = P_\mu \cup P_{\mu'}$ be of maximal length among all paths in $\mathcal{P}$.

P_μ and $P_{\mu'}$ are matchings of $G^\pi_{I'}$ that do not match the same set of agents. Hence, by the proof of Lemma 1(iv), $w(P_\mu) \neq w(P_{\mu'})$. If $w(P_\mu) > w(P_{\mu'})$, then P is an augmenting path for

μ' in $G^\pi_{I'}$. Since i is unmatched in μ' and P has maximal length, $\mu' - P_{\mu'} \cup P_\mu$ is a matching of $G^\pi_{I'}$. The weight of this matching is larger than that of μ', contradicting that μ' has maximal weight. Similarly, if $w(P_\mu) < w(P_{\mu'})$, then $|P_{\mu'}| \geq |P_\mu|$, so that the initial vertex of P is i and the terminal vertex is an agent $j \in N$ that is unmatched in μ. Hence, $\mu - P_\mu \cup P_{\mu'}$ is a matching of G^π_I, which has larger weight than μ, a contradiction. □

We summarize the conclusions of Lemmas 1, 2, and 3 in the following theorem.

Theorem 1. *Any allocation rule that chooses a maximal weight matching of G^π_I for every instance I*

(i) complies with acceptability requirements,

(ii) is of maximum size among feasible matchings,

(iii) respects priorities,

(iv) is strategyproof, and

(v) is monotonic in quotas.

6 Computational Aspects

In this section, we discuss the computational complexity of finding maximal weight matchings. We show that a maximal weight matching of G^π_I can be computed in strongly polynomial-time. Hence, an allocation rule that chooses maximal weight matchings can be implemented efficiently. The polynomial-time guarantee follows from the fact that a maximum weight b-matching with upper capacities on the nodes can be computed in strongly polynomial time even if edge weights are rational (Schrijver, 2003, Theorem 32.5).

Reduction from b-matchings to matchings One can avoid having to compute a maximal weight b-matching and use a standard algorithm for finding maximal weight matchings. We can expand G^π_I by cloning each vertex $c \in C$ exactly q_c times. Cloning means replacing the vertex by several vertices with the same neighbors and weights on the adjacent edges. In the expanded graph, we have a vertex for each of the units reserved for a category rather than a single vertex for each category. We can compute a maximal weight matching of the expanded graph using the Hungarian Algorithm (see, e.g., Kuhn, 2010) for bipartite graphs, which runs in polynomial-time. It is straightforward to see that a maximal weight matching of the expanded graph corresponds in the obvious way to a maximal weight b-matching in the original reservation graph.

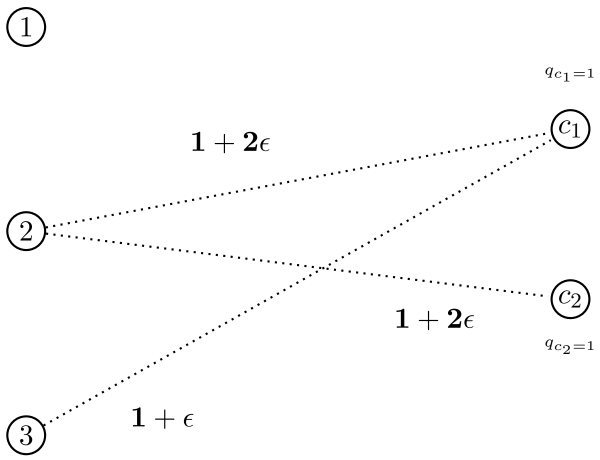

Figure 4: The modified reservation graph G_I for the instance I in Example 1. The dotted lines correspond to edges in E. The numbers close to the edges are their weights according to w.

Avoiding high-precision weights Although maximal weight matchings of the reservation graph can be computed in strongly polynomial-time, our approach requires using weights of high precision (the space requirement is still $|C|n \log n$ bits) because of the exponentially small weights of the weight function w_π. To avoid these high-precision weights, we provide an alternative polynomial-time algorithm that gives the same outcome. It iteratively computes maximum weight b-matchings of at most n graphs. These graphs are the same as the reservation graph G_I^π except that the weights w_π awarded based on the baseline ordering are omitted.

