Efficient Computationally Tractable School Choice Mechanisms

Andrew McLennan* Shino Takayama† Yuki Tamura‡

February 12, 2025

Abstract

We propose applying the Generalized Constrained Probabilistic Serial (GCPS) mechanism of Balbuzanov (2022) to school choice problems with dichotomous school priorities. For priorities that are neither dichotomous nor strict, we develop a market clearing cutoffs (MCC) mechanism. These mechanisms avoid the inefficiencies resulting from student proposes deferred acceptance when the schools' priorities do not embody actual social values. They work especially well when each student has a safe school that will certainly admit her if she is not admitted to a school she prefers. They are strategy-proof in the large (Azevedo and Budish, 2019) and thus highly resistant to manipulation. Software implementing our algorithms has satisfactory running times even for very large problems.

Keywords: School choice, safe schools, object allocation, deferred acceptance, sd-efficiency, strategy-proof, strategy-proof in the large, probabilistic serial mechanism, Hall's marriage theorem.

1 Introduction

Many school systems around the world now use mechanisms that pass from the students' reported preferences to assignments of students to schools. The *Boston* (or *immediate acceptance*) mechanism[1] was one of the first mechanisms used in school choice.

*School of Economics, University of Queensland, a.mclennan@economics.uq.edu.au

†School of Economics, University of Queensland, s.takayama1@uq.edu.au

‡Department of Economics, Ecole Polytechnique, CREST, Institut Polytechnique de Paris, yuki.tamura@polytechnique.edu

[1]The Boston mechanism begins by assigning as many students to their favorite (according to the submitted rankings) schools as possible. It then assigns as many of the remaining students as possible to their second favorite schools, then as many of the students who still remain as possible to their third choice, and so forth.

Since it is possible that a student can (for example) greatly increase her chance of being accepted at her second favorite school if she ranks it as her favorite, the Boston mechanism is not strategy-proof, and the likelihood that other students are not reporting their preferences truthfully makes it strategically tricky, with high stakes.

In a seminal paper Abdulkadiroğlu and Sönmez (2003) propose the application, to school choice, of two strategy-proof mechanisms based on matching theory. The *student proposes deferred acceptance* (DA) mechanism[2] was originally proposed by Gale and Shapley (1962), and it has been widely adopted for school choice and similar problems around the world. The *top trading cycles* (TTC) mechanism[3] was originated by David Gale, as described by Shapley and Scarf (1974), and although it has some superior theoretical properties, it has found less practical acceptance. These mechanisms require that each school have a preference ordering over possible students, called a *priority*, that is strict. If these do not represent actual social preferences (perhaps because they are generated randomly, simply to fulfill the requirements of the mechanism) then DA can yield assignments that are inefficient. These losses can be quantitatively significant: in a study of New York City data Abdulkadiroğlu et al. (2009) found a Pareto improving reassignment that gave almost 4300 students a school they preferred.

This paper proposes alternative mechanisms. Each of these mechanisms generates probabilities of assigning each student to each school that are *feasible*: for each student, the sum of her assignment probabilities is one, and for each school the sum of its assignment probabilities does not exceed its capacity. Each mechanism then *implements* the assignment probabilities by generating a random deterministic assignment with a probability distribution that realizes each of the assignment probabilities.

A feasible matrix of assignment probabilities is *sd-efficient* if there is no other feasible matrix of assignment probabilities that gives each student a probability distribution over schools that first order stochastically dominates[4] the given probability distribution,

[2]At the outset in DA each student applies to her favorite school. Each school with more applicants than its capacity rejects the lowest priority applicants beyond the number it can serve. In each subsequent round each student who was rejected in the preceding round applies to her favorite school among those that have not rejected her, and each school retains the highest priority applicants, up to its capacity, among those who have applied in all rounds, and rejects all others. The process continues in the same manner until there is a round with no rejections.

[3]In TTC each student points to her favorite school and each school points to its highest priority student. The resulting directed graph has at least one cycle, each student in a cycle is assigned to the school she points to, and she is removed from the mechanism, along with the seat she claimed in her school. This process is then repeated with the remaining students and seats, and it continues in this manner until all students have been assigned. When different schools have different priorities, the role of priorities in TTC is hard to grasp. (See Leshno and Lo (2020).)

[4]For a strict preference on a set of objects O, one probability distribution on O *first order stochastically dominates* a second probability distribution if, for each $o \in O$, the probability of an object that is at least as desirable as o is at least as large for the first distribution as for the second. If this is the case and the two distributions are not the same, then some such inequality holds strictly, and we say that the

with strict domination for some students. Any probability distribution over deterministic assignments that implements an sd-efficient matrix of assignment probabilities assigns all probability to assignments that are ex post efficient. Whereas DA randomizes at the stage of priority assignment, we avoid inefficiencies by deferring randomization until after sd-efficient assignment probabilities have been determined.

The schools have *dichotomous priorities* if, for each school, each student is either *eligible* to attend that school or she is not, and the schools give equal consideration to all eligible students. Eligibility may be based on characteristics such as test scores and residential location, and the student may also be ineligible if she does not rank the school highly enough in her reported preference. For dichotomous priorities we propose the *generalized constrained probabilistic serial* (GCPS) mechanism of Balbuzanov (2022), which is a generalization of the probabilistic serial (PS) mechanism of Bogomolnaia and Moulin (2001) (henceforth BM). Balbuzanov shows that the allocations produced by the GCPS mechanism are sd-efficient. Our main contribution with respect to the GCPS mechanism is to show that, in its application to school choice, it is computationally tractable.

The schools' priorities are not dichotomous if the schools have nontrivial preferences over eligible students. We assume that each school has a finite set of priority classes, which are ranked. A common example is that highest priority is given to students with a sibling at the school who live in the school's walk zone, second priority is given to students outside the walk zone with a sibling at the school, third priority is given to students in the walk zone without a sibling at the school, and other eligible students have lowest priority. To apply DA in such a setting it is necessary to assign strict priorities that refine the given coarse priorities, and again this can result in inefficiencies.

The *market clearing cutoffs* (MCC) mechanism is a variant of a mechanism proposed by Azevedo and Leshno (2016). The MCC mechanism computes a system of cutoff priority classes and numerical cutoffs that measure the extent to which students in the cutoff priority classes are rationed. The allowed consumption of a student at a school is zero if her priority at the school is below the school's cutoff, it is one if her priority at the school is above the school's cutoff, and when her priority at the school is the school's cutoff, it is determined by the school's numerical cutoff. The computed cutoff priority classes and numerical cutoffs are such that for each school, the total demand (computed in the natural way) is not greater than the school's capacity, and the school does not restrict admission when its total demand is less than its capacity. The MCC mechanism generalizes DA because the two mechanisms coincide when the schools'

domination is *strict*.

priorities are strict.

The assignment probabilities produced by the MCC mechanism need not be sd-efficient relative to the schools' priorities. There are simple algorithms that pass from a matrix of assignment probabilities that is market clearing, relative to a system of cutoffs, to a Pareto improving matrix of assignment probabilities that is sd-efficient within the set of matrices of assignment probabilities that are market clearing relative to the cutoffs. An *enhanced MCC mechanism* is a mechanism that first computes the MCC cutoffs and assignment probabilities, and then applies such an algorithm.

Strictly speaking, the GCPS and MCC mechanisms are not strategy-proof for the students, but they are still quite resistant to manipulation. Intuitively, opportunities to manipulate are uncommon, the potential gains from manipulation are small, and attempting to manipulate is risky, since whenever the manipulation changes the outcome, the student is to some extent receiving what she asks for rather than what she most prefers. (Enhanced MCC mechanisms have an additional, minor, possibility for manipulation.) Formally, we will show that the GCPS and MCC mechanisms are *strategy-proof in the large* (Azevedo and Budish, 2019) which means that any expected gains from misreporting, for students without precise information about the others' types, vanish asymptotically as the number of students per school increases. The underlying intuition is that any benefit of a manipulation must be derived from its impact on price-like variables whose probability distribution, from the point of view of the student, is only slightly affected by the student's own report. Azevedo and Budish discuss numerous examples of mechanisms that are not strategy-proof, but are strategy-proof in the large, and which work well in practice, as well as examples of mechanisms that are not strategy-proof in the large for which there is empirical evidence that real world participants strategically misreport.

In practice (e.g. Pathak (2017)) transparency and straightforward incentives are required in order for school choice mechanisms to be accepted by parents and school administrators. DA is quite difficult for lay people to understand,[5] and the Boston mechanism continued to be widely used for a long time, in spite of its theoretical drawbacks, because it is more intuitive than DA. The main thrust of the GCPS and MCC mechanisms is very straightforward and intuitive: each student is given what she says she wants, to the extent possible. The computation of "the extent possible" is admittedly far from simple, but students and parents should find it easy to understand that manipulation has high risks and meager potential gains.

[5] Ashlagi and Gonczarowski (2018) show that DA cannot be formulated in a way that makes it obviously strategy-proof (Li, 2017) for the students. Bó and Hakimov (2019) give references to an extensive body of experimental evidence and field data showing that even though truthful revelation is a dominant strategy in DA, misreporting of preferences is common.

In practice almost all school choice mechanisms limit the number of schools that a student can rank, and in large districts such restrictions seem unavoidable. We focus on versions of the GCPS, MCC, and enhanced MCC mechanisms that finish in a single round (other possibilities are described below) and do not assign any student to a school she did not rank. Specifically, we assume that each student is assigned a *safe school* which is guaranteed to accept the student if she is not admitted to a school she prefers, and which she may be required to attend if other schools do not admit her. Some systems (e.g., the state of Victoria in Australia) have *neighborhood priority* in which each student's safe school is the one whose district contains her residence, and of course there are many other possibilities.

In the case of the GCPS mechanism we assume that for each school, the number of students for whom that school is the safe school is not greater than the school's capacity. In the case of the MCC we assume that for each student, the number of students that have equal or higher priority at the safe school is not greater than the school's capacity. (When priorities are dichotomous, this requirement is too stringent, which is why the MCC mechanism and enhanced MCC mechanisms are not applicable in that case.) In order for the GCPS and MCC mechanisms to compute feasible assignments, such assignments must exist, and our assumptions insure that assigning each student to her safe school is a feasible assignment.

Mechanisms that would be strategy-proof without restrictions on the number of schools that can be ranked become manipulable when such restrictions are imposed. Haeringer and Klijn (2009) study the Nash equilibria of matching based mechanisms with such limitations. Calsamiglia et al. (2010) is an experimental study of the effects of constraining the number of schools that can be ranked, for DA and TTC; a main finding is that constraints have a large negative effect on manipulability, and reduce efficiency and stability while increasing segregation.

When there are safe schools each student submits only a ranking of those schools she weakly prefers to her safe school. If, for each student, the number of such schools is not greater than the number of schools she is allowed to rank, then DA (for which safe schools are also possible) becomes strategy-proof, and GCPS and MCC become strategy-proof in the large, because in effect the student is allowed to rank all schools. Of course having safe schools that students are likely to find desirable is consistent with the main goal of school choice, which is to assign students to schools they would like to attend. It seems likely that the guaranteed lower bounds on outcomes provided by safe schools will be intuitively appealing to parents.

In the New York City High School Match as of 2006 (Pathak, 2006) each student submitted a ranking of up to 12 schools. Of the roughly 100,000 participants, over 8,000 were unmatched after the main round, in the sense that they were not offered a seat by any school they ranked. These students submitted new rank ordered lists for the supplementary round, in which schools with unfilled capacity participated. Students who did not receive a seat in the supplementary round were assigned administratively. We do not know the particular considerations that motivated this design. (Perhaps neighborhood priority would have impeded a goal of school desegregation since there was a high degree of de facto residential segregation.)

The important point for us is that GCPS and MCC can also be employed in multiround systems: each student's safe school in the first round is participation in the second round, for each student in the second round the safe school is participation in the third round or administrative assignment, and so forth. In the remainder we assume a single round because this setting is simple, but rich enough to encompass most relevant technical issues. Issues related to multiround systems are not analyzed here.

1.1 Related Literature

The literature on school choice is now vast; Abdulkadiroğlu and Andersson (2022) is a recent survey. In this section we survey some of the literature that is most closely related to our work.

In response to the inefficiencies observed by Abdulkadiroğlu et al. (2009), a rather extensive literature (Erdil and Ergin, 2008, Kesten, 2010, Tang and Yu, 2014, Kesten and Ünver, 2015, Che and Tercieux, 2019, Dur et al., 2019, Ehlers and Morrill, 2020, Troyan et al., 2020, Tang and Zhang, 2021, Reny, 2022, Cerrone et al., 2024) studies how the outcome of DA might be adjusted ex post. In fact there are theoretical barriers to improving efficiency by manipulating the breaking of ties in the schools' rankings. Gale and Shapley (1962) show that DA yields the best outcome for each student that can be achieved in any allocation without justified envy for the given priorities. Improving on results of Kesten (2006) and Erdil and Ergin (2008), Theorem 1 of Abdulkadiroğlu et al. (2009) asserts that, for any member of a large class of tie breaking rules, there is no mechanism that is both strategy-proof for that tie breaking rule and gives outcomes that weakly Pareto dominate those produced by DA.

A different approach to improving efficiency is to use mechanisms which (unlike DA and TTC, precisely because they are strategy proof) are responsive to cardinal preferences. Miralles (2009) and Abdulkadiroğlu et al. (2011) study the Boston mechanism from this point of view, and He et al. (2018) propose a version of the Hylland and Zeckhauser (1979) pseudo-market concept for school choice problems when the schools have coarse priorities, as in Section 6.

Although it is not widely used, TTC continues to be a topic of research. Some papers (Morrill, 2015, Hakimov and Kesten, 2018, Grigoryan, 2023) proposed modified versions of the mechanism. Leshno and Lo (2020) analyze it in terms of cutoffs for the schools.

The GCPS, MCC, and enhanced MCC mechanisms may be applied to domains other than school choice. For example, motivated by matching of medical residents with hospitals in Japan and similar problems, Kamada and Kojima (2015, 2017) study mechanisms in which regional caps on the number of residencies are implemented by imposing caps on the number of residencies at individual hospitals in the region. This can lead to a hospital rejecting applicants as a result of the hospital's cap even though other hospitals in the region have unfilled vacancies. They propose a more flexible version of DA in which some hospitals are allowed to exceed their caps if the total number of doctors matched to the region is below the region's cap. Similar effects can be achieved by running our mechanisms repeatedly while adjusting the caps of individual hospitals.

One way to implement affirmative action objectives has been suggested by Abdulkadiroğlu and Sönmez (2003). For example, a school may be divided into three subschools, one with 30% of the seats that is reserved for minority students, one with 30% of the seats that is reserved for majority students, and one with 40% of the seats that accepts all students. "Hard" upper and lower bounds for the percentages of students of different types are extensively used in practice, but Kojima (2012) and Hafalir et al. (2013) point out that they lead to conflicts with other objectives, and Ehlers et al. (2014) suggest implementing affirmative action goals using soft bounds. Such an approach can be implemented, at least informally, by running our mechanisms multiple times while adjusting the parameters to better reconcile competing objectives.

We now describe the PS mechanism of BM and subsequent generalizations. BM study the problem of assigning a different object from a finite set to each of finitely many agents, based on their reported strict ordinal preferences. BM provide an intuitive description of the PS mechanism in which each object is regarded as a perfectly divisible cake of unit size. At each moment in the unit interval of time each agent consumes, at unit speed, probability of her favorite cake, among those that have not yet been fully consumed. Provided that there are at least as many objects as agents, at time 1 each agent has a probability distribution over the objects, and for each object the sum of the assignment probabilities is not greater than one. Among the most important theoretical results are that the PS mechanism is sd-efficient but not strategy-proof.

Extensions of BM's cake eating procedure have been proposed in (at least) six other papers. Using the method of network flows (which is also used in the proof of Theorem 1 in Appendix A) Katta and Sethuraman (2006) extend the PS mechanism to profiles of preferences with indifferences. Their mechanism has both the PS mechanism for strict preferences and the mechanism proposed by Bogomolnaia and Moulin (2004) for matching problems with dichotomous preferences as special cases. Bogomolnaia (2015) provides a welfarist characterization of it.

Kojima (2009) studies perhaps the simplest extension of BM in which agents receive multiple objects. Each agent receives $r \geq 2$ objects, and the number of objects is r times the number of agents. The mechanism is shown to be sd-efficient and envy-free, but not weakly strategy-proof, as this concept is defined by BM.

Yılmaz (2010) studies house allocation problems with existing tenants, which are object allocation problems in which some objects have owners who can insist on not receiving a worse object. He proposes the special case of the mechanism studied here for that problem, and in particular he recognizes the relationship between Hall's marriage theorem, its generalization by Gale (1957), and the set of feasible allocations. His algorithm is generalized by Athanassoglou and Sethuraman (2011) to problems in which agents have fractional endowments. Yılmaz (2009) uses the methods of Katta and Sethuraman (2006) to extend the mechanism to the domain of preferences with indifferences.

Budish et al. (2013) (BCKM) study problems in which there are constraints that require that certain sums of probabilities are bounded, either below, in which case the constraint is a *floor constraint*, or above, in which case it is a *ceiling constraint*. For a problem with only ceiling constraints in which there is a "null object" (e.g., being unemployed, unhoused, or unschooled) that is available in infinite supply, and which is not involved in any constraint, they propose a *generalized probabilistic serial* (GPS) mechanism. As in BM, at each moment in $[0, 1]$ each agent increases her probability of her favorite available object. When a ceiling constraint binds with equality, the sets of available objects are revised by disallowing further consumption of probabilities that would violate a constraint. Since the null object is always available, each agent's set of available objects is always nonempty. Thus at time 1 each agent has total probability one, and the GPS assignment is defined as the probability shares that have been consumed by each agent at time 1. BCKM also developed an algorithm (described for our special case in Online Appendix C) for *implementation*, which passes from a matrix of assignment probabilities to a random deterministic assignment whose distribution realizes the given probabilities.