More precisely, for a given instance I, the corresponding *modified* reservation graph $G_I = (N \cup C, E, w, (q_c))$ has the same set of vertices and edges and the same quotas as the reservation graph G_I^π. We weight of an edge $\{i, c\} \in E$ is

$$w(\{i,c\}) = 1 + w_c(\{i,c\})$$

where w_c is defined as in Section 5. The modified reservation graph for the instance in Example 1 is depicted in Figure 4.

To state our algorithm, we also define a graph $H_I^{N'}$ that is parametrized by a subset of agents N'. It is identical to G_I except that for each $i \in N'$ and each c with $i \succ_c \emptyset$, $w_i(c) = 2n + 1$.

The algorithm is specified in pseudo-code Algorithm 1. The main idea of the algorithm is to work on the graph G_I and iteratively go through the agents in the baseline ordering and build the working set of selected agents N^*. The set of agents in N^* can all be matched in some maximum weight matching of G_I. For the new agent i under consideration, we check whether i can be matched in a maximum weight matching of G_I that also matches the agents currently in N^*. This is checked by modifying the weights of G_I to consider the graph $H_I^{N^* \cup \{i\}}$. The graph $H_I^{N^* \cup \{i\}}$ is the same as graph G_I except that a large weight of $2n$ is added to the edges adjacent to agents in $N^* \cup \{i\}$. This ensures that there is a maximum weight matching in G_I that matches the agents in $N^* \cup \{i\}$ if and only if the maximum weight matching of $H_I^{N^* \cup \{i\}}$

has weight $mw(H_I^{N^* \cup \{i\}}) = mw(G_I) + 2n(|N^*| + 1)$, where $mw(G)$ is the weight of a maximum weight matching in a graph G.

Algorithm 1

Input: $I = (N, C, (\succsim_c), (q_c))$; a baseline ordering π over agents in N

Output: A matching of G_I.

1: Construct the modified reservation graph G_I.
2: Compute the weight $mw(G_I)$ of a maximum weight matching of G_I.
3: Selected set of agents $N^* \leftarrow \emptyset$
4: **for** agent $i \notin N^*$ down the ordering π **do**
5: Consider graph $H_I^{N^* \cup \{i\}}$ that is identical to G_I except that for all $j \in N$ and $c \in C$,

$$w(\{j, c\}) = \begin{cases} 1 + 2n + w_c(\{j, c\}) & \text{if } j \in N^* \cup \{i\} \\ 1 + w_c(\{j, c\}) & \text{if } j \notin N^* \cup \{i\} \end{cases}$$

6: **if** $mw(H_I^{N^* \cup \{i\}}) = mw(G_I) + 2n(|N^*| + 1)$ **then**
7: Add i to N^*
8: **end if**
9: **end for**
10: Compute a maximum weight b-matching μ of $H_I^{N^*}$.
11: **return** μ

Example 3 (Illustration of Algorithm 1)**.** Suppose the baseline ordering is $\pi = 123$. In that case the algorithm first computes G_I as shown in Figure 2. The maximum weight matching of the graph is $\{\{3, c_1\}, \{2, c_2\}\}$. It is the unique maximum weight matching. Therefore, we are unable to add agent 1 to N^*. We then check 2 and 3 which are able to be added to N^*.

7 Discussion

We presented an allocation rule that applies to resource allocation problems in which the resources are reserved for categories, each of which has a priority ranking over agents. The rule has several properties that are desirable in applications. It is fair in the sense that it complies with the acceptability requirements and respects the priorities, it is efficient in that it yields maximum size matchings, and it never incentivizes agents to under-report which categories they are eligible for. Furthermore, weakly increasing quotas does not harm the agents who get a unit.

We can reframe the main theorem in the context of school choice (Abdulkadiroğlu and Sönmez, 2003) by viewing agents as students and categories as schools. The students are indifferent between all schools that are acceptable for them. The schools have priorities over the students. Then Theorem 1 reads as follows.