Balbuzanov (2022) generalizes the BCKM mechanism by allowing the set of feasible allocations to be an arbitrary polytope in the nonnegative orthant of the space of matrices of assignment probabilities. (Echenique et al. (2021) follow this approach in their study of pseudo-market equilibria with constraints.) We specialize this mechanism to a setting that is slightly more general than school choice.

1.2 Structure of the Paper

We briefly describe the structure of the remainder. Sections 2-5 describe the GCPS mechanism and our algorithm for computing it. Section 2 gives an overview, and definitions of key technical concepts. Section 3 states our generalization of Hall's marriage theorem. During the allocation process there can be a *critical pair* consisting of a set J of agents and a set P of objects such that the agents in J must be assigned all of the remaining capacity of the objects in P. Section 4 studies such pairs. Section 5 describes the algorithm for computing the GCPS allocation.

Section 6 defines the MCC mechanism, and describes a method of computing it. Section 7 explains that GCPS allocations are sd-efficient (this result is due to Balbuzanov) and also efficient in relation to other orderings of the set of probability measures on objects derived from an ordinal preference that correspond to the limits of extreme risk loving and extreme risk averse cardinal preferences. As mentioned earlier, the MCC mechanism does not produce efficient assignment probabilities, and in Section 7 we explain how to pass to assignment probabilities that are efficient relative to the given priorities, thereby defining enhanced MCC mechanisms.

Section 8 presents examples illustrating why the GCPS and MCC mechanisms are not strategy-proof, and then defines strategy-proofness in the large, and proves that the GCPS and MCC mechanisms satisfy this condition. Section 9 considers the fairness properties of GCPS allocations. Section 10 provides some concluding remarks.

Appendix A contains some proofs (of Theorem 1 and Propositions 2 and 4) not presented in the body of the paper, and Appendix B proves a result from Section 8. Online Appendix C describes a special case of an algorithm of BCKM that implements a matrix of assignment probabilities by passing to a random deterministic assignment whose distribution realizes the given probabilities. Online Appendix D gives a brief informal description of the software package *GCPS MCC Schools*, which implements the algorithms described in Sections 5 and 6 and Online Appendix C. The software has satisfactory running times for the largest contemporary school choice problems, and is ready for practical application.

2 Overview of the GCPS Mechanism

We begin the technical exposition with a high level overview of the GCPS mechanism. We first review some basic definitions.

A *polytope* Q may be defined to be the convex hull of a finite set of points, or as an intersection of finitely many closed half spaces that happens to be bounded. To avoid technical detail our discussion in this paragraph assumes that Q is full dimensional, in the sense that its affine hull is the entire Euclidean space of which it is a subset. Among the finite systems of weak linear inequalities that may be used to define Q, there is a unique (up to rescaling of inequalities by multiplication by positive scalars) such system that is minimal, and that is contained in any other such system. Its elements are the *facet inequalities* of Q. For each facet inequality the corresponding *facet* is the subset of Q on which the facet inequality holds with equality. A subset of Q is a *face* if it is Q itself, the null set, or the intersection of some set of facets. A polytope Q is the convex hull of a finite set of points, and among the finite sets whose convex hulls are Q, there is a unique such set that is minimal in the sense that it is contained in any other such set, whose elements are the *vertices* of Q. The vertices of Q may also be described as its extreme points, where an *extreme point* of Q is a point that cannot be expressed as a convex combination of other points of Q.

In Balbuzanov (2022) the set of feasible allocations is a given polytope Q in the nonnegative orthant of the space of matrices of assignment probabilities. Let R be the intersection of the nonnegative orthant with the sum of Q and the nonpositive orthant. That is, a point in the nonnegative orthant is in R if and only if it lies below some point of Q. The GCPS allocation process is a piecewise linear function $p\colon [0,1] \to R$. It begins with $p(0)$ equal to the origin and increases each student's probability of receiving her favorite object, among those she is allowed to consume, until one of the facet inequalities of R is encountered.

A key result (Balbuzanov's Proposition 1) is that the facet inequalities of R (other than the nonnegativity conditions) require that weighted sums of probabilities, with nonnegative weights, not exceed certain quantities. When the process encounters one or more facet inequalities, each student's set of allowed objects is updated by disallowing further consumption of probabilities that would result in one of these facet inequalities being violated. The process then continues, with each student increasing the probability of receiving her favorite allowed object until additional facet inequalities of R are encountered, and again the students' sets of allowed objects are updated. (For the problems we study each student's set of allowed objects is always nonempty.) Eventually the process arrives at a point $p(1) \in Q$ that is, by definition, the GCPS allocation.

A computational implementation of the GCPS mechanism must have a way of detecting when the allocation process encounters a facet of R. One possible implementation first passes to the description of Q as a convex hull of vertices. The vertices of R are all the points obtained from vertices of Q by changing some of the components to zero, and one may then pass from this set of vertices to the description of R as an intersection of finitely many half spaces. The computational problem of passing from the description of a polytope as a convex hull of vertices to its description as an intersection of half spaces, and the reverse computation, are well studied, and efficient softwares for these tasks are available. (See Section 3 of Balbuzanov (2022).) However, even if the number of bounding inequalities of Q and the number of bounding inequalities of R are small, large data structures can arise at intermediate stages of the computation. For example, for the problem of assigning n objects to n agents the numbers of facet inequalities of Q and R are constant multiples of n, but Q has $n!$ vertices.

Our first main result is a generalization of Hall's marriage theorem. For a class of problems somewhat more general than school choice problems it gives a set of inequalities, in closed form, that constitute a necessary and sufficient condition for the nonemptiness of Q. A direct consequence is a result giving a set of inequalities that contains the facet inequalities of R. The number of such inequalities is $2^{|O|}$, where O is the set of schools and $|O|$ is its cardinality. An algorithm that monitors these inequalities has reasonable running times when $|O| \leq 25$.

A second main innovation is another algorithm (described in Section 5) for computing the GCPS allocation. In addition to computing p, it computes a piecewise linear path $\overline{p}\colon [0,1] \to Q$ such that $p(t) \leq \overline{p}(t)$ for all t, by iteratively computing the linear segments of the combined path $(p, \overline{p})\colon [0,1] \to R \times Q$. Each linear segment continues in the same direction until the time at which continuing further would result in the violation of one of the constraints. At such a time there is a polynomial time procedure that either finds a new direction for $\overline{p}$ that allows continuation of $(p, \overline{p})$ with the given direction of p, or finds a set of agents J and a set of objects P such that the only feasible allocations give the agents in J all of the remaining resources in P and their maximum allowed consumptions of objects in the complement of P. The algorithm is recursive in the sense that its continuation combines the application of the algorithm to the continuation for J and P with the application of the algorithm to the continuation for the complements of these sets. The algorithm has been implemented (see Online Appendix D) and computational experience shows that it is capable to handling very large school choice problems.

3 A Generalized Hall's Marriage Theorem

In this section we introduce the formal framework, state the generalization of Hall's theorem, and provide useful characterizations of Q and R.

A *communal endowment economy* (CEE) is a quintuple $E = (I, O, r, q, g)$ in which I is a nonempty finite set of *agents*, O is a nonempty finite set of *objects*, $r \in \mathbb{R}^I_+$, $q \in \mathbb{R}^O_+$, and $g \in \mathbb{R}^{I \times O}_+$. We say that r_i is i's *requirement*, q_o is the *quota* of o, and g_{io} is *i's o-max*. We say that E is *integral* if $r \in \mathbb{Z}^I_+$, $q \in \mathbb{Z}^O_+$, and $g \in \mathbb{Z}^{I \times O}_+$. In comparison with most models of random assignment, the matrix g is the main novelty, and we will see that it may represent several things and be used in various ways.

An *allocation* for I and O is a matrix $p \in \mathbb{R}_+^{I \times O}$. Such a p is *integral* if $p \in \mathbb{Z}_+^{I \times O}$. A *partial allocation* for E is an allocation p such that $\sum_o p_{io} \le r_i$ for all i, $\sum_i p_{io} \le q_o$ for all o, and $p_{io} \le g_{io}$ for all i and o. A *feasible allocation* is a partial allocation m such that $\sum_o m_{io} = r_i$ for all i. A partial allocation p is *possible* if there is a feasible allocation m such that $p \le m$. Let Q be the set of feasible allocations, and let R be the set of possible partial allocations.

For $J \subset I$ and $P \subset O$ let $J^c = I \setminus J$ and $P^c = O \setminus P$ be the complements. We say that E satisfies the *generalized marriage condition* (GMC) if, for every $J \subset I$ and $P \subset O$,

$$\sum_{i \in J} r_i \le \sum_{i \in J} \sum_{o \in P^c} g_{io} + \sum_{o \in P} q_o.$$

We will refer to this relation as the *GMC inequality* for (J, P). Note that the GMC inequality for $(\{i\}, \emptyset)$ is $r_i \le \sum_o g_{io}$, and the GMC inequality for (I, O) is $\sum_i r_i \le \sum_o q_o$. The GMC is obviously necessary for the existence of a feasible allocation. Our first main result is:

Theorem 1. The CEE E has a feasible allocation if and only if it satisfies the GMC.

Our proof of Theorem 1 (in Appendix A) is an application of the max-flow min-cut theorem of Ford and Fulkerson (1956). Hall's marriage theorem, the Gale supply-demand theorem, and the max-flow min-cut theorem are three members of a large and important class of results in combinatorial matching theory that are equivalent in the informal sense that relatively simple arguments (described in detail by Reichmeider (1978, 1985)) allow one to pass from any member of the class to any other. As yet another member of this class, Theorem 1 does not provide distinctly novel mathematical information. Its primary significance here, and perhaps more generally, is that the test it provides is in closed form.

Several types of CEE occur in our discussion. A *Hall marriage problem* is a CEE such that for all i and o, $r_i = 1$, $q_o = 1$, and $g_{io} \in \{0, 1\}$. In this case elements of I are *boys* and elements of O are *girls*. Intuitively a Hall marriage problem is a bipartite graph with an edge connecting boy i to girl o if i and o are compatible. The set of *neighbors* of boy i is $N_g(i) = \{\, o \in O : g_{io} = 1 \,\}$, and for $J \subset I$ we set $N_g(J) = \bigcup_{i \in J} N_g(i)$. We say that E satisfies the *marriage condition* if $|J| \le |N_g(J)|$ for all $J \subset I$. The GMC inequality for J and $P = N_g(J)$ gives this inequality. Conversely, for a given $J \subset I$, the contribution of $o \in N_g(J)$ to the right hand side of the GMC inequality is minimized if $o \in P$, and the contribution of $o \in N_g(J)^c$ is minimized if $o \in P^c$, so if the GMC is satisfied for J and $N_g(J)$, then it is satisfied for J and any P. Therefore $|J| \leq |N_g(J)|$ for all J implies that the GMC is satisfied, so Theorem 1 implies that E has a feasible allocation if and only if the marriage condition is satisfied.

For a Hall marriage problem an integral feasible allocation is called a *matching*. (Each of the boys has a different partner.) Hall's marriage theorem asserts that a Hall marriage problem has a matching if and only if it satisfies the marriage condition. To pass from a feasible allocation to a matching one can repeatedly adjust the allocation along paths of fractional allocations that alternate between boys and girls, and either form a loop or pass from one incompletely allocated girl to another. A more precise and general version of this argument is given in Online Appendix C.

A *Gale supply-demand CEE* is a CEE E such that $g_{io} \in \{0, r_i\}$ for all $i \in I$ and $o \in O$. The Gale (1957) supply-demand theorem[6] is the special case of Theorem 1 for a Gale supply-demand CEE.

We say that E is a *school choice CEE* if $r_i = 1$ and $g_{io} \in \{0, 1\}$ for all i and o, and we write $E = (I, O, 1, q, g)$ to indicate that this is the case. In a school choice CEE elements of I are *students*, elements of O are *schools*, and each student must receive a seat in some school. In an integral school choice CEE each school has an integral number of seats, and for each student i and school o, $g_{io} = 1$ if i is eligible to attend o, and otherwise $g_{io} = 0$.

For $i \in I$ let

$$\alpha_i = \{\, o \in O : g_{io} > 0 \,\}$$

be the set of objects that are *possible* for i, and for $P \subset O$ let $J_P = \{\, i \in I : \alpha_i \subset P \,\}$ be the set of agents who cannot be allocated objects outside of P. If E is an integral school choice CEE, then for any $P \subset O$, J_P minimizes the difference between the right hand side and the left hand side of the GMC inequality. Therefore E satisfies the GMC if and only if, for each $P \subset O$, $|J_P| \leq \sum_{o \in P} q_o$.

The next result gives a finite collection of inequalities, in closed form, that contains the facet inequalities of R. If p is a partial allocation, let

$$E - p = (I, O, r', q', g')$$

be the derived CEE in which $r'_i = r_i - \sum_o p_{io}$, $q'_o = q_o - \sum_i p_{io}$, and $g'_{io} = g_{io} - p_{io}$. If p is a partial allocation, m is an allocation, and $p \leq m$, then m is a feasible allocation for E if and only if $m - p$ is a feasible allocation for $E - p$. Thus a partial allocation p

[6] Although this result is attributed to Gale (1957) by Yılmaz (2010), and perhaps others, this exact formulation does not appear in Gale's paper. The paper does consider slightly more complicated problems, and it is easy to see that this result can be obtained from Gale's methods in the same manner.

is possible if and only if $E - p$ has a feasible allocation, which of course is the case if and only if $E - p$ satisfies the GMC. Substituting the definitions above into the GMC inequality for $E - p$ and (J, P), then simplifying, gives

$$\sum_{i \in J^c} \sum_{o \in P} p_{io} \leq \sum_{o \in P} q_o + \sum_{i \in J} \sum_{o \in P^c} g_{io} - \sum_{i \in J} r_i. \tag{1}$$

Proposition 1. *R is the set of partial allocations p such that (1) holds for all $J \subset I$ and $P \subset O$.*

4 Critical Pairs

In this section we work with a given CEE E that satisfies the GMC. For $J \subset I$ and $P \subset O$ we say that the pair (J, P) is *critical* for E if $(J, P) \neq (\emptyset, \emptyset)$ and it satisfies the GMC inequality for (J, P) with equality:

$$\sum_{i \in J} r_i = \sum_{i \in J} \sum_{o \in P^c} g_{io} + \sum_{o \in P} q_o.$$

We refer to this condition as the *GMC equality* for (J, P). Our goal in this section is to understand the relationship between critical pairs and feasible allocations, and how the various critical pairs for E are related to each other.

We say that E is *critical* if (I, O) itself is a critical pair, which is the case if and only if $\sum_i r_i = \sum_o q_o$, so that any feasible allocation consumes all of the available resources. We say that E is *simple* if there are no critical pairs (J, P) with $(J, P) \neq (I, O)$.

Evidently, if (J, P) is critical for E, then any feasible allocation m gives the agents in J all of the endowment of objects in P and also as much of the objects in P^c as g allows. Conversely, if m is a feasible allocation such that $\sum_{i \in J} m_{io} = q_o$ for all $o \in P$ and $m_{io} = g_{io}$ for all $i \in J$ and $o \in P^c$, then

$$\sum_{i \in J} r_i = \sum_{i \in J} \sum_{o} m_{io} = \sum_{i \in J} \sum_{o \in P^c} m_{io} + \sum_{o \in P} \sum_{i \in J} m_{io} = \sum_{i \in J} \sum_{o \in P^c} g_{io} + \sum_{o \in P} q_o.$$

Lemma 1. For $J \subset I$ and $P \subset O$ the following are equivalent:

(a) (J, P) is critical for E;

(b) There is a feasible allocation m such that $\sum_{i \in J} m_{io} = q_o$ for all $o \in P$ and $m_{io} = g_{io}$ for all $i \in J$ and $o \in P^c$;

(c) For every feasible allocation m, $\sum_{i \in J} m_{io} = q_o$ for all $o \in P$ and $m_{io} = g_{io}$ for all $i \in J$ and $o \in P^c$.

The next result gives a key property of critical pairs. Its proof (in Appendix A) applies the last result.

Proposition 2. The set of critical pairs for E is a lattice in the sense that if (J,P) and (J',P') are critical pairs, then so are $(J \cup J', P \cup P')$ and $(J \cap J', P \cap P')$.

Now suppose that (J,P) is critical for E. Let

$$E_{(J,P)} = (J, O, r|_J, q', g|_{J \times O}) \quad \text{and} \quad E^{(J,P)} = (J^c, P^c, r|_{J^c}, q'', g|_{J^c \times P^c})$$

where $q'_o = q_o$ if $o \in P$, $q'_o = \sum_{i \in J} g_{io}$ if $o \in P^c$, and $q'' \colon P^c \to \mathbb{R}_+$ is the function $q''_o = q_o - \sum_{i \in J} g_{io}$. Clearly $E_{(J,P)}$ is critical, and $E^{(J,P)}$ is critical if and only if E is critical.

Any feasible allocation for E is the sum of a feasible allocation for $E_{(J,P)}$ and a feasible allocation for $E^{(J,P)}$, so $E_{(J,P)}$ and $E^{(J,P)}$ satisfy the GMC. Conversely, any sum of a feasible allocation for $E_{(J,P)}$ and a feasible allocation for $E^{(J,P)}$ is a feasible allocation for E. Thus a critical pair splits the given allocation problem into two smaller problems of the same type. This is very important because it allows our algorithm to be recursive.

We say that (J,P) is a *minimal critical pair* for E if there is no critical pair (J',P') for E with $J' \subset J$, $P' \subset P$, and $(J',P') \neq (J,P)$. The next result (whose proof follows easily from the discussion above and is therefore left as an exercise) implies that if (J,P) is a minimal critical pair for E, then $E_{(J,P)}$ is simple.

Lemma 2. If $J' \subset J$ and $P' \subset P$, then (J',P') is critical for E if and only if it is critical for $E_{(J,P)}$.

Since (J,P) is a critical pair for E, any feasible allocation m has $m_{io} = 0$ for all $i \in J^c$ and $o \in P$, and in this sense $g_{io} > 0$ is illusory. We say that E is *tight* if $g_{io} = 0$ for all critical pairs (J,P) and all $i \in J^c$ and $o \in P$. The (J,P)*-tightening of* E is $E' = (I, O, q, r, g')$ where $g'_{io} = 0$ if $i \in J^c$ and $o \in P$, and otherwise $g'_{io} = g_{io}$. Since E satisfies the GMC, it has a feasible allocation m, which necessarily has $m_{io} = 0$ for all $i \in J^c$ and $o \in P$, so it is a feasible allocation for E', and consequently E' satisfies the GMC.

A *tightening sequence* for E is a sequence $(J_1,P_1),\ldots,(J_\ell,P_\ell)$ for which there is a sequence $E_0 = E, E_1, \ldots, E_\ell$ of CEE's such that for each $j = 1,\ldots,\ell$, (J_j,P_j) is a critical pair for E_{j-1} and E_j is the (J_j,P_j)-tightening of E_{j-1}. By induction each E_j satisfies the GMC.

The following result is obvious:

Lemma 3. If $E = (I, O, r, q, g)$ satisfies the GMC, (J, P) is a critical pair for E, $g' \leq g$, $E' = (I, O, r, q, g')$, and E' satisfies the GMC, then (J, P) is a critical pair for E'.

In view of the last result, if $(J_1, P_1), \dots, (J_\ell, P_\ell)$ and $(J'_1, P'_1), \dots, (J'_{\ell'}, P'_{\ell'})$ are tightening sequences, then so is $(J_1, P_1), \dots, (J_\ell, P_\ell), (J'_1, P'_1), \dots, (J'_{\ell'}, P'_{\ell'})$. Therefore starting with E and repeatedly tightening with respect to critical pairs, including pairs that become critical as a result of the tightening, until no further tightening is possible, leads to a tight CEE that is independent of the order of tightening, that we call the *tightening of* E.

5 The GCPS Allocation

We now define the GCPS precisely and describe how it can be computed. We work with a fixed CEE $E = (I, O, r, q, g)$ that satisfies the GMC and a profile $\succ = (\succ_i)_{i \in I}$ of strict preferences over O. Recall that for each i, $\alpha_i = \{\, o : g_{io} > 0 \,\}$. For each o let $\omega_o = \{\, i : o \in \alpha_i \,\}$. Let e_i be i's most preferred element of α_i. Let $T = \max_i r_i$. Let θ be the matrix whose entry θ_{io} is 1 if $o = e_i$ and zero otherwise. Let $t^* \geq 0$ be the number such that $\theta t \in R$ if $0 \leq t \leq t^*$ and $\theta t \notin R$ if $t > t^*$, and let $p \colon [0, t^*] \to \mathbb{R}_+^{I \times O}$ be the function $p(t) = \theta t$.

There is a pair (J, P) such that $p(t)$ satisfies inequality (1) for (J, P) strictly if $0 \leq t < t^*$ and violates it if $t > t^*$. At time t^* the requirements of the agents in J can exactly be met by giving them all that they are allowed to consume of objects in P^c and all of the remaining objects in P. The *GCPS allocation* $GCPS(E, \succ)$ is defined[7] (recursively) to be the sum of $p(t^*)$ and the GCPS allocations of $(E - p(t^*))_{(J,P)}$ and $(E - p(t^*))^{(J,P)}$.

The main computational challenge is to compute t^* and a pair (J, P) that becomes critical at that time. According to Proposition 1, the facet inequalities of R, other than those associated with the r_i and g_{io}, are a subset of those given by (1) for various J and P. For a given $P \subset O$ it is easy to find the $J \subset I$ that minimizes the difference between the two sides of (1) for (J, P). An algorithm that searches over all $P \subset O$ has been implemented, and works reasonably well for moderate (roughly $|O| \leq 25$) numbers of schools. It has a computational burden that is roughly proportional to the number $2^{|O|}$ of subsets of O, which makes it unsuitable for very large school choice problems.

Next, we note that t^* is the value of the linear program

$$\max t \quad \text{subject to} \quad (m, t) \in Q \times [0, 1] \text{ and } \theta t \leq m.$$

[7] Proposition 2 and Lemma 2 imply that the choice of (J, P) does not matter if more than one pair becomes critical at t^*, because the overall effect of the recursive descent is to decompose according to the minimal pairs that become critical at time t^*.

This is conceptually significant because there are polynomial time algorithms for linear programming (Khachian, 1979, Karmarkar, 1984) which in turn will imply that there is a polynomial time algorithm for the computation of the GCPS mechanism. However, actual computational experience shows that this approach works quite poorly, even for fairly small problems[8].

We now describe a computational procedure that is not as well founded theoretically, but which works quite well in practice. It computes a piecewise linear function $\overline{p}\colon [0,t^*] \to Q$ such that $p(t) \leq \overline{p}(t)$ for all t. By repeatedly trying to find a way to continue $\overline{p}$, we eventually compute t^* and a pair (J,P) that becomes critical at time t^*.

We first need to find an initial point $\overline{p}(0) \in Q$. A feasible allocation is a *maximal flow* of the network (N_E, A_E) defined in the proof of Theorem 1 in Appendix A. The problem of computing a maximum flow of a network is very well studied, and the literature continues to advance (e.g., Chen et al. (2022)). In practice the push-relabel algorithm of Goldberg and Tarjan (1988) is satisfactory in the sense of not adding significantly to the overall computational burden.

Suppose that we have computed p and $\overline{p}$ on an interval $[0,t_0]$. We will describe a procedure that searches for a matrix $\overline{\theta} \in \mathbb{Z}^{I\times O}$ such that

$$p(t_0) + \theta\varepsilon \leq \overline{p}(t_0) + \overline{\theta}\varepsilon \in Q \tag{$*$}$$

for sufficiently small $\varepsilon > 0$. If the search succeeds, then we let t_1 is the largest number such that $p(t_0) + \theta(t-t_0) \leq \overline{p}(t_0) + \overline{\theta}(t-t_0) \in Q$ for all $t \in [t_0,t_1]$, we define p and $\overline{p}$ on the interval $[t_0,t_1]$ by setting $p(t) = p(t_0) + \theta(t-t_0)$ and $\overline{p}(t) = \overline{p}(t_0) + \overline{\theta}(t-t_0)$, and we iterate the computation with t_1, $p(t_1)$, and $\overline{p}(t_1)$ in place of t_0, $p(t_0)$, and $\overline{p}(t_0)$. If the search fails, our search will uncover a polynomial time algorithm for computing a critical pair for $E - p(t_0)$, so that so $t^* = t_0$.

Fix $\overline{\theta} \in \mathbb{Z}^{I\times O}$. If $\overline{p}(t_0) + \overline{\theta}\varepsilon \in Q$ for small $\varepsilon > 0$ then:

(a) For all i and o, if $o \notin \alpha_i$, then $\overline{\theta}_{io} = 0$.

(b) For all i, $\sum_o \overline{\theta}_{io} = 0$.

Furthermore, $p(t_0) + \theta\varepsilon \leq \overline{p}(t_0) + \overline{\theta}\varepsilon \leq g$ for small $\varepsilon > 0$ if and only if, for all i and o:

(c) If $\overline{p}_{io}(t_0) = p_{io}(t_0)$, then $\overline{\theta}_{io} \geq 0$, and $\overline{\theta}_{io} \geq 1$ if $o = e_i$.

[8] One issue is that, in the standard formulation of a linear program as $\max cx$ subject to $Ax \leq b$ the matrix A has a column for each school, a row for each school's capacity constraint, a row for each student i for the inequality $\sum_{j\in\alpha_i\setminus\{e_i\}} m_{ij} \leq 1$, and a row for each student i for the inequality $t \leq 1 - \sum_{j\in\alpha_i\setminus\{e_i\}} m_{ij}$. Thus A has $(\sum_i |\alpha_i| - |I|)(|O| + 2|I|)$ entries, and thus more than one hundred billion entries when $|I| = 100,000$, $|O| = 500$, and the average value of $|\alpha_i|$ is 6.

(d) If $\overline{p}_{io}(t_0) = g_{io}$, then $\overline{\theta}_{io} \le 0$.

If $\overline{\theta}$ satisfies (a)–(d), then $\overline{p}(t_0) + \overline{\theta}\varepsilon \in Q$ for small $\varepsilon > 0$ if and only if, for all o:

(e) If $\sum_i \overline{p}_{io}(t_0) = q_o$, then $\sum_i \overline{\theta}_{io} \le 0$.

Our search for a $\overline{\theta}$ satisfying (a)–(e) begins by defining an initial $\overline{\theta}^0$ as follows. For each i, if $\overline{p}_{ie_i}(t_0) > p_{ie_i}(t_0)$, then we set $\overline{\theta}^0_{io} = 0$ for all o, and if $\overline{p}_{ie_i}(t_0) = p_{ie_i}(t_0)$, then we set $\overline{\theta}^0_{ie_i} = 1$, we set $\overline{\theta}^0_{io_i} = -1$ for an arbitrary $o_i \ne e_i$ such that $\overline{p}_{io_i}(t_0) > p_{io_i}(t_0)$, and we set $\overline{\theta}^0_{io} = 0$ for all other o. By construction $\overline{\theta}^0$ satisfies (a)–(d).

More generally, suppose that a $\overline{\theta} \in \mathbb{Z}^{I\times O}$ satisfying (a)–(d) is given. For $o \in O$ let

$$J(o) = \{\, i \in \omega_o : \text{if } \overline{p}_{io}(t_0) = p_{io}(t_0), \text{ then } \overline{\theta}_{io} > 0, \text{ and } \overline{\theta}_{io} > 1 \text{ if } o = e_i \,\}$$

be the set of i such that decreasing $\overline{\theta}_{io}$ by one does not result in a violation of (a) or (b). For $i \in I$ let

$$P(i) = \{\, o \in \alpha_i : \text{either } \overline{\theta}_{io} < 0 \text{ or } \overline{p}_{io}(t_0) < g_{io} \,\}$$

be the set of o such that increasing $\overline{\theta}_{io}$ by one does not result in a violation of (b) or (c).

A *pivot* for $\overline{\theta}$ is a sequence $o_0, i_1, o_1, \ldots, o_{h-1}, i_h, o_h$ such that $o_0, \ldots, o_h$ are distinct elements of O, $i_1, \ldots, i_h$ are distinct elements of I, and:

(a') $\sum_i \overline{p}_{io_0}(t_0) = q_o$ and $\sum_i \overline{\theta}_{io_0} > 0$;

(b') for each $g = 1, \ldots, h$, $i_g \in J(o_{g-1})$ and $o_g \in P(i_g)$;

(c') either $\sum_i \overline{p}_{i_h o_h}(t_0) < q_{o_h}$ or $\sum_i \overline{\theta}_{i_h o_h} < 0$.

Given such a pivot, we can define $\overline{\theta}'$ by setting

$$\overline{\theta}'_{i_g o_{g-1}} = \overline{\theta}_{i_g o_{g-1}} - 1 \quad \text{and} \quad \overline{\theta}'_{i_g o_g} = \overline{\theta}_{i_g o_g} + 1$$

for $g = 1, \ldots, h$ and $\overline{\theta}'_{io} = \overline{\theta}_{io}$ for all other (i, o). Since $\overline{\theta}$ satisfies (a) above and $o_{g-1}, o_g \in \alpha_{i_g}$ for all g, $\overline{\theta}'$ satisfies (a). Clearly $\overline{\theta}'$ satisfies (b) because $\overline{\theta}$ satisfies (b) and for each g the decrease of $\overline{\theta}_{i_g o_{g-1}}$ balances the increase of $\overline{\theta}_{i_g o_g}$. Since $\overline{\theta}$ satisfies (c) and (d) above, (b') implies that $\overline{\theta}'$ also satisfies (c) and (d).

We have $\sum_i \overline{\theta}'_{io_0} = \sum_i \overline{\theta}_{io_0} - 1$, $\sum_i \overline{\theta}'_{io_g} = \sum_i \overline{\theta}_{io_g}$ for $g = 1, \ldots, h-1$, and if $\sum_i \overline{p}_{i_h o_h}(t_0) = q_{o_h}$, then $\sum_i \overline{\theta}_{io_h} < 0$ and thus $\sum_i \overline{\theta}_{io_h} \le 0$. Therefore replacing $\overline{\theta}$ with $\overline{\theta}'$ reduces (in an obvious sense) the extent to which (e) is violated. If it is possible, repeating this construction eventually produces an element of $\mathbb{Z}^{I\times O}$ satisfying (a)–(e).

We now describe how the algorithm searches for pivots. Supposing that (e) does not hold, fix o_0 such that $\sum_i \overline{p}_{io_0}(t_0) = q_o$ and $\sum_i \overline{\theta}_{io_0} > 0$. We set $P_0 = \{o_0\}$ and define sets $J_1, P_1, J_2, P_2, \ldots$ by continuing inductively with

$$J_g = \bigcup_{o \in P_{g-1}} J(o) \setminus \bigcup_{f<g} J_f \quad \text{and} \quad P_g = \bigcup_{i \in J_g} P(i) \setminus \bigcup_{f<g} P_f.$$

If there such an h, let h be the smallest integer such that there is an $o_h \in P_h$ such that either $\sum_i \overline{p}_{io_h}(t_0) < q_{o_h}$ or $\sum_i \overline{\theta}_{io_h} < 0$. We can construct a pivot by choosing $i_h \in J_h$ such that $o_h \in P(i_h)$, choosing $o_{h-1} \in P_{h-1}$ such that $i_h \in P(o_{h-1})$, choosing $i_{h-1} \in J_{h-1}$ such that $o_{h-1} \in P(i_{h-1})$, and so forth.

Now suppose that there is no h and $o_h \in P_h$ such that either $\sum_i \overline{p}_{io_h}(t_0) < q_{o_h}$ or $\sum_i \overline{\theta}_{io_h} < 0$. Let $J = \bigcup_g J_g$ and $P = \bigcup_g P_g$. If $o \in P$ and $i \in J^c$, then $i \notin J(o)$, so $\overline{p}_{io}(t_0) = p_{io}(t_0)$. If $i \in J$ and $o \in P^c$, then $o \notin P(i)$, so $\overline{p}_{io}(t_0) = g_{io}$. Thus $\overline{p}(t_0) - p(t_0)$ is a feasible allocation for $E - p(t_0)$ that gives all of the resources in P to agents in J and that gives $g_{io} - p_{io}(t_0)$ to $i \in J$ whenever $o \in P^c$. By Lemma 1, (J, P) is a critical pair for $E - p(t_0)$. Thus our procedure for finding a pivot fails only when $t_0 = t^*$, in which case it computes a critical pair.

Since the computational procedure descends recursively to a small number of simpler subproblems, in assessing its theoretical complexity, the remaining issue is the complexity of using this process to compute a critical pair for $E - p(t^*)$. The number of times that pivots need to be computed while computing t_1 or (J, P) is bounded by the initial value $\sum_{o \in \tilde{P}} \sum_i \overline{\theta}^0_{io}$ where

$$\tilde{P} = \{\, o : \sum_i \overline{p}_{io}(t_0) = q_o \text{ and } \sum_i \overline{\theta}^0_{io} > 0 \,\},$$

which is in turn bounded by the number of i such that $\overline{p}_{ie_i}(t_0) = p_{ie_i}(t_0)$. Without going into any detail, it is clear that the computation of either a pivot or a critical pair has polynomially bounded complexity.

In the linear programming approach, linear programming is used to compute t^*. The procedure above gives a polynomial time method of computing a critical pair for $E - p(t^*)$, so (in conjunction with a polynomial time algorithm for linear programming) it gives a polynomial time algorithm for computing the GCPS allocation.

For our procedure the remaining issue is the number of times that a transition such as the one from t_0 to t_1 is performed. Unfortunately we have been unable to make progress on this problem, and in fact we have been unable to show that our procedure is an algorithm in the sense of necessarily halting after finitely many steps. (As far as we know, it could zigzag infinitely many times while approaching some point of Q

asymptotically.) As we explain in Online Appendix D, it has been implemented and has acceptable running times for problems at the scale of the world's largest school choice problems.

Comparison with linear programming is suggestive. Both our procedure and the simplex method ascend within polytopes along piecewise linear paths, with the next direction at an endpoint of a line segment of the path determined by local information. The famous Klee-Minty cube (Klee and Minty, 1972) shows that the worst case running time of the simplex algorithm is very bad. However, such anomalies are unheard of in practical applications, and the simplex method continues to be competitive with leading versions of interior point methods. A number of theoretical explanations of the typical running time of the simplex algorithm (e.g., Spielman and Teng (2004)) have been advanced.

6 The Market Clearing Cutoffs Mechanism

The *market clearing cutoffs* (MCC) mechanism is applicable when the schools' priorities are not dichotomous. When the schools' priorities are strict, it is equivalent to DA, but when the schools' priorities are coarse, *enhanced MCC mechanisms* (defined in the next section) avoid the inefficiencies that arise when DA is applied after choosing strict priorities that refine the given priorities.

Azevedo and Leshno (2016) study a model with a continuum of students in which the schools' cutoffs are analogous to prices. Our framework is similar, insofar as our agents are, in effect, infinitely divisible. At a technical level, up to a certain point the discussion below follows the corresponding material in Appendix A of their paper. Similar ideas appear in Abdulkadiroğlu et al. (2015).

We take as given a CEE $E = (I, O, r, q, g)$, a profile $\succ = (\succ_i)$ of preferences over O, and a matrix $e = (e_{io})$ of *priorities* that are nonnegative integers, where the objects "prefer" agents with greater priorities.

A *coarse profile* for an object o is a nonnegative integer C_o. Let $\bar{e}$ be an integer that is at least as large as any e_{io}, let $\bar{r}$ be a real number that is at least as large as any r_i, and let $\Psi = \{0, \ldots, \bar{e}\} \times [0, \bar{r}]$. Elements of Ψ are *fine cutoffs*. For i, o, and $\psi_o = (C_o, \rho_o) \in \Psi$, the maximum allowed consumption of o by i is

$$g_{io}(e_{io}, \psi_o) = \begin{cases} 0, & e_{io} < C_o, \\ \min\{\bar{r} - \rho_o, g_{io}\}, & e_{io} = C_o, \\ g_{io}, & e_{io} > C_o. \end{cases}$$

The demand of agent i, as a function of a profile of fine cutoffs $\psi \in \Psi^O$, is generated by having i consume as much of the favorite object as allowed, then as much of the second favorite object as allowed, and so forth, until either she has fulfilled her requirement or she has consumed as much of each object as she is allowed to consume. These demands $D_{io}(\psi)$ of i for the various objects o can be defined implicitly by requiring that:

(a) $0 \le D_{io}(\psi) \le g_{io}(e_{io}, \psi_o)$ for all o;

(b) $\sum_o D_{io}(\psi) \le r_i$;

(c) If either $\sum_o D_{io}(\psi) < r_i$ or there is an o' such that $o' \prec_i o$ and $D_{io'}(\psi) > 0$, then $D_{io}(\psi) = g_{io}(e_{io}, \psi_o)$.

Let $D(\psi)$ be the matrix with entries $D_{io}(\psi)$. For $o \in O$ let $D_o(\psi) = \sum_i D_{io}(\psi)$.

We completely order the elements of Ψ according to restrictiveness by specifying that for $\psi = (C, \rho)$ and $\psi = (C', \rho')$, $\psi \le \psi'$ if either $C < C'$ or $C = C'$ and $\rho' \le \rho$. As usual, $\psi < \psi'$ means that $\psi \le \psi'$ and $\psi' \neq \psi$. For given $\psi_{-o} \in \Psi^{O \setminus \{o\}}$, D_{io} is a nonincreasing function of ψ_o with respect to this order. An obvious but important property of demand, which Azevedo and Leshno (2016) call *gross substitutes*, is that when $o \neq o'$, for a fixed $\psi_{-o'}$, D_{io} is a nondecreasing function of $\psi_{o'}$. It follows that for given ψ_{-o}, D_o is a nonincreasing function of ψ_o, and when $o \neq o'$, for given $\psi_{-o'}$, D_o is a nondecreasing function of $\psi_{o'}$.

We endow Ψ with the topology induced by our ordering: a set is open if it is a union of open intervals (ψ, ψ') and half open intervals $[(0,0), \psi)$ and $(\psi, (\bar{e}, \bar{r})]$, defined as usual. Clearly each D_{io} is continuous with respect to the product topology of Ψ^O.

Let

$$I_o(\psi_{-o}) = \{\, \psi_o : D_o(\psi_o, \psi_{-o}) \le q_o \text{ and } D_o(\psi_o, \psi_{-o}) = q_o \text{ if } \psi_o > (0,0) \,\}$$

be the set of ψ_o that either equate supply and demand for o or allow any eligible agent i to consume up to g_{io} of o without exhausting o's quota. Since D_o is continuous and nonincreasing as a function of ψ_o, and $D_o((\bar{e}, \bar{r}), \psi_{-o}) = 0$, $I_o(\psi_{-o})$ is a closed subinterval of Ψ. It is nonempty, either by the intermediate value theorem if $D_o((0,0), \psi_{-o}) \ge q_o$, or because $D_o((0,0), \psi_{-o}) < q_o$ so that $(0,0) \in I_o(\psi_{-o})$.

Let $J \colon \Psi^O \to \Psi^O$ be the function with component functions

$$J_o(\psi) = \begin{cases} \min I_o(\psi_{-o}), & \psi_o < \min I_o(\psi_{-o}), \\ \psi_o, & \psi_o \in I_o(\psi_{-o}) \\ \max I_o(\psi_{-o}), & \max I_o(\psi_{-o}) < \psi_o. \end{cases}$$

The set of fixed points of J is $\mathcal{F} = \{ \psi \in \Psi : \psi_o \in I_o(\psi) \text{ for all } o \}$. In the usual sense $\mathcal{F}$ is the set of ψ such that markets clear:

$$\mathcal{F} = \{ \psi \in \Psi : \text{for each } o,\ D_o(\psi) \leq q_o \text{ and } \psi_o = (0,0) \text{ if } D_o(\psi) < q_o \}.$$

We note that this concept embeds a variant of stability: there is no agent-object pair (i, o) such that i is eligible to consume o in the sense that $e_{io} \geq C_o$, i is able to consume additional o insofar as $D_{io}(\psi) < g_{io}$, i is assigned a positive amount of an object that i likes less than o, and either o has unused capacity or $e_{io} > C_o$.

We partially order Ψ^O by specifying that $\psi \leq \psi'$ if $\psi_o \leq \psi'_o$ for all o. For $\psi, \psi' \in \Psi^O$, $\psi \vee \psi'$ and $\psi \wedge \psi'$ are the elements of Ψ^O with components $(\psi \vee \psi')_o = \max\{\psi_o, \psi'_o\}$ and $(\psi \wedge \psi')_o = \min\{\psi_o, \psi'_o\}$. A set $L \subset \Psi^O$ is a *lattice* if, for all $\psi, \psi' \in L$, $\psi \vee \psi'$ and $\psi \wedge \psi'$ are elements of L, and it is a *complete lattice* if, in addition, for every $S \subset L$, L contains a least upper bound $\bigvee S$ and a greatest lower bound $\bigwedge S$ of S. For $\emptyset \neq S \subset \Psi^O$, the elements of Ψ^O with components $\inf\{ \psi_o : \psi \in S \}$ and $\sup\{ \psi_o : \psi \in S \}$ are the greatest lower bound and least upper bound of S, and the elements $\psi^l, \psi^u \in \Psi^O$ with components $\psi^l_o = (0,0)$ and $\psi^u_o = (\bar{e}, \bar{r})$ for all o are the least upper bound and greatest lower bound of $\emptyset$, so Ψ^O is itself a complete lattice.

Lemma 4. J is a weakly increasing function: if $\psi \leq \psi'$, then $J(\psi) \leq J(\psi')$.

Proof. For $o' \neq o$, $D_o(\psi)$ is a nondecreasing function of $\psi_{o'}$, so the lower and upper bounds of $I_o(\psi_{-o})$ are nondecreasing functions of $\psi_{o'}$, and thus $J_o(\psi)$ is a nondecreasing function of each of these components. Of course for given ψ_{-o}, $J_o(\psi)$ is also a nondecreasing function of ψ_o. □

Now Tarski's fixed point theorem implies that $\mathcal{F}$ is a complete (and thus nonempty) lattice. The *market clearing cutoffs* (MCC) *mechanism* allocation for E, $\succ$, and e is

$$MCC(E, \succ, e) = D(\bigwedge \mathcal{F}).$$

In general $MCC(E, \succ, e)$ need not be a feasible allocation. For example, if agent i's eligible objects all have high demand, it can happen that for each of them the only possible values of the coarse cutoff are above the agent's priority at the object, in which case $MCC_{io}(E, \succ, e) = 0$ for all o. If, for an agent i, there is some object o such that $g_{io} \geq r_i$ and $\sum_{j : e_{jo} \geq e_{io}} r_j \leq q_o$, then o is a *safe object* (for MCC) for i. If every agent has a safe object, then $MCC(E, \succ, e)$ is a feasible allocation. In the context of school choice this notion of a safe school is too restrictive when priorities are dichotomous because it requires that every student is eligible at a school whose number of eligible students is not greater than its capacity, and for this reason the GCPS mechanism is more appropriate in that case.

A natural method of approximating the minimal fixed point of a function satisfying the hypotheses of the Tarski fixed point theorem is to iterate the function beginning at a point below the minimal fixed point: set $\psi^0 = \psi^l$, and for $t > 0$ set $\psi^t = J(\psi^{t-1})$. Since $\psi^0 \le \psi^1$, $0 \le \bigwedge \mathcal{F}$, J is monotonic, and $\bigwedge \mathcal{F}$ is a fixed point, by induction $\psi^t \le \psi^{t+1} \le \bigwedge \mathcal{F}$ for all t. Since $\{\psi^t\}$ is monotonic and bounded above, it must converge to a point $\psi^* \le \bigwedge \mathcal{F}$. Since D is continuous, ψ^* must be a fixed point of J, so $\psi^* = \bigwedge \mathcal{F}$. (Simple examples show that this procedure need not converge in finitely many steps.) We expect the distance from the iterate to the solution to decrease geometrically, and last part of the description of an algorithm is to specify that it terminates when the distance is less than some chosen quantity. This algorithm has been implemented (see Online Appendix D) and is fast enough for practical purposes.[9]

Example 1. Suppose that $I = \{1, 2, 3\}$, $O = \{a, b, c\}$, $r_i = 1$, $q_o = 1$, $g_{io} = 1$, and $e_{io} = 0$ for all i and o, with preferences $a \succ_1 c \succ_1 b$, $a \succ_2 b \succ_2 c$, and $c \succ_3 a \succ_3 b$. Let $\psi^0 = (\psi^0_a, \psi^0_b, \psi^0_c)$ where (with $\overline{r} = 1$) $\psi^0_a = \psi^0_b = \psi^0_c = (0, 0)$. Then $D_a(\psi^0) = 2$, $D_b(\psi^0) = 0$, and $D_c(\psi^0) = 1$, so $\psi^1 = (\psi^1_a, \psi^0_b, \psi^0_c)$ where $\psi^1_a = (0, \frac{1}{2})$. We have $D_a(\psi^1) = 1$, $D_b(\psi^0) = \frac{1}{2}$, and $D_c(\psi^0) = \frac{3}{2}$, so $\psi^2 = (\psi^1_a, \psi^0_b, \psi^2_c)$ where $\psi^2_c = (0, \frac{1}{2})$. Now $D_a(\psi^2) = \frac{3}{2}$, $D_b(\psi^2) = \frac{1}{2}$, and $D_c(\psi^2) = 1$, so $\psi^3 = (\psi^3_a, \psi^0_b, \psi^2_c)$ where $\psi^3_a = (0, \frac{2}{3})$. Since $D_a(\psi^3) = D_b(\psi^3) = D_c(\psi^3) = 1$, the MCC allocation is $\frac{1}{3}a + \frac{1}{6}b + \frac{1}{2}c \to 1$, $\frac{1}{3}a + \frac{2}{3}b \to 2$, and $\frac{1}{3}a + \frac{1}{6}b + \frac{1}{2}c \to 3$.

Now suppose that we have a school choice CEE: $r_i = 1$ for all i and $g_{io} \in \{0, 1\}$ for all i and o. If the priorities are strict ($e_{io} \neq e_{i'o}$ for all o and distinct i and i' such that $g_{io} = 1 = g_{i'o}$) then the computational procedure above coincides with the student proposes deferred acceptance algorithm. Thus the MCC mechanism agrees with DA when the priorities are strict.

An alternative to the MCC mechanism is to assign the allocation $D(\bigvee \mathcal{F})$ to E, $\succ$, and e. In the school choice context this coincides with school proposes deferred acceptance when priorities are strict. The main motivations for preferring DA to school proposes deferred acceptance are that it is strategy-proof for the students and gives the student optimal stable matching. However, we will see that in general the MCC mechanism is not strategy-proof, and it will be evident that the alternate given by $D(\bigvee \mathcal{F})$ is

[9]There is an active literature (e.g., Fearnly et al. (2022) and references therein) on other algorithms for computing Tarski fixed points in finite search spaces. For functions on continuous search spaces, complexity results are likely to depend on additional hypotheses imposed on the function, and it seems that such problems are less well studied.

also strategy-proof in the large, as this concept is defined in Section 8, so $D(\bigvee \mathcal{F})$ may be preferred if the social values expressed by the priorities are sufficiently important.

He et al. (2018) present a variant of the Hylland and Zeckhauser (1979) pseudo-market mechanism for this problem that also achieves an allocation that fulfills the given coarse priorities. It takes the agents' cardinal utilities as inputs (the MCC mechanism avoids the practical difficulties associated with the elicitation of cardinal utilities) and its allocations are efficient in a stronger sense than sd-efficiency, and fair in a strong sense (He et al., 2015) that is defined in terms of cardinal preferences. Insofar as the allocation it produces is part of a Kakutani fixed point, algorithms for computing it are likely to be significantly harder to program, and may have unacceptable running times on large problems.

7 Efficiency

This section studies the efficiency properties of the GCPS and MCC mechanisms. At the outset we should mention that mechanisms that are *ordinal* (that is, based on the agents' reports of ordinal preferences) and nondictatorial often allow allocations that are inefficient in comparison with those produced by mechanisms taking cardinal utility functions as inputs (Featherstone and Niederle, 2008, Miralles, 2009, Abdulkadiroğlu et al., 2011, Troyan, 2012, Abdulkadiroğlu et al., 2015, Ashlagi and Shi, 2016, He et al., 2018).

We work with a CEE $E = (I, O, r, q, g)$ and a fixed profile of preferences $\succ$. Our first objective is to show that if E satisfies the GMC, then the GCPS mechanism applied to E and $\succ$ yields an allocation that is efficient in a strong sense. The MCC mechanism does not yield allocations that are efficient in this sense, but there will be algorithms that pass from the MCC allocation to allocations that are efficient relative to the given priorities.

For $i \in I$, an *allocation for* i is a vector $m_i \in \mathbb{R}^O_+$ such that $m_i \le g_i$ and $\sum_o m_{io} = r_i$. The *stochastic dominance relation* $sd(\succ_i)$ on allocations for i derived from $\succ_i$ is defined by $m'_i \; sd(\succ_i) \; m_i$ if $\sum_{p \succeq_i o} m'_{ip} \ge \sum_{p \succeq_i o} m_{ip}$ for all $o \in O$. Usually in applications of this concept m_i is a probability distribution on O, but the concept makes perfect sense in our more general context, and standard arguments generalize straightforwardly to show that $m'_i \; sd(\succ_i) \; m_i$ if and only if $\sum_o m'_{io} u_i(o) \ge \sum_o m_{io} u_i(o)$ for any cardinal utility function $u_i : O \to \mathbb{R}$ such that for all $o, o' \in O$, $u_i(o) \ge u_i(o')$ if and only if $o \succeq_i o'$.

Two other well-studied extensions of a given preference to preferences over lotteries relate to lexicographic preferences (Cho, 2016; Schulman and Vazirani, 2015; Cho and Doğan, 2016; Saban and Sethuraman, 2014; Cho, 2018). The first extension, which is called the *downward lexicographic* extension (dl-extension) compares two i-allocations as follows. One of the i-allocations is preferred if it assigns a higher amount of the most preferred object than the other. If the two i-allocations assign the same amount of the most preferred object, then the one that is preferred is the one that assigns the greater amount of the second most preferred object. If the two amounts are equal again, then the i-allocation that assigns a greater amount of the third most preferred object is preferred, and so on. The second extension, which is called the *upward lexicographic* extension (ul-extension) is a dual of the dl-extension. It lexicographically minimizes amounts of less preferred objects, starting from the least preferred object.[10] The dl- and ul-extensions yield preferences that represent the limits of standard vNM utility functions with extreme risk loving and extreme risk aversion, respectively.

For $d \in \{sd, dl, ul\}$, a feasible allocation m' *d-dominates* another feasible allocation m if $m'_i \, d(\succ_i) \, m_i$ for all i and there is some i such that $m'_i \neq m_i$. A feasible allocation m is *d-efficient* if there is no feasible allocation that d-dominates it. The following result is essentially due to Cho and Doğan (2016). (Lemma 3 of BM is a precursor.) We provide no proof because it is easy to see that their proof works, essentially without any modification, in our more general setting.

Lemma 5. sd-efficiency, dl-efficiency, and ul-efficiency are equivalent.

The next result now follows from the sd-efficiency of the GCPS allocation, which is a special case of Proposition 3 of Balbuzanov (2022).

Theorem 2. For $d \in \{sd, dl, ul\}$, the GCPS allocation for E and $\succ$ is *d-efficient*.

We now study the constrained (by the priorities) efficiency of the MCC mechanism. Let e be a matrix of priorities, and let $\psi \in \Psi^O$ be a profile of fine cutoffs. An allocation m_i for i is an *(e, ψ)-allocation for i* if, for all o, $m_{io} \leq g_{io}(e_{io}, \psi_o)$. For $d \in \{sd, dl, ul\}$ and (e, ψ)-allocations m_i and m'_i for i, m_i is *(e, ψ, d)-dominated* by m'_i if $m'_i \, d(\succ_i) \, m_i$. If, in addition, $m'_i \neq m_i$, then m'_i *strictly (e, ψ, d)-dominates* m_i. The following is an obvious consequence of the construction of individual demand.

Lemma 6. For each $d \in \{sd, dl, ul\}$, every (e, ψ)-allocation for i is (e, ψ, d)-dominated by $D_i(\psi)$.

[10] Formally, the *downward lexicographic relation* $dl(\succ_i)$ on allocations for i derived from $\succ_i$ is defined by specifying that $m'_i \, dl(\succ_i) \, m_i$ if either $m'_i = m_i$ or there is an $o \in O$ such that $\sum_{p \succeq_i o'} m'_{ip} = \sum_{p \succeq_i o'} m_{ip}$ for all $o' \in O$ such that $o' \succ_i o$ and $\sum_{p \succeq_i o} m'_{ip} > \sum_{p \succeq_i o} m_{ip}$. The *upward lexicographic relation* $ul(\succ_i)$ is defined by specifying that $m'_i \, ul(\succ_i) \, m_i$ if either $m'_i = m_i$ or there is an $o \in O$ such that $\sum_{o' \succeq_i p} m'_{ip} = \sum_{o' \succeq_i p} m_{ip}$ for all $o' \in O$ such that $o \succ_i o'$ and $\sum_{o \succeq_i p} m'_{ip} < \sum_{o \succeq_i p} m_{ip}$.

A partial allocation p is an (e,ψ)-*allocation* if each p_i is an (e,ψ)-allocation for i. For $d \in \{sd, dl, ul\}$, a (e,ψ)-allocation p is (e,ψ,d)-*efficient* if there is no (e,ψ)-allocation p' such that p'_i (e,ψ,d)-dominates p_i for all i, and p'_i strictly (e,ψ,d)-dominates p_i for some i. Lemma 6 implies that:

Proposition 3. If $p^* = MCC(E,\succ,e)$, and ψ is a vector of fine cutoffs such that $D(\psi) = p^*$, then, for any $d \in \{sd, dl, ul\}$, p^* is (e,ψ,d)-efficient.

Example 2. Suppose that $I = \{1,2,3,4\}$ and $O = \{a,b,c\}$, with $r_i = 1$ and $g_{io} = 1$ for all i and o. Suppose that $q_a = q_b = 1$ and $q_c = 4$, $e_{ia} = e_{ib} = 0$ and $e_{ic} = 1$ for all i, and the preferences are $a \succ_1 b \succ_1 c$, $a \succ_2 b \succ_2 c$, $b \succ_3 a \succ_3 c$, and $b \succ_4 a \succ_4 c$. The MCC fine cutoffs are $\psi^0_a = \psi^0_b = (0, 3/4)$ and $\psi^0_c = (0,0)$, and the MCC allocation is $\frac{1}{4}a + \frac{1}{4}b + \frac{1}{2}c \to i$ for all i.

Let $\pi\colon \psi \mapsto (C_o)_{o\in O}$ be the natural projection onto profiles of coarse priorities, and for $\psi \in \Psi^O$ with components $\psi_o = (C_o, \rho_o)$ let ψ^* be the element of Ψ^O with components $\psi^*_o = (C_o, \bar{r})$. As Example 2 illustrates, if i and j are both marginal at o and o', in the sense that $e_{io}, e_{jo} = C_o$ and $e_{io'}, e_{jo'} = C_{o'}$, then it can happen that p^*_{io}, $p^*_{io'}$, p^*_{jo}, and $p^*_{jo'}$ are all positive. If $o \succ_i o'$ and $o' \succ_j o$, then a Pareto improving trade is possible. Since i and j have already been judged to be equally worthy of o and o', such a trade is unambiguously welfare improving. More complicated trades may also be possible, and the social desirability of these is less clear. Allowing all mutually improving trades that do not give objects to agents who are unqualified to consume them leads to the notion of d-efficiency within the set of feasible allocations that fulfill $\pi(\psi)$, which is (e,ψ^*,d)-efficiency.

Let m be a feasible (e,ψ)-allocation. We form a directed graph $G = (V, A)$ whose set of *vertices* is $V = \{\,(i,o) \in I \times O : m_{io} > 0\,\}$ and whose set A of *arcs* is the set of $((i,o),(j,p)) \in V \times V$ such that $p \succ_i o$, $e_{ip} \geq C_p$, and $m_{ip} < g_{ip}$. The following is a variant of Lemma 3 of Bogomolnaia and Moulin (2001). Since the argument is simple, and similar to theirs, we do not provide a proof.

Lemma 7. For any $d \in \{sd, dl, ul\}$, m is (e,ψ^*,d)-efficient if and only if G is acyclic.

Suppose we have found a cycle $(i_1,o_1),\ldots,(i_k,o_k)$ of G. Let

$$\Delta = \min\{m_{i_1,o_1},\ldots,m_{i_k,o_k}, g_{i_1,o_2} - m_{i_1,o_2},\ldots,g_{i_{k-1},o_k} - m_{i_{k-1},o_k}, g_{i_k,o_1} - m_{i_k,o_1}\},$$

and let m' be the allocation obtained from m by decreasing each m_{i_h,o_h} by Δ and increasing each $m_{i_h,o_{h+1}}$ (and m_{i_k,o_1}) by Δ. If $G' = (V',A')$ is the directed graph for m', defined as above, then A' is a proper subset of A. Therefore repeating this maneuver leads, in finitely many steps, to an (e, ψ^*, d)-efficient allocation. In fact there are a great many algorithms that pass from an arbitrary allocation that fulfills ψ to an (e, ψ^*, d)-efficient allocation. How to do this in a way that might be regarded as fairest or most symmetric seems like an interesting, but complicated, issue. We define an *enhanced MCC mechanism* to be any mechanism that first computes the MCC allocation p^* and then uses some algorithm to pass from p^* to an (e, ψ^*, d)-efficient allocation. In the final allocation of Example 1 the only mutually beneficial trade is between agents 1 and 3, and the unique enhanced MCC allocation is $\frac{2}{3}a + \frac{1}{6}b + \frac{1}{6}c \to 1$, $\frac{1}{3}a + \frac{2}{3}b \to 2$, and $\frac{1}{6}b + \frac{5}{6}c \to 3$

Superficially the passage from the MCC allocation to an enhanced MCC allocation resembles the procedure of Kesten (2010) (and the extensive literature (Tang and Yu, 2014, Dur et al., 2019, Ehlers and Morrill, 2020, Troyan et al., 2020, Tang and Zhang, 2021, Reny, 2022, Cerrone et al., 2024) following up on his work) which passes from the DA allocation, which is a deterministic assignment, to an efficient assignment. However, the closest point of comparison is actually with the GCPS mechanism applied without safe schools, with each student ranking all schools, in a setting in which every student is eligible to attend any school, and the number of students is not greater than the sum of the schools' capacities, so that each student is guaranteed a seat in some school. Kesten's mechanism begins by generating priorities, which we may assume are random, and running DA. The allocation is then modified by repeated trades along improving cycles, in a way that requires that some agents relinquish their priorities, but that never leads to worse outcomes for those who do so, and thus may be voluntary, or at least imagined to be voluntary. In the GCPS mechanism randomization is deferred to the implementation (Online Appendix C) stage, and it is never necessary to endow students with things that they will subsequently be asked to relinquish. In addition, GCPS achieves sd-efficiency, which is stronger than ex post efficiency.

8 Strategy-Proofness

We now discuss the issue of manipulability for the GCPS, MCC, and enhanced MCC mechanisms. The predominant character of the GCPS and MCC mechanisms is that the student is given probability of a seat in the schools that she says she most prefers, to the extent possible. In order for a manipulation to succeed, the manipulation must affect the student's outcome, which means that the student is actually consuming the schools in the order of reported preference rather than actual preference. Thus we should expect that opportunities to successfully manipulate are rare, the gains from successful manipulations are slight, and attempts at manipulation are risky, often resulting in the student getting what she said she wants instead of what she actually wants.

A mechanism (understood as a map from profiles of revealed preferences to probability distributions over outcomes) is *strategy-proof* if reporting a false preference never results in higher expected utility, or, equivalently, always gives a distribution over outcomes that is stochastically dominated, for the actual preferences, by the distribution resulting from truthful revelation. We will see that none of our mechanisms is strategy-proof in this exact sense.

We begin with a simple example illustrating how GCPS can fail to be strategy-proof. Suppose that a and b are the agent's first and second favorite object, with $q_a = q_b = 1$, and there are $A - 1$ other people who have a as their favorite and $B - 1$ other people who have b as their favorite, where $1 < A < B$. Further, assume that no one outside the set of agents who have a as their favorite will ever consume any a and no one outside the set of agents who have b as their favorite will ever consume any b. If the agent reports the truth she will receive $\frac{1}{A}$ units of a and none of b. If she reports that b is her favorite and a is her second favorite, then she will consume b between time 0 and time $\frac{1}{B}$ while $\frac{A-1}{B}$ units of A are being consumed by others, and then she will receive $\frac{1}{A}(1 - \frac{A-1}{B})$ units of a, so her total consumption of a and b will be $\frac{1}{A}(1 + \frac{1}{B})$. This can be an improvement if the utility difference between a and the agent's third favorite object is more than A times the utility difference between a and b.

Proposition 1 of BM asserts (among other things) that the PS mechanism is *weakly strategy-proof*: reporting false preference never gives an allocation that is strictly sd-preferred to the allocation resulting from truthful revelation. Kojima (2009) gives an example (p. 138) that shows weak strategy-proofness does not extend to the allocation of $r \geq 2$ objects per agent.

There are three different ways a student might try to manipulate: a) reporting that some of the schools that are actually worse than the safe school are better than it; b) reporting that some of the schools that are actually better than the safe school are worse; c) reordering of the schools that are better than the safe school. A manipulation attempt of type a) will be called an *augmentation*; following Roth and Rothblum (1999), a manipulation attempt of type b) will be called a *truncation*; a manipulation attempt of type c) will be called a *reordering*.

Yılmaz (2010) (Example 5) presents the following example of an unambiguously successful truncation manipulation of GCPS for a house allocation problem with existing tenants.[11] There are three homeowners and three houses, with 1 endowed with a, 2

[11]Theorem 3 of Cho (2018) asserts that the PS mechanism is dl-strategy proof, which means that manipulation never results in a dl-better allocation. This example shows that Cho's result does not extend to house allocation problems with existing tenants.

endowed with b, and 3 endowed with c, and preferences

$$b \succ_1 c \succ_1 a, \quad a \succ_2 b, \quad b \succ_3 a \succ_3 c.$$

In the GCPS process b is exhausted at time $\frac{1}{2}$, which results in $P = \{a\}$ becoming critical, with $J_P = \{2\}$, so the GCPS allocation gives $\frac{1}{2}b + \frac{1}{2}c$ to 1, a to 2, and $\frac{1}{2}b + \frac{1}{2}c$ to 3. If 1 reports the preference $b \succ'_1 a$ (i.e., $b \succ'_1 a \succ'_1 c$) then $P = \{a, b\}$ is critical at time 0, and the allocation gives b to 1, a to 2, and c to 3.

As Yılmaz points out, this manipulation continues to be possible in problems obtained by replicating each agent a certain number of times. For example, suppose that there are four copies of agents of type 1 and 5 copies each of agents of types 2 and 3. With truthful revelation, $\{a, b\}$ becomes critical at time $\frac{4}{9}$, and the allocation is $\frac{4}{9}b + \frac{5}{9}c$ to agents of types 1 and 3 and $\frac{4}{5}a + \frac{1}{5}b$ to agents of type 2. Again, an agent of type 1 who deviates to $b \succ'_1 a$ gets only b.

For a simple example of how MCC can fail to be strategy-proof, consider that in Example 1, if agent 1 reports the preference $a \succ_1 b \succ_1 c$, then the MCC allocation is $\frac{1}{2}a + \frac{1}{2}b \to 1$, $\frac{1}{2}a + \frac{1}{2}b \to 2$, and $c \to 3$. This can be a successful manipulation if 1 is predominantly concerned about the probability of a.

The remainder of the section presents two positive theoretical results concerning the manipulability of the GCPS and MCC mechanisms. Our main result is that both the GCPS and MCC mechanisms are strategy-proof in the large, roughly as this concept is defined by Azevedo and Budish (2019) (AB). In order to explain this we briefly review their (appropriately modified) definitions, beginning with notation for probability.

For a measurable space X, $\Delta(X)$ is the set of probability measures on X. If X is finite, then $\Delta^\circ(X)$ is the set of elements of $\Delta(X)$ that assign positive probability to each element of X. For $x \in X$, let $\delta_x \in \Delta(X)$ be the *Dirac measure* that assigns all probability to x.

In our model there is a finite set T of *ordinal preference types*. In our model there is also a finite set O of *objects*. Each $t \in T$ there is a set U_t of vNM utility functions $u_t \colon O \to [0, 1]$ that are consistent with t; let u_t also denote the induced function on $\Delta(O)$.

For each integral *market size* $n \geq 1$ let $\chi^n \colon T^n \to \Delta(T)$ be the function

$$\chi^n(t_1, \dots, t_n) = \sum_{i=1}^{n} \tfrac{1}{n}\delta_{t_i},$$

and let $\Delta^n(T) = \chi^n(T^n)$. Define $\theta^n : T \times \Delta(T) \to \Delta(\Delta^n(T))$ by

$$\theta^n(t_1, m) = \sum_{t_{-1} \in T^{n-1}} \delta_{\chi^n(t_1, t_{-1})} \cdot \Pr(t_{-1} | t_{-1} \sim iid(m))$$

where $iid(m)$ is the measure on T^{n-1} generated by $n-1$ independent draws distributed according to m. For each n, $t, t' \in T$, and $m \in \Delta(T)$ the *total absolute difference* between the distributions generated by t and t' is

$$D^n(t, t', m) = \sum_{\tau \in \Delta^n(T)} |\theta^n(t, m)(\tau) - \theta^n(t', m)(\tau)|.$$

Lemma A.1 of AB asserts that for any $m \in \Delta^\circ(T)$ and $\varepsilon > 0$ there is a constant $C(m, \varepsilon) > 0$ such that $D^n(t, t', m) < C(m, \varepsilon) n^{-1/2+\varepsilon}$ for all n and $t, t' \in T$.

A (direct) *mechanism* is a sequence $\{\Phi^n\}$ where $\Phi^n : T^n \to \Delta(O)^n$. Let Σ_n be the set of permutations $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$. For $\sigma \in \Sigma_n$ let σ also denote the function $\sigma : T^n \to T^n$ given by $\sigma(t_1, \ldots, t_n) = (t_{\sigma(1)}, \ldots, t_{\sigma(n)})$, and define $\sigma : \Delta(O)^n \to \Delta(O)^n$ similarly. For the time being attention is restricted to mechanisms that are *anonymous*, meaning that for each n, Φ_n is invariant with respect to permutations of its arguments:

$$\Phi^n(\sigma(t_1, \ldots, t_n)) = \sigma(\Phi^n(t_1, \ldots, t_n))$$

for all n, $\sigma \in \Sigma_n$, and $(t_1, \ldots, t_n) \in T^n$. Anonymity implies that for each n there is a function $\tilde{\Phi}^n : T \times \Delta^n(T) \to \Delta(O)$ such that $\Phi^n_i(t_1, \ldots, t_n) = \tilde{\Phi}^n(t_i, \chi^n(t_1, \ldots, t_n))$ for all $(t_1, \ldots, t_n)$ and i. Define $\phi^n : T \times \Delta(T) \to \Delta(O)$ by

$$\phi^n(t, m) = \sum_{\tau \in \Delta^n(O)} \tilde{\Phi}^n(t, \tau) \cdot \theta^n(t, m)(\tau).$$

The mechanism $\{\Phi^n\}$ is *strategy-proof in the large* (SP-L) if, for any $m \in \Delta^\circ(T)$ and $\varepsilon > 0$, there is an integer n_0 such that for all $n > n_0$, $t, t' \in T$, and $u_t \in U_t$,

$$u_t(\phi^n(t, m)) > u_t(\phi^n(t', m)) - \varepsilon.$$

As AB explain (p. 88) this notion weakens strategy-proofness in two ways: a) manipulation is only required to be approximately unrewarding asymptotically as n increases; b) agents are assumed to have weak information concerning the distribution of reports. Of course for school choice there are typically many students per school, and students and their parents do not have detailed information concerning the preferences of others.

The reason that the GCPS and MCC mechanisms are SP-L is familiar: each agent's outcome depends on her own report and on equilibrating variables that depend on the distribution of all agents' reports. Let P denote a space of equilibrating variables. We assume that there are functions $\xi \colon T \times P \to \Delta(O)$ and $f^n \colon \Delta^n(T) \to P$ for each n such that $\tilde{\Phi}^n(t, \tau) = \xi(t, f^n(\tau))$ for all $t \in T$ and $\tau \in \Delta^n(T)$.

We also assume that $u_t(\xi(t, p)) \geq u_t(\xi(t', p))$ for all $t, t' \in T$, $u_t \in U_t$, and $p \in P$. For the GCPS mechanism an element of P is a schedule of times at which pairs become critical, and for the MCC mechanism an element of P is a profile of fine cutoffs, so this is indeed the case. Since any benefit of manipulation must be the result of its effect on the distribution of distributions of types, Lemma A.1 of AB now implies that $\{\Phi^n\}$ is SP-L. Although the details of our framework are somewhat different, this result can also be understood, at least in principle, as a consequence of Theorem 1 of AB because our assumptions imply that $\{\Phi^n\}$ is *envy-free*: for all n, $(t_1, \ldots, t_n) \in T^n$, $i, j = 1, \ldots, n$, and $u_{t_i} \in U_{t_i}$ and $u_{t_j} \in U_{t_j}$, if $p = f^n(\chi^n(t_1, \ldots, t_n))$, then

$$u_{t_i}(\Phi_i^n(t_1, \ldots, t_n))) = u_{t_i}(\xi(t_i, p)) \geq u_{t_i}(\xi(t_j, p)) = u_{t_i}(\Phi_j^n(t_1, \ldots, t_n)).$$

As AB explain in their Appendix C, these definitions and results extend straightforwardly to mechanisms that are semi-anonymous, as defined by Kalai (2004), in which the set of agents is divided into finitely many groups, with each group having a different set of ordinal preference types. For the GCPS mechanism a natural group would be a set of students that have the same safe school, and for the MCC mechanism a natural group would be a set of students that have the same priorities. In this sense we now have:

Theorem 3. The GCPS and MCC mechanisms are SP-L.

We should mention that an enhanced MCC mechanism may allow a manipulation in which a student upgrades the ranking of a school that is popular, in order to increase her allocation of it coming from the MCC mechanism, because she is confident that during the reallocation-to-attain-efficiency phase she will be able to trade this allocation for probability of schools she desires. The effect of this type of manipulation on the fairness of the mechanism is ambiguous: the manipulating student may be regarded as taking advantage of a flaw of the mechanism, or she may be regarded as partially compensating for a penalty that the mechanism imposes on students whose preferences are unusual. On the whole this type of manipulation seemingly does little to change the overall character of an enhanced MCC mechanism.

Our other theoretical result is that manipulation of the GCPS mechanism by augmentation is unambiguously unsuccessful. We fix an integral school choice CEE $E =$

$(I, O, 1, q, g)$ that satisfies the GMC and a profile $\succ$ of preferences over O. We assume that the safe school is the $\succ_j$-worst element of α_j. We also fix a particular student $i \in I$ whose possible deviations from truthful reporting are studied.

Theorem 4. Let $\alpha'_i = \alpha_i \cup \{o^*\}$, where o^* is an element of $O \setminus \alpha_i$, and let $\succ'_i$ be a preference over O that has α'_i as the set of schools weakly preferred to the safe school, and that agrees with $\succ_i$ on α_i. Let $\succ' = (\succ'_i, \succ_{-i})$. Then

$$GCPS_i(\succ)\, sd(\succ_i)\, GCPS_i(\succ').$$

The primary interest of this result is probably its proof (in Appendix B) which lays out a detailed analytic framework for understanding the impact of one agent changing her consumption of an object incrementally.

9 Fairness

We now briefly consider fairness properties of the GCPS and MCC mechanisms. Let $E = (I, O, 1, q, g)$ be a school choice CEE that satisfies the GMC, and let $\succ$ be a profile of preferences. It is obvious from the definition that the GCPS mechanism satisfies anonymity (the outcomes do not depend on the ordering of the agents, or their "names") and equal treatment of equals (the GCPS gives the same allocations to i and j if $r_i = r_j$, $g_i = g_j$, and $\succ_i = \succ_j$).

The other fairness property considered by BM is envy-freeness. They show that the PS mechanism is envy-free in the strong sense that if m_i and m_j are the allocations of the PS mechanism for $\succ$, then $m_i\, sd(\succ_i)\, m_j$. It is not reasonable to expect this if the two agents have different opportunities, and in recognition of this Abdulkadiroğlu and Sönmez (2003) introduced a notion of no justified envy. This concept takes on different meanings depending on the setting. (For a recent discussion see Romm et al. (2020).) We follow Yılmaz (2010) in the context of school choice. If E is a school choice CEE with $g \in \{0,1\}^{I \times O}$, we say that $m \in Q$ *has no justified envy* if, for all $i, j \in I$, if $\alpha_i \subset \alpha_j$ and $o_i \succ_i o_j$ for all $o_i \in \alpha_i$ and $o_j \in \alpha_j \setminus \alpha_i$, then $m_i\, sd(\succ_i)\, m_j$. Intuitively, i's envy of j is not justified if i is not eligible to attend a desirable element of α_j, or if j can demand a seat in a desirable element of α_i because less desirable elements of α_i are not in α_j.

Proposition 4. If E is a school choice CEE with $g \in \{0,1\}^{I \times O}$, then $GCPS(\succ)$ has no justified envy.

In addition to E and $\succ$, suppose that e is a matrix of priorities. As above, it is easy to see that the MCC mechanism satisfies anonymity and equal treatment of equals. In this context we say that a feasible allocation m *has no justified envy* if, for all $i, j \in I$, if $e_{io} \geq e_{jo}$ for all o, $\alpha_i \subset \alpha_j$, and $o_i \succ_i o_j$ for all $o_i \in \alpha_i$ and $o_j \in \alpha_j \setminus \alpha_i$, then $m_i \, sd(\succ_i) \, m_j$. Suppose that c is a profile of fine cutoffs and m is a feasible allocation that fulfills c. If $e_{io} \geq e_{jo}$ for all o, $\alpha_i \subset \alpha_j$, and $o_i \succ_i o_j$ for all $o_i \in \alpha_i$ and $o_j \in \alpha_j \setminus \alpha_i$, then $m_{io} \geq m_{jo}$ for all $o \in \alpha_i$ such that $\sum_{o' \succeq_i o} m_{io'} < 1$, and thus $m_i \, sd(\succ_i) \, m_j$. Therefore:

Proposition 5. If E is a school choice CEE with $g \in \{0, 1\}^{I \times O}$, then $MCC(E, \succ, e)$ has no justified envy.

10 Concluding Remarks

This paper has studied two mechanisms that can be applied to school choice. We have focused on the case in which each student is endowed with a safe school, to which admission is guaranteed in the event that the student is not admitted to a school she prefers. Such guarantees seem intuitively attractive from the point of view of welfare, and they simplify the student's decision and improve incentives to reveal preferences truthfully. Nevertheless the mechanisms are somewhat more general than the school choice setting, and can be applied in other settings.

Using a novel generalization of Hall's marriage theorem, and an innovative algorithm, we have provided a tractable method of computing the GCPS mechanism of Balbuzanov (2022) for school choice. This mechanism is appropriate when the schools' priorities are dichotomous. The allocation it produces is sd-efficient.

For priorities that are coarse, but not dichotomous, the MCC mechanism computes a profile of fine cutoffs such that for each school, total demand is not greater than the school's quota, and the school does not restrict admission if its total demand is less than its quota. An enhanced MCC mechanism combines the MCC mechanism with an algorithm that passes from the allocation produced by the MCC mechanism to an allocation that is sd-efficient within the set of allocations that do not give any probability of schools to students who are unqualified, in the sense that their priority is less than the school's coarse cutoff.

These mechanisms are not strategy-proof, but intuition strongly suggests that manipulation is difficult, and necessarily entails some risk of receiving probability of admission to less preferred schools. This intuition is confirmed by the formal result that the GCPS mechanism and the MCC mechanism are SP-L. Azevedo and Budish (2019) point to several examples of mechanisms that are not strategy-proof, but are strategy-proof in the large, and which work well in practice.

A software package (described in Online Appendix D) for computing the mechanisms has been developed. Extensive testing has shown that the implementations of the algorithms are reliable, and have acceptable running times, even at the scale of the world's largest school choice problems. In this sense our mechanisms are ready for practical application.

There are many directions for further research. A possibility we hope to explore is to modify the GCPS mechanism so that instead of consuming probability of desirable objects, the agents may discard probability of undesirable objects, which seems appropriate for problems, perhaps such as chore assignment, in which the agents' main concern is to avoid the objects that are most noxious for them.

References

Abdulkadiroğlu, A. and Andersson, T. (2022), School choice. NBER Working Paper 29822.

Abdulkadiroğlu, A., Che, Y.-K. and Yasuda, Y. (2011), 'Resolving conflicting preferences in school choice: The Boston mechanism reconsidered', *Am. Econ. Rev.* **101**, 399–410.

Abdulkadiroğlu, A., Che, Y.-K. and Yasuda, Y. (2015), 'Expanding choice in school choice', *Am. Econ. J-Microecon* **7**, 1–42.

Abdulkadiroğlu, A., Pathak, P. A. and Roth, A. E. (2009), 'Strategy-proofness versus efficiency with indifferences: Redesigning the NYC high school match', *Am. Econ. Rev.* **99**, 1954–1978.

Abdulkadiroğlu, A. and Sönmez, T. (2003), 'School choice: a mechanism design approach', *Am. Econ. Rev.* **93**, 729–747.

Ahuja, R. K., Magnanti, T. L. and Orlin, J. B. (1993), *Network Flows: Theory, Algorithms, and Applications*, Prentice Hall.

Akbarpour, M. and Nikzad, A. (2020), 'Approximate random allocation mechanisms', *Rev. Econ. Stud* **87**, 2473–2510.

Ashlagi, I. and Gonczarowski, Y. A. (2018), 'Stable matching mechanisms are not obviously strategy proof', *J. Econ. Theory* **177**, 405–425.

Ashlagi, I. and Shi, P. (2016), 'Optimal allocation without money: an engineering approach', *Manag. Sci.* **62**, 1078–1097.

Athanassoglou, S. and Sethuraman, J. (2011), 'House allocation with fractional endowments', *Int. J. Game Theory* **40**, 481–513.

Azevedo, E. M. and Budish, E. (2019), ‘Strategy-proofness in the large’, *Rev. Econ. Stud.* **86**, 81–116.

Azevedo, E. M. and Leshno, J. D. (2016), ‘A supply and demand framework for two-sided matching markets’, *J. Polit. Econ.* **124**, 1235–1268.

Balbuzanov, I. (2022), ‘Constrained random matching’, *J. Econ. Theory* **203**. Article 105472.

Bó, I. and Hakimov, R. (2019), ‘Iterative versus standard deferred acceptance: Experimental evidence’, *The Econ. J.* **130**, 356–392.

Bogomolnaia, A. (2015), ‘Random assignment: Redefining the serial rule’, *J. Econ. Theory* **158**, 308–315.

Bogomolnaia, A. and Moulin, H. (2001), ‘A new solution to the random assignment problem’, *J. Econ. Theory* **100**(2), 295–328.

Bogomolnaia, A. and Moulin, H. (2004), ‘Random matching under dichotomous preferences’, *Econometrica* **72**(1), 257–279.

Budish, E., Che, Y.-K., Kojima, F. and Milgrom, P. (2013), ‘Designing random allocation mechanisms: Theory and practice’, *Amer. Econ. Rev.* **103**, 585–623.

Calsamiglia, C., Haeringer, G. and Klijn, F. (2010), ‘Constrained school choice: An experimental study’, *Am. Econ. Rev.* **100**, 1860–1874.

Cerrone, C., Hermstrüwer, Y. and Kesten, O. (2024), ‘School choice with consent: An experiment’, *Econ. J.* **134**, 1760–1805.

Che, Y.-K. and Tercieux, O. (2019), ‘Efficiency and stability in large matching markets’, *J. Polit. Econ.* **127**, 2301–2342.

Chen, L., Kyng, R., Liu, Y., Gutenberg, M. and Sachdeva, S. (2022), ‘Maximum flow and minimum-cost flow in almost-linear time’, *arXiv preprint* .

Cho, W. J. (2016), ‘Incentive properties for ordinal mechanisms’, *Game. Econ. Behav.* **95**, 168–177.

Cho, W. J. (2018), ‘Probabilistic assignment: an extension approach’, *Soc. Choice Welfare* **51**(1), 137–162.

Cho, W. J. and Doğan, B. (2016), ‘Equivalence of efficiency notions for ordinal assignment problems’, *Econ. Lett.* **146**, 8–12.

Dur, U., Gitmez, A. A. and Yilmaz, Ö. (2019), 'School choice under partial fairness', *Theor. Econ.* **14**, 1309–1346.

Echenique, F., Miralles, A. and Zhang, J. (2021), 'Constrained pseudo-market equilibrium', *Am. Econ. Rev.* **111**, 3699–3732.

Ehlers, L., Hafalir, I. E., Yenmez, M. B. and Yildirim, M. A. (2014), 'School choice with controlled choice constraints: Hard bounds versus soft bounds', *J. Econ. Theory* **153**, 648–683.

Ehlers, L. and Morrill, T. (2020), '(il)legal assignments in school choice', *R. Econ. Studies* **87**, 1837–1875.

Erdil, A. and Ergin, H. (2008), 'What's the matter with tie breaking? Improving efficiency in school choice', *Am. Econ. Rev.* **98**, 669–689.

Fearnly, J., Pálvölgvi, D. and Savani, R. (2022), 'A faster algorithm for finding Tarski fixed points', *ACM Trans. Algorithms* **18**, 1–23. Article 23.

Featherstone, C. R. and Niederle, M. (2008), Ex ante efficiency in school choice mechanisms: An experimental investigation. National Bureau of Economic Research (NBER) Working Paper 14618.

Ford, L. and Fulkerson, D. (1956), 'Maximal flow through a network', *Canadian J. Math.* **8**, 399–404.

Gale, D. (1957), 'A theorem on flows in networks', *Pacific J. Math.* **7**, 1073–1082.

Gale, D. and Shapley, L. (1962), 'College admissions and the stability of marriage', *Am. Math. Mon.* **65**, 9–15.

Goldberg, A. V. and Tarjan, R. E. (1988), 'A new approach to the maximum flow problem', *J. ACM* **35**, 921–940.

Grigoryan, A. (2023), 'Top trading cycles with reordering: improving match priority in school choice', *Soc. Choice Welfare* **60**, 383–396.

Haeringer, G. and Klijn, F. (2009), 'Constrained school choice', *J. Econ. Theory* **144**, 1921–1947.

Hafalir, I. E., Yenmez, M. B. and Yildirim, M. A. (2013), 'Effective affirmative action in school choice', *Theor. Econ.* **8**, 325–363.

Hakimov, R. and Kesten, O. (2018), 'The equitable top trading cycles mechanism for school choice', *Int. Econ. Rev.* **59**, 2219–2258.

He, Y., Li, S. and Yan, J. (2015), 'Evaluating assignment without transfers: A market perspective', *Econ. Letters* **133**, 40–44.

He, Y., Miralles, A., Pycia, M. and Yan, J. (2018), 'A pseudo-market approach to allocation with priorities', *Am. Econ. J-Microecon* **10**(3), 272–314.

Hylland, A. and Zeckhauser, R. (1979), 'The efficient allocation of individuals to positions', *J. Polit. Econ.* **87**(2), 293–314.

Kalai, E. (2004), 'Large robust games', *Econometrica* **72**, 1631–1665.

Kamada, Y. and Kojima, F. (2015), 'Efficient matching under distributional constraints: Theory and applications', *Amer. Econ. Rev.* **105**, 67–99.

Kamada, Y. and Kojima, F. (2017), 'Stability concepts in matching under distributional constraints', *J. Econ. Theory* **168**, 107–142.

Karmarkar, N. (1984), A new polynomial-time algorithm for linear programming, *in* 'Proceedings of the 16th ACM Symposium on Theory of Computing', STOC '84, ACM, NewYork, NY, USA, pp. 302–311.

Katta, A.-K. and Sethuraman, J. (2006), 'A solution to the random assignment problem on the full preference domain', *J. Econ. Theory* **131**, 231–250.

Kesten, O. (2006), 'On two competing mechanisms for priority-based allocation problems', *J. Econ. Theory* **127**, 155–171.

Kesten, O. (2010), 'School choice with consent', *Q. J. Econ.* **127**, 1297–1346.

Kesten, O. and Ünver, M. U. (2015), 'A theory of school-choice lotteries', *Theor. Econ.* **10**, 543–595.

Khachian, L. (1979), 'A polynomial algorithm in linear programming', *Soviet Math. Dokl.* **20**, 191–194.

Klee, V. and Minty, G. (1972), How good is the simplex algorithm?, *in* O. Sisha, ed., 'Inequalities III', Academic Press, New York.

Kojima, F. (2009), 'Random assignment of multiple indivisible objects', *Math. Soc. Sci.* **57**, 134–142.

Kojima, F. (2012), 'School choice: Impossibilities for affirmative action', *Games Econ. Behav.* **75**, 685–693.

Kojima, F. and Manea, M. (2010), 'Incentives in the probabilistic serial mechanism', *J. Econ. Theory* **145**, 293–314.

Leshno, J. D. and Lo, I. (2020), The cutoff structure of top trading cycles in school choice. Columbia University Working Paper.

Li, S. (2017), 'Obviously strategy-proof mechanisms', *Amer. Econ. Rev.* **107**, 3257–3287.

Miralles, A. (2009), School choice: The case for the Boston mechanism, *in* 'Auctions, Market Mechanisms, and Their Applications 2009', pp. 58–60.

Morrill, T. (2015), 'Two simple variations of top trading cycles', *Theor. Econ.* **60**, 123–140.

Pathak, P. (2006), Lotteries in school assignment. Harvard University.

Pathak, P. A. (2017), What really matters in designing school choice mechanisms, *in* 'Advances in Economics and Econometrics', Cambridge University Press, Cambridge, pp. 176–214.

Reichmeider, P. F. (1978), The equivalence of some combinatorial matching theorems. Ph.D. thesis, Adelphi University.

Reichmeider, P. F. (1985), *The Equivalence of Some Combinatorial Matching Theorems*, Polygonal Publishing House, Washington.

Reny, P. J. (2022), 'Efficient matching in the school choice problem', *Am. Econ. Rev.* **112**, 2025–2043.

Romm, A., Roth, A. E. and Shorrer, R. I. (2020), Stability vs. no justified envy. draft, Stanford University.

Roth, A. E. and Rothblum, U. G. (1999), 'Truncation strategies in matching markets–in search of advice for participants', *Econometrica* **67**, 21–43.

Saban, D. and Sethuraman, J. (2014), 'A note on object allocation under lexicographic preferences', *J. Math. Econ.* **50**, 283–289.

Schulman, L. J. and Vazirani, V. V. (2015), 'Allocation of divisible goods under lexicographic preferences', *arXiv preprint* .

Shapley, L. S. and Scarf, H. (1974), ‘On cores and indivisibility’, *J. Math. Econ.* **1**, 23–38.

Spielman, D. and Teng, S.-H. (2004), ‘Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time’, *Journal of Association of Computing Machinery* **51**, 385–463.

Tang, Q. and Yu, J. (2014), ‘A new perspective on Kesten’s school with consent idea’, *J. Econ. Theory* **154**, 543–561.

Tang, Q. and Zhang, Y. (2021), ‘Weak stability and pareto efficiency in school choice’, *Econ. Theory* **71**, 533–552.

Troyan, P. (2012), ‘Comparing schoool choice mechanisms by interim and ex-ante welfare’, *Game. Econ. Behav.* **75**, 936–947.

Troyan, P., Delacrétaz, D. and Kloosterman, A. (2020), ‘Essentially stable matchings’, *Game. Econ. Behav.* **120**, 370–390.

Yılmaz, Ö. (2009), ‘Random assignment under weak preferences’, *Game. Econ. Behav.* **66**, 546–558.

Yılmaz, Ö. (2010), ‘The probabilistic serial mechanism with private endowments’, *Game. Econ. Behav.* **69**, 475–491.

A Miscellaneous Proofs

Proof of Theorem 1. Our proof of Theorem 1 applies the theory of network flows. (Ahuja et al. (1993) provides a general introduction and overview.) Let (N, A) be a directed graph (N is a finite set of *nodes* and $A \subset N \times N$ is a set of *arcs*) with distinct distinguished nodes s and t, called the *source* and *sink* respectively. For the sake of simplicity and clarity of intuition (the formal analysis can be more general) we assume that $(n, s), (t, n), (n, n) \notin A$ for all $n \in N$, and that $(n', n) \notin A$ whenever $(n, n') \in A$.

A *flow* is a function $f \colon N \times N \to \mathbb{R}$ such that:

(a) for all n and n', if $(n, n') \notin A$, then $f(n, n') \leq 0$;

(b) for all n and n', $f(n, n') = -f(n', n)$;

(c) $\sum_{n' \in N} f(n, n') = 0$ for all $n \in N \setminus \{s, t\}$.

Clearly (a) and (b) imply that $f(s,n), f(n,t) \geq 0$ for all n, and that $f(n,n') = 0$ if neither (n,n') nor (n',n) is in A. Summing (b) over n and n', then applying (c), gives

$$0 = \sum_{n \in N} \sum_{n' \in N} f(n,n') = \sum_{n' \in N} f(s,n') + \sum_{n' \in N} f(t,n'),$$

so (b) allows us to define the *value* of f to be

$$|f| = \sum_{n \in N} f(s,n) = \sum_{n \in N} f(n,t) \geq 0.$$

A *capacity* is a function $c\colon N \times N \to [0,\infty]$ such that $c(n,n') = 0$ whenever $(n,n') \notin A$. A *cut* is a set $S \subset N$ such that $s \in S$ and $t \in S^c$ where $S^c = N \setminus S$ is the complement. For a capacity c, the *capacity* of S is

$$c(S) = \sum_{(n,n') \in S \times S^c} c(n,n').$$

A flow f is *bounded* by a capacity c if $f(n,n') \leq c(n,n')$ for all (n,n'), and it is a *maximum flow* for c if it is maximal for $|f|$ among the flows bounded by c. When f is bounded by c and S is a cut we have

$$|f| = \sum_{n' \in N} f(s,n') = \sum_{(n,n') \in S \times N} f(n,n') = \sum_{(n,n') \in S \times S^c} f(n,n') \leq c(S).$$

(The first equality is the definition of $|f|$, the second is (c) for $n \in S \setminus \{s\}$, and the third applies (b).) The max-flow min-cut theorem (Ford and Fulkerson, 1956) asserts that for a given c, there is a flow f bounded by c and a cut S such that $|f| = c(S)$.

We consider a particular directed graph (N_E, A_E) in which the set of nodes is $N_E = \{s\} \cup I \cup O \cup \{t\}$ and the set of arcs is

$$A_E = \{\, a_i : i \in I \,\} \cup \{\, a_{io} : i \in I, o \in O \,\} \cup \{\, a_o : o \in O \,\}$$

where, for $i \in I$ and $o \in O$, $a_i = (s,i)$, $a_{io} = (i,o)$, and $a_o = (o,t)$. Let c_E be the capacity in which:

$$c_E(a_i) = r_i, \quad c_E(a_{io}) = g_{io}, \quad c_E(a_o) = q_o.$$

Suppose that S is a cut, and let $J = I \cap S$ and $P = O \cap S$. An arc can go from a node in S to a node in S^c by going from s to J^c, from J to P^c, or from P to t, so

$$c_E(S) = \sum_{i \in J^c} r_i + \sum_{i \in J} \sum_{o \in P^c} g_{io} + \sum_{o \in P} q_o.$$

If f is a flow, there is an allocation p such that $p_{io} = f(a_{io})$ for all i and o. Conversely, if p is an allocation, the unique flow f_p such that $f_p(a_{io}) = p_{io}$ for all i and o has $f_p(a_i) = \sum_o p_{io}$ for all i and $\sum_i p_{io} = f_p(a_o)$ for all o. Evidently p is a partial allocation if and only if f_p is bounded by c_E, and it is a feasible allocation if and only if, in addition, $f_p(a_i) = r_i$ for all i, which is the case if and only if $|f_p| = \sum_i r_i$.

Thus there is a feasible allocation if and only if the maximum value of a flow bounded by c_E is $\sum_i r_i$, and the max flow-min cut theorem implies that this is the case if and only if the minimum capacity of a cut for c_E is $\sum_i r_i$. Since $c_E(\{s\}) = \sum_i r_i$, this is the case if and only if, for every cut S, $c_E(S) \geq \sum_i r_i$. For every $J \subset I$ and $P \subset O$, $\{s\} \cup J \cup P$ is a cut, and every cut has this form, so there is a feasible allocation if and only if $\sum_i r_i \leq c_E(\{s\} \cup J \cup P)$ for all $J \subset I$ and $P \subset O$. Subtracting $\sum_{i \in J^c} r_i$ from both sides reveals that this is the GMC inequality for J and P. □

Proof of Proposition 2. Let m be a feasible allocation for E. For each $o \in P \cup P'$ we have either $\sum_{i \in J} m_{io} = q_o$ and $m_{io} = 0$ for all $i \in J^c$ or $\sum_{i \in J'} m_{io} = q_o$ and $m_{io} = 0$ for all $i \in J'^c$, so $\sum_{i \in J \cup J'} m_{io} = q_o$. For each $i \in J \cup J'$ and $o \in (P \cup P')^c = P^c \cap P'^c$ we have $m_{io} = g_{io}$, either because $i \in J$ and $o \in P^c$ or because $i \in J'$ and $o \in P'^c$. In view of Lemma 1, $(J \cup J', P \cup P')$ is critical.

For each $o \in P \cap P'$ we have $\sum_{i \in J} m_{io} = q_o$ and $m_{io} = 0$ for all $i \in J^c$ and $\sum_{i \in J'} m_{io} = q_o$ and $m_{io} = 0$ for all $i \in J'^c$, so $\sum_{i \in J \cap J'} m_{io} = q_o$. For each $i \in J \cap J'$ and $o \in (P \cap P')^c = P^c \cup P'^c$ we have $m_{io} = g_{io}$, either because $i \in J$ and $o \in P^c$ or because $i \in J'$ and $o \in P'^c$. Again Lemma 1 implies that $(J \cap J', P \cap P')$ is critical. □

Proof of Proposition 4. Suppose that $i, j \in I$, $\alpha_i \subset \alpha_j$, and $o_i \succ_i o_j$ for all $o_i \in \alpha_i$ and $o_j \in \alpha_j \setminus \alpha_i$. At each time t during the allocation process there are sets of schools $P_i(t)$ and $P_j(t)$ such that for times slightly greater than t, i is required to consume from $P_i(t)$ and j is required to consume from $P_j(t)$, and either $P_i(t) = P_j(t)$ or $P_i(t) \cap P_j(t) = \emptyset$. If $P_i(t) = P_j(t)$, then i weakly prefers her favorite element of $\alpha_i \cap P_i(t)$ to every element of $\alpha_j \cap P_i(t)$. If, at some time t, we start to have $P_i(t) \cap P_j(t) = \emptyset$, it must be because $P_i(t)$ is part of a critical pair, with $\alpha_i \subset P_i(t)$ and $\alpha_j \setminus P_i(t) \neq \emptyset$. From this time going forward i will be consuming an element of α_i and j will be consuming an element of $\alpha_j \setminus \alpha_i$. Thus at every time i is consuming a school that she weakly prefers to the school that j is consuming, so $GCPS_i(E, \succ) \, sd(\succ_i) \, GCPS_j(E, \succ)$. □

B Eating Function Analysis

In this Appendix we prove Theorem 4. This proof is based on a detailed analysis of the consequences of manipulation in terms of its effect on the continuous time eating process of BM, as generalized by Kojima and Manea (2010) and here.

We fix a school choice CEE $E = (I, O, 1, q, g)$ that satisfies the GMC and a profile $\succ = (\succ_j)_{j\in I}$ of strict preferences over O. Fixing $i \in I$, let o^* be an element of $O \setminus \alpha_i$, let $\alpha'_i = \alpha_i \cup \{o^*\}$, and let $\succ'_i$ be a preference over O that has α'_i as the set of schools weakly preferred to i's safe school, and that agrees with $\succ_i$ on α_i. We wish to show that the augmentation manipulation of reporting $\succ'_i$ rather than $\succ_i$ results in an allocation for i that is weakly $sd(\succ_i)$ worse. For $\rho \in [0, 1]$ let $E^\rho = (I, O, 1, q, g^\rho)$ where $g^\rho_{io^*} = \rho$ and $g^\rho_{jo} = g_{jo}$ for all $(j, o) \neq (i, o^*)$. Fixing $\tilde{o} \in \alpha_i$, we will show that i's total consumption of schools that are $\succ_i$-weakly preferred to $\tilde{o}$ is weakly decreasing as ρ increases, or, equivalently, that the consumption of schools that are $\succ'_i$-weakly preferred to $\tilde{o}$ does not increase more rapidly than i's consumption of o^* increases.

For a general (not necessarily school choice) CEE E, $j \in I$, and $t \in [0, 1]$, an *eating schedule* on $[0, t)$ is a function $e_j \colon [0, t) \to O$ that is piecewise constant (i.e., changes objects finitely many times) and right continuous: for any $t' \in [0, t)$ there is an $\varepsilon > 0$ such that $e_j(t'') = e_j(t')$ for all $t'' \in [t', t' + \varepsilon)$. For such an e_j, $o \in O$, and $t' \in [0, t]$ let

$$p_{jo}(e_j, t') = \int_0^{t'} \mathbf{1}_{e_j(s)=o}\, ds,$$

and let $\tau_{jo}(e_j) = \sup\{\, t' : p_{jo}(e_j, t') < g_{jo} \,\}$.

An *eating function* on $[0, t)$ is a vector $e = (e_j)_{j\in I}$ of eating schedules on $[0, t)$. For $t' \in [0, t]$ let $p(e, t') \in \mathbb{R}^{I\times O}_+$ be the allocation with components $p_{jo}(e_j, t')$. For $J \subset I$, $P \subset O$, and $t' \in [0, t]$ let

$$s_{(J,P)}(e, t') = \sum_{o\in P} q_o + \sum_{i\in J} \sum_{o\in P^c} g_{io} - \sum_{i\in J} r_i - \sum_{i\in J^c} \sum_{o\in P} p_{io}(e, t'),$$

and let $\tau_{(J,P)}(e) = \sup\{\, t' : s_{(J,P)}(e, t') > 0 \,\}$. For $j \in I$ and $t' \in [0, t]$ let

$$\alpha_j(e, t') = \alpha_j \setminus \Big(\{\, o : p_{jo}(e_j, t') \geq g_{io} \,\} \cup \bigcup_{J\subset I, P\subset O \,:\, s_{(J,P)}(e,t') \leq 0 \text{ and } j \in J^c} P \Big)$$

be the set of objects that are still available to j at time t'. Note that $\alpha_j(e, \cdot)$ is right continuous. Let $e^{\succ}_j(e, t')$ be the $\succ_j$-best element of $\alpha_j(e, t)$. We say that e_j is *myopic* for e if, for all $t' \in [0, t)$, $e_j(t') = e^{\succ}_j(e, t')$.

Lemma 8. If E satisfies the GMC, then for each $t \in (0,1]$ there is a unique eating function e on $[0,t)$ such that each e_j is myopic for e on $[0,1)$.

Proof. For sufficiently small $\varepsilon > 0$, if, for each j, e_j is the constant function on $[0,\varepsilon)$ with value $e_j^{\succ}(0)$, and $e = (e_j)_j$, then for all $t' \in [0,\varepsilon)$, each e_j is myopic for e. Therefore there is a $\bar{t} \in (0,t]$ and a vector of eating schedules e on $[0,\bar{t})$ such that each e_j is myopic for e.

Suppose that e' is also a vector of eating schedules on $[0,\bar{t})$ such that each e'_j is myopic for e'. For each j we have $e_j(0) = e_j^{\succ}(e,0) = e_j^{\succ}(e',0) = e'_j(0)$, so e_j and e'_j agree on the degenerate interval $[0,0]$. If $\hat{t} \in [0,\bar{t})$ and each e_j and e'_j agree on $[0,\hat{t}]$, then $\alpha_j(e,\hat{t}) = \alpha_j(e',\hat{t})$ for all $j \in I$, so for some $\varepsilon > 0$, each e_j and e'_j agree on $[0,\hat{t}+\varepsilon)$. Therefore, for the given $\bar{t}$, the vector e is unique.

If $\bar{t} < 1$, then, since E satisfies the GMC, $p(e,\bar{t})$ is a possible allocation, and thus each $\alpha_j(e,\bar{t})$ is nonempty, so for some $\varepsilon > 0$ we can extend e to $[0,\bar{t}+\varepsilon)$ by setting $e_j(t') = e_j^{\succ_j}(e,\bar{t})$ for all $t' \in [\bar{t},\bar{t}+\varepsilon)$, and each extended e_j will be myopic for the extended e. It follows that there is a unique maximal $\bar{t}$, which must be t. □

For $\rho \in [0,1]$ let e^{ρ} be the eating function on $[0,1)$ given by Lemma 8 for E^{ρ} and $\succ'$. The quantities $p_{jo}(e^{\rho},t)$ are piecewise linear functions of (ρ,t), and each $\tau_{io}(e^{\rho})$ and $\tau_{(J,P)}(e^{\rho})$ are piecewise linear functions of ρ. (If, for some $\bar{t}$, these conditions hold for $(\rho,t) \in [0,1] \times [0,\bar{t}]$, then they also hold on $[0,\bar{t}+\varepsilon]$ for some $\varepsilon > 0$, so they hold everywhere.)

For $\rho_0 \in (0,1)$ and $\varepsilon \in (0,\min\{\rho_0, 1-\rho_0\})$ let

$$\mathcal{I}(\rho_0,\varepsilon) = (\rho_0 - \varepsilon, \rho_0 + \varepsilon) \quad \text{and} \quad \mathcal{J}(\rho_0,\varepsilon) = [\rho_0, \rho_0 + \varepsilon).$$

We say that $\rho_0 \in (0,1)$ is *semigeneric* if, for sufficiently small $\varepsilon > 0$, there are affine functions

$$t_0, t_1, \ldots, t_K \colon \mathcal{J}(\rho_0,\varepsilon) \to [0,1]$$

such that $0 \equiv t_0 < t_1 < \cdots < t_K \equiv 1$ and for each j and k there is an $o_{jk} \in O$ such that $e_j^{\rho}(t) = o_{jk}$ for all $\rho \in \mathcal{J}(\rho_0,\varepsilon)$ and $t \in [t_{k-1}(\rho), t_k(\rho))$. We say that ρ_0 is *generic* if, for sufficiently small $\varepsilon > 0$, there are such affine functions on the domain $\mathcal{I}(\rho_0,\varepsilon)$. Basic properties of piecewise linear functions imply that there are finitely many elements of $[0,1]$ that are semigeneric but not generic, and that all other points are generic.

We now fix a semigeneric ρ_0, ε, and $t_0,\ldots,t_K$ as above. We assume that this collection is minimal in the sense that for each $k = 1,\ldots,K-1$ there is some j such that $o_{jk} \neq o_{j,k+1}$. For each $k = 1, \ldots, K$ let

$$\mathcal{P}_k = \{ (J, P) : t_k(\rho) = \tau_{(J,P)}(e^\rho) \text{ for all } \rho \in \mathcal{I}(\rho_0, \varepsilon) \}.$$

For $(J, P) \in \mathcal{P}_k$ let $L_{k,(J,P)} = \{ j \in J^c : o_{jk} \in P \}$.

Our claim will follow if i's total consumption of goods $\succ'$-weakly preferred to $\tilde{o}$ does not increase more rapidly than i's consumption of o^* on $\mathcal{J}(\rho_0, \varepsilon)$. Note that when $\tilde{o} \succ'_i o^*$ the claim holds because Lemma 8 implies that the eating function up to the time when i stops eating $\tilde{o}$ does not depend on ρ. Henceforth we assume that $o^* \succ'_i \tilde{o}$. Similarly, the claim is immediate if i's consumption of o^* does not vary as ρ varies in $\mathcal{J}(\rho_0, \varepsilon)$. Therefore we may assume that there is some $k^* < K$ such that $\tau_{io^*}(e^\rho_i) = t_{k^*}(\rho)$ for $\rho \in \mathcal{J}(\rho_0, \varepsilon)$.

Let o^{**} be the object that agent i starts consuming immediately after ceasing consumption of o^*. Consumption of schools weakly preferred to $\tilde{o}$ may continue until time 1, and our claim also follows easily in this case. Since E is a school choice CEE, consumption of o^{**} and $\tilde{o}$ by i can only end before time 1 if there are integers $k^{**} \leq \tilde{k} < K$ and pairs $(J^{**}, P^{**}) \in \mathcal{P}_{k^{**}}$ and $(\tilde{J}, \tilde{P}) \in \mathcal{P}_{\tilde{k}}$ such that $o^{**} \in P^{**}$, $\tilde{o} \in \tilde{P}$, and $i \in L_{k^{**},(J^{**},P^{**})} \cap L_{\tilde{k},(\tilde{J},\tilde{P})}$. Henceforth we assume this, and that k^{**} and $\tilde{k}$ are the smallest such integers.

For each k let σ_k be the number such that $t_k(\rho) = t_k(\rho_0) + \sigma_k(\rho - \rho_0)$ for all $\rho \in \mathcal{J}(\rho_0, \varepsilon)$. Evidently $\sigma_{k^*} = 1$, and our goal is to show that $\sigma_{\tilde{k}} \leq 1$. In fact we will show that $\sigma_k \leq 1$ for all k.

For each j and $k = 1, \ldots, K$ there is a number $\kappa_{j,k-1}$ such that

$$p_{jo_{jk}}(e^\rho, t) = p_{jo_{jk}}(e^{\rho_0}, t_{k-1}(\rho_0)) + \kappa_{j,k-1}(\rho - \rho_0) + t - t_{k-1}(\rho)$$

for all $\rho \in \mathcal{J}(\rho_0, \varepsilon)$ and $t \in [t_{k-1}(\rho), t_k(\rho))$. Clearly $\kappa_{jk} = \kappa_{j,k-1}$ when $o_{j,k+1} = o_{jk}$ and $\kappa_{jk} = -\sigma_k$ when $o_{j,k+1} \neq o_{jk}$. For $j \in L_{k,(J,P)}$ and $\rho \in \mathcal{J}(\rho_0, \varepsilon)$ we have

$$\begin{aligned} p_{jo_{jk}}(e^\rho, t_k(\rho)) &= p_{jo_{jk}}(e^{\rho_0}, t_k(\rho_0)) + (p_{jo_{jk}}(e^\rho, t_k(\rho)) - p_{jo_{jk}}(e^\rho, t_k(\rho^0))) \\ &\quad + (p_{jo_{jk}}(e^\rho, t_k(\rho^0)) - p_{jo_{jk}}(e^{\rho_0}, t_k(\rho^0))) \\ &= p_{jo_{jk}}(e^{\rho_0}, t_k(\rho_0)) + (\sigma_k + \kappa_{j,k-1})(\rho - \rho_0). \end{aligned}$$

Lemma 9. If $(J, P) \in \mathcal{P}_k$, then

$$\sum_{j \in L_{k,(J,P)}} \kappa_{j,k-1} = -|L_{k,(J,P)}|\sigma_k.$$

Proof. In view of the equation above, the claim follows from the fact that the quantity

$$\sum_{j \in L_{k,(J,P)}} \sum_{o \in P} p_{jo}(e^\rho, t_k(\rho)) = C + (\sum_{j \in L_{k,(J,P)}} \kappa_{j,k-1} + |L_{k,(J,P)}|\sigma_k) \cdot (\rho - \rho_0)$$

does not depend on ρ, where $C = \sum_{j \in L_{k,(J,P)}} \sum_{o \in P} p_{jo}(e^{\rho_0}, t_k(\rho_0))$. □

Lemma 10. If $k \neq k^*, k^{**}$, then $\sum_j \kappa_{jk} = \sum_j \kappa_{j,k-1}$, and $\sum_j \kappa_{jk^{**}} = 1 + \sum_j \kappa_{j,k^{**}-1}$.

Proof. If $k \neq k^*, k^{**}$ and (J, P) is a minimal element of $\mathcal{P}_k$, then $\kappa_{jk} = \sigma_k$ for all $j \in L_{k,(J,P)}$, so

$$\sum_{j \in L_{k,(J,P)}} \kappa_{jk} = -|L_{k,(J,P)}|\sigma_k = \sum_{j \in L_{k,(J,P)}} \kappa_{j,k-1}.$$

Since the set of critical pairs is a lattice (Proposition 2) the set of $L_{k,(J,P)}$ such that (J, P) is a minimal element of $\mathcal{P}_k$ is a partition of $\bigcup_{(J,P) \in \mathcal{P}_k} L_{k,(J,P)}$. For j outside this union we have $\kappa_{jk} = \kappa_{j,k-1}$ because $o_{jk} = o_{j,k-1}$. Therefore summing the equation above gives the first claim.

If (J, P) is a minimal element of $\mathcal{P}_{k^{**}}$ such that $i \in J^c$ and $o^{**} \in P$, then $\kappa_{jk^{**}} = \sigma_{k^{**}}$ for all $j \in L_{k^{**},(J,P)} \setminus \{i\}$, and $\kappa_{ik^*} + \kappa_{ik^{**}} = \sigma_{k^{**}}$ and $\kappa_{ik^*} = -1$, so

$$\sum_{j \in L_{k^{**},(J,P)}} \kappa_{j,k^{**}-1} = -|L_{k^{**},(J,P)}|\sigma_{k^{**}} = -1 + \sum_{j \in L_{k^{**},(J,P)}} \kappa_{j,k^{**}}.$$

Now summing as above gives the second claim. □

Lemma 11. If $0 \leq k < k^*$ then $\sigma_k = 0$ and $\kappa_{jk} = 0$ for all j. We have $\sigma_{k^*} = 1$, $\kappa_{ik^*} = -1$, and $\kappa_{jk^*} = 0$ for all $j \neq i$. If $k^* < k < k^{**}$, then $\kappa_{jk} \leq 0$ for all j, and $\sum_j \kappa_{jk} = -1$. If $k^{**} \leq k$, then $\sum_j \kappa_{jk} = 0$, and $\sum_{j \in I_k^-} \kappa_{jk} \geq -1$ and $\sum_{j \in I_k^+} \kappa_{jk} \leq 1$, where $I_k^- = \{ j : \kappa_{jk} < 0 \}$ and $I_k^+ = \{ j : \kappa_{jk} > 0 \}$.

Proof. The claims for $k < k^*$ follow from the fact that e^ρ does not depend on t for $t < t_{k^*}(\rho)$. The definitions give $\sigma_{k^*} = 1$, $\kappa_{ik^*} = -1$, and $\kappa_{jk^*} = 0$ for all $j \neq i$. From the last result and induction, $\sum_j \kappa_{jk} = -1$ if $k^* \leq k < k^{**}$ and $\sum_j \kappa_{jk} = 0$ if $k^{**} \leq k$.

If $(j, k) \neq (i, k^{**})$, then either $\kappa_{jk} = \kappa_{j,k-1}$ (if $o_{jk} = o_{j,k-1}$) or

$$\kappa_{jk} = -\sigma_k = \frac{\sum_{j' \in L_{k,(J,P)}} \kappa_{j',k-1}}{|L_{k,(J,P)}|}$$

is an average, where (J, P) is an element of $\mathcal{P}_k$ such that $j \in L_{k,(J,P)}$. For $k < k^{**}$, $\kappa_{jk} \leq 0$ follows by induction. This equation with k^{**} in place of k also holds when

$(J, P) \in \mathcal{P}_{k^{**}}$ and $j \in L_{k,(J,P)}$, so $\kappa_{jk^{**}} \leq 0$ holds for $j \neq i$. Since $\sum_j \kappa_{jk^{**}} = 0$, it follows that $\sum_{j \in I^-_{k^{**}}} \kappa_{jk^{**}} \geq -1$ and $\sum_{j \in I^+_{k^{**}}} \kappa_{jk^{**}} \leq 1$. When $k > k^{**}$ this averaging cannot increase $-\sum_{j \in I^-_k} \kappa_{jk}$ or $\sum_{j \in I^+_k} \kappa_{jk}$, so induction implies that these quantities are not greater than 1. □

The last result implies that $\kappa_{jk} \geq -1$ for all j and k, so Lemma 10 implies that $\sigma_k \leq 1$ for all k. As we noted previously, this implies the desired result.

For Online Publication

C Implementation

In this Appendix we consider the problem of passing from a matrix of assignment probabilities to a random deterministic assignment whose distribution realizes the given probabilities, showing that this is possible, and describing an algorithm for this task.

Let $E = (I, O, 1, q, g)$ be an integral school choice CEE that satisfies the GMC, and let Q be its set of feasible allocations. BCKM say that $m \in Q$ is *implementable* if its assignment probabilities are those resulting from some probability distribution over deterministic assignments[12]. Recalling that the vertices of Q are its extreme points, we see that every element of Q is implementable if and only if each of its vertices is a deterministic assignment, which is to say that its entries are elements of $\{0, 1\}$.

As we explain in detail below, Theorem 1 of BCKM has the following result as a special case, and in turn this result has the Birkhoff-von Neumann theorem as a special case.

Theorem 5. Each vertex of Q is integral.

We quickly review the related concepts and results of BCKM. A *constraint set* is a nonempty subset of $I \times O$, and a *constraint structure* $\mathcal{H}$ is a set of constraint sets. A vector of quotas $\mathbf{q} = (q_S, q^S)_{S \in \mathcal{H}}$ is *integral* if $q_S, q^S \in \mathbb{Z}$ for all S. An allocation m is *feasible* under $\mathbf{q}$ if $q_S \le \sum_{io \in S} m_{io} \le q^S$ for all $S \in \mathcal{H}$. Let $\mathcal{M}_{\mathbf{q}}$ be the set of feasible allocations for $\mathbf{q}$. If $\mathcal{H}$ contains all singletons, then $\mathcal{M}_{\mathbf{q}}$ is bounded, hence a polytope.

The constraint structure $\mathcal{H}$ is *universally implementable* if, whenever $\mathbf{q}$ is integral, each vertex of $\mathcal{M}_{\mathbf{q}}$ is integral. A constraint structure is a *hierarchy* if, for all $S, S' \in \mathcal{H}$, we have $S \subset S'$ or $S' \subset S$ or $S \cap S' = \emptyset$, and $\mathcal{H}$ is a *bihierarchy* if there are hierarchies $\mathcal{H}_1$ and $\mathcal{H}_2$ such that $\mathcal{H}_1 \cup \mathcal{H}_2 = \mathcal{H}$ and $\mathcal{H}_1 \cap \mathcal{H}_2 = \emptyset$. Theorem 1 of BCKM asserts that if $\mathcal{H}$ is a bihierarchy, then it is universally implementable.

Let $\mathcal{H}^1 = \{\{i\} \times O : i \in I\}$, $\mathcal{H}^2 = \{\{(i, o)\} : (i, o) \in I \times O\}$, and $\mathcal{H}^3 = \{I \times \{o\} : o \in O\}$, corresponding to the constraints $\sum_o m_{io} = 1$, $0 \le m_{io} \le g_{io}$, and $\sum_i m_{io} \le q_o$ respectively. We can show that $\mathcal{H} = \mathcal{H}^1 \cup \mathcal{H}^2 \cup \mathcal{H}^3$ is a bihierarchy either by setting $\mathcal{H}_1 = \mathcal{H}^1 \cup \mathcal{H}^2$ and $\mathcal{H}_2 = \mathcal{H}^3$ or by setting $\mathcal{H}_1 = \mathcal{H}^1$ and $\mathcal{H}_2 = \mathcal{H}^2 \cup \mathcal{H}^3$, so our Theorem 5 follows from their Theorem 1.

The practical implementation of a random allocation depends not only on the existence of a representation of it as a convex combination of pure allocations, but also on

[12] Recently Akbarpour and Nikzad (2020) expanded the scope of this concept by studying a notion of approximate implementation that is appropriate when some constraints need not be satisfied exactly.

an efficient algorithm for generating a random pure allocation with a probability distribution that averages to the given allocation. To this end we describe the argument in Appendix B of the Online Appendices of Budish et al., which they attribute to Tomomi Matsui and Akihisa Tamura, as it applies to our setting.

We work with the directed graph (N_E, A_E) defined in the proof of Theorem 1 in Appendix A. Recall that the set of nodes is $N_E = \{s\} \cup I \cup O \cup \{t\}$, where s and t are artificial nodes called the source and sink. The set of arcs is

$$A_E = \{\, a_i : i \in I \,\} \cup \{\, a_{io} : i \in I, o \in O \,\} \cup \{\, a_o : o \in O \,\}$$

where, for $i \in I$ and $o \in O$, $a_i = (s, i)$, $a_{io} = (i, o)$, and $a_o = (o, t)$. If $m \in Q$, let $C(m) = \{\, a_{io} : m_{io} \notin \mathbb{Z} \,\}$ and $D(m) = \{\, a_o : \sum_i m_{io} \notin \mathbb{Z} \,\}$. The *nonintegrality set* of m is $B(m) = C(m) \cup D(m) \subset A_E$.

Recall that the *floor* of a real number x is the largest integer that is not greater than x, and the *ceiling* of x is the smallest integer that is not less than x. When x is an integer, it is both the floor and ceiling of itself.

Proposition 6. If the nonintegrality set of $m \in Q$ is nonempty, then there are $m^0, m^1 \in Q \setminus \{m\}$ such that m is a convex combination of m^0 and m^1, and for both $h = 0, 1$:

(a) For each i and o, m^h_{io} is between the floor and the ceiling of m_{io}.

(b) For each o, $\sum_i m^h_{io}$ is between the floor and the ceiling of $\sum_i m_{io}$.

(c) The nonintegrality set of m^h is a proper subset of the nonintegrality set of m.

Proof. An *allowed path* is a sequence $n_1, \ldots, n_h$ of distinct nodes in $I \cup O \cup \{t\}$ such that $h > 2$, for all $g = 1, \ldots, h-1$ either $(n_g, n_{g+1}) \in B(m)$ or $(n_{g+1}, n_g) \in B(m)$, and either $(n_h, n_1) \in B(m)$ or $(n_1, n_h) \in B(m)$. Given such a cycle, for each i and o, if $a_{io} = (n_g, n_{g+1})$ ($a_{io} = (n_{g+1}, n_g)$) for some g, then we say that a_{io} is a *forward* (*backward*) *arc*. For $\gamma \in \mathbb{R}$ let $m^\gamma \in \mathbb{R}^{I \times O}$ be the matrix with components

$$m^\gamma_{io} = \begin{cases} m_{io} + \gamma, & a_{io} \text{ is a forward arc,} \\ m_{io} - \gamma, & a_{io} \text{ is a backward arc,} \\ m_{io}, & \text{otherwise.} \end{cases}$$

Let α be the smallest positive number such that one of the following occurs:

(a) $m^\alpha_{io} \in \mathbb{Z}$ for some $a_{io} \in C(m)$.

(b) $\sum_i m^\alpha_{io} \in \mathbb{Z}$ for some $a_o \in D(m)$.

Let β be the smallest positive number such that $m^{-\beta}$ satisfies one of these conditions. Let $m^0 = m^\alpha$ and $m^1 = m^{-\beta}$, so that $m = \frac{\beta}{\alpha+\beta}m^0 + \frac{\alpha}{\alpha+\beta}m^1$.

For each i and g such that $n_g = i$, (i, n_{g-1}) is a backward arc and (i, n_{g+1}) is a forward arc, so $\sum_o m^\gamma_{io} = \sum_o m_{io} = r_i$ for all γ. Since E is integral, it follows that m^0 and m^1 satisfy all the constraints defining Q. It is now easy to see that m^0 and m^1 satisfy (a)–(c) of the statement.

The remainder of the proof describes an algorithm for constructing an allowed cycle. An *allowed path* is a sequence $n_1, \ldots, n_h$ of distinct nodes in $I \cup O \cup \{t\}$ such that for all $g = 1, \ldots, h-1$ either $(n_g, n_{g+1}) \in B(m)$ or $(n_{g+1}, n_g) \in B(m)$. By hypothesis there are n_1 and n_2 such that $(n_1, n_2) \in B(m)$. Therefore we may suppose that an allowed path $n_1, \ldots, n_g$ has already been constructed.

To show that there is some $n_{g+1} \neq n_{g-1}$ such that either $(n_g, n_{g+1}) \in B(m)$ or $(n_{g+1}, n_g) \in B(m)$, we enumerate cases: a) if $(n_g, n_{g-1}) \in C(m)$, then (since $\sum_o m_{io} = 1$) there is an $o \in O \setminus \{n_{g-1}\}$ such that $(n_g, o) \in C(m)$; b) if $(n_{g-1}, n_g) \in C(m)$, then either there is an $i \in I \setminus \{n_{g-1}\}$ such that $(i, n_g) \in C(m)$, or $(n_g, t) \in D(m)$; c) if $(n_g, n_{g-1}) \in D(m)$, then there is an i such that $(i, n_g) \in C(m)$; d) f