Theorem 2. *Consider the school choice problem where the students partition the schools into acceptable and unacceptable schools. Then, there is an allocation rule that only matches students to acceptable schools, has not justified envy, is non-wasteful, matches the maximal feasible number of students, and is strategyproof for students.*

Our allocation rule involves a baseline ordering π over the agents, which gives rise to a natural ordering in which patients are allocated units. We can go down the ordering π and give a unit to the agent that was matched by the allocation rule.

We assumed that the categories and their capacities are primitives of the model. A separate research problem is to decide on the distribution of the units over the categories with the aim to reduce the pandemic for the society. Finally, it will be useful to consider a more fine-grained model that allows quantifying how much a patient needs a unit.

Acknowledgements

This material is based on work supported by the Deutsche Forschungsgemeinschaft under grant BR 5969/1-1. The authors thank Adi Ganguly for helpful comments.

References

A. Abdulkadiroğlu and T. Sönmez. School choice: A mechanism design approach. *American Economic Review*, 93(3):729–747, 2003.

A. Abdulkadiroğlu and T. Sönmez. Ordinal efficiency and dominated sets of assignments. *Journal of Economic Theory*, 112(1):157–172, 2003.

D. Abraham, A. Blum, and T. Sandholm. Clearing algorithms for barter exchange markets: Enabling nationwide kidney exchanges. In *Proceedings of the 8th ACM Conference on Electronic Commerce (ACM-EC)*, pages 295–304. ACM Press, 2007.

M. O. Afacan, I. Bó, and B. Turhan. Assignment maximization, 2020.

S. Ahmadi, F. Ahmed, J. P. Dickerson, M. Fuge, and S. Khuller. An algorithm for multi-attribute diverse matching. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pages 3–9, 2020.

F. Ahmed, J. P. Dickerson, and M. Fuge. Diverse weighted bipartite b-matching. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 35–41. AAAI Press, 2017.

T. Andersson and L. Ehlers. Assigning refugees to landlords in Sweden: Stable maximum matchings. Technical Report 2016-08, Université de Montréal Papyrus Institutional Repository, 2016.

O. Aygün and I. Bó. College admission with multidimensional privileges: The Brazilian affirmative action case. *American Economic Journal: Microeconomics*, 2020.

O. Aygün and B. Turhan. Dynamic reserves in matching markets: Theory and applications. *Journal of Economic Theory*, 188, 2020.

H. Aziz. Mechanisms for house allocation with existing tenants under dichotomous preferences. *Journal of Mechanism and Institution Design*, 2018.

H. Aziz, S. Gaspers, and Z. Sun. Mechanism design for school choice with soft diversity constraints. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI)*, 2020.

P. Biró, T. Fleiner, R. W. Irving, and D. F. Manlove. The college admissions problem with lower and common quotas. *Theoretical Computer Science*, 411(34):3136 – 3153, 2010.

P. Biró, D. F. Manlove, and S. Mittal. Size versus stability in the marriage problem. *Theoretical Computer Science*, 411(16-18):1828–1841, 2010.

G. Bognar and I. Hirose. *The Ethics of Health Care Rationing: An Introduction*. Routledge, 2014.

A. Bogomolnaia and H. Moulin. Size versus fairness in the assignment problem. *Games and Economic Behavior*, 90:119–127, 2015.

L. Bruce and R. Tallman. Promoting racial equity in covid-19 resource allocation. *Journal of Medical Ethics*, 2021.

J. P. Dickerson, K. A. Sankararaman, A. Srinivasan, and P. Xu. Balancing relevance and diversity in online bipartite matching via submodularity. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1877–1884, 2019.

F. Echenique and M. B. Yenmez. How to control controlled school choice. *American Economic Review*, 105(8):2679–94, August 2015.

L. Ehlers, I. E. Hafalir, M. B. Yenmez, and M. A. Yildirim. School choice with controlled choice constraints: Hard bounds versus soft bounds. *Journal of Economic Theory*, 153:648—683, 2014.

E. J. Emanuel, G. Persad, R. Upshur, B. Thome, M. Parker, A. Glickman, C. Zhang, C. Boyle, M. Smith, and J. P. Phillips. Fair allocation of scarce medical resources in the time of covid-19. *New England Journal of Medicine*, 382(21):2049–2055, 2020.

S. Fink. The hardest questions doctors may face: Who will be saved? who won't?, 2020.

D. Gale and L. S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.

Y. A. Gonczarowski, N. Nisan, L. Kovalio, and A. Romm. Matching for the Israeli "Mechinot" gap year: Handling rich diversity requirements. In *Proceedings of the 20th ACM Conference on Economics and Computation*, pages 321–321, 2019.

M. Goto, A. Iwasaki, Y. Kawasaki, R. Kurata, Y. Yasuda, and M. Yokoo. Strategyproof matching with regional minimum and maximum quotas. *Artificial intelligence*, 235:40–57, 2016.

A. Grigoryan. Effective, fair and equitable pandemic rationing. Technical Report 3646539, SSRN, Nov 2020.

E. J. Heo. Equity and diversity in college admissions. In J. Laslier, H. Moulin, R. Sanver, and W. Zwicker, editors, *The Future of Economic Design*. Springer, 2019.

Y. Kamada and F. Kojima. Efficient matching under distributional constraints: Theory and applications. *The American Economic Review*, 105(1):67–99, 2015.

Y. Kamada and F. Kojima. Recent developments in matching with constraints. *The American Economic Review*, 107(5):200–204, 2017.

F. Kojima. New directions of study in matching with constraints. In J.-F. Laslier, H. Moulin, R. Sanver, and W. S. Zwicker, editors, *The Future of Economic Design*. Springer-Verlag, 2019. Forthcoming.

P. Krysta, D. F. Manlove, B. Rastegari, and J. Zhang. Size versus truthfulness in the house allocation problem. In *Proceedings of the 15th ACM Conference on Economics and Computation (ACM-EC)*, pages 453–470. ACM Press, 2014.

H. W. Kuhn. The Hungarian method for the assignment problem. In M. Jünger, T. M. Liebling, D. Naddef, G. L. Nemhauser, W. R. Pulleyblank, G. Reinelt, G. Rinaldi, and L. A. Wolsey, editors, *50 Years of Integer Programming 1958-2008 - From the Early Years to the State-of-the-Art*, pages 29–47. Springer, 2010.

R. Kurata, N. Hamada, A. Iwasaki, and M. Yokoo. Controlled school choice with soft bounds and overlapping types. *Journal of Artificial Intelligence Research*, 58:153–184, 2017.

E. Martin. Rationing in healthcare. Technical Report 8, Deeble Institute, 2015.

NASEM-National Academies of Sciences, Engineering, and Medicine. Framework for Equitable Allocation of COVID-19 Vaccine. Technical report, Washington, DC, 2020.

P. A. Pathak, T. Sönmez, M. U. Ünver, and M. B. Yenmez. Fair Allocation of Vaccines, Ventilators and Antiviral Treatments: Leaving No Ethical Value Behind in Health Care Rationing. Boston College Working Papers in Economics 1015, Boston College Department of Economics, July 2020. URL https://ideas.repec.org/p/boc/bocoec/1015.html.

G. Persad, M. E. Peek, and E. J. Emanuel. Prioritizing groups for access to covid-19 vaccines. *JAMA*, 324(16):1601–1602., 2020.

A. E. Roth. Deferred acceptance algorithms: history, theory, practice, and open questions. *International Journal of Game Theory*, 36:537—569, 2008.

A. E. Roth and M. A. O. Sotomayor. *Two-Sided Matching: A Study in Game Theoretic Modelling and Analysis*. Cambridge University Press, 1990.

A. Schrijver. *Combinatorial Optimization - Polyhedra and Efficiency*. Springer, 2003.

T. Sönmez and M. B. Yenmez. Affirmative action with overlapping reserves. Manuscript, 2020. URL http://fmwww.bc.edu/EC-P/wp990.pdf.

T. Sönmez, P. A. Pathak., M. U. Ünver, G. Persad, R. D. Truog, and D. B. White. Categorized priority systems: A new tool for fairly allocating scarce medical resources in the face of profound social inequities. *CHEST*, 2020.

R. D. Truog, C. Mitchell, and G. Q. Daley. The toughest triage — allocating ventilators in a pandemic. *New England Journal of Medicine*, 382(21):1973–1975, 2020.

WHO. A global framework to ensure equitable and fair allocation of COVID-19 products and potential implications for COVID-19 vaccines. Technical report, 2020.

8 Appendix

Recall that all units are identical and the agents are indifferent between receiving a unit through any category. Hence, for the agents, the relevant characteristic of a matching is the set of matched agents. We show that the maximal sets of agents (with respect to set inclusion) matched in a matching that complies with the acceptability requirements, is non-wasteful, and respects priorities, all have the same cardinality. By Lemma 1, there exists a maximal beneficiary assignment satisfying these three axioms. Hence, every agent that is matched in some matching satisfying the three axioms is matched in some matching that additionally is a maximal beneficiary assignment.

Lemma 4. *Let $\mathcal{M}$ be the set of all matchings for an instance I that comply with the acceptability requirements, are non-wasteful, and respect priorities. Then the maximal elements of $\{\mu^{-1}(C)\colon \mu \in \mathcal{M}\}$ (with respect to set inclusion) have the same cardinality.*

Proof. Let $\mathcal{I}$ be the collection of be maximal elements of $\{\mu^{-1}(C)\colon \mu \in \mathcal{M}\}$ and A and B in $\mathcal{I}$. Let μ and μ' be matchings in $\mathcal{M}$ so that μ matches the agents in A ($\mu^{-1}(C) = A$) and μ' matches the agents in B. It suffices to show that every maximal path (with respect to set inclusion) that alternates between edges in μ and μ' has even length.

Let $i \in A - B$. Among all paths that have i as initial vertex and alternate between edges in μ and μ', let P be one that is maximal.

Suppose P has *odd* length, so that it is defined by a sequence $i = j_0, c_1, j_1, c_2, \ldots, j_{n-1}, c_n$. (Note that $|(\mu')^{-1}(c_n)| < q_{c_n}$ by maximality of $|P|$.) Let $S = \{l \in \{1, \ldots, n\}$: there is $j \in N$ with $\mu'(j) = \emptyset$ and $j \succ_{c_l} j_{l-1}\}$.

If $S \neq \emptyset$, let $l_0 = \max S$ and $j \in N$ maximal with respect to $\succsim_{c_{l_0}}$ among all agents who are unmatched in μ'. Let μ'' be the matching that results from alternating μ' along the path $j, c_{l_0}, \ldots, j_{n-1}, c_n$. We claim that $\mu'' \in \mathcal{M}$. We see that μ''

(i) complies with the eligibility requirements since μ complies with the eligibility requirements and $j \succ_{c_{l_0}} j_{c_{l_0-1}} \succ_{c_{l_0}} \emptyset$.

(ii) respects priorities. For suppose $k \in N$ is unmatched in μ'' and $k \succ_c k'$ for some $k' \in N$ and $c \in C$. Then k is unmatched in μ' and $k \neq j$. Hence, since μ' respects priorities, we only need to consider $c = c_l$ and $k' = j_{l-1}$ for some $l \in \{l_0+1, \ldots, n\}$ or $c = c_{l_0}$ and $k' = j$. By the choice of l_0 and j, $\mu''(c_l) = j_{l-1} \succsim_{c_l} k$ for all $l \in \{l_0 + 1, \ldots, n\}$ and $j \succsim_{c_{l_0}} k$.

(iii) is non-wasteful since μ' is non-wasteful, μ'' matches every agent that is matched in μ', and μ'' matches to every category at least as many agents as μ'.

Hence, $\mu'' \in \mathcal{M}$. The fact that μ'' matches agents $B \cup \{j\}$ contradicts the maximality of B.

If $S = \emptyset$, alternating μ' along P yields a matching in $\mathcal{M}$ that matches $B \cup \{i\}$. The arguments are the same as in the previous case.

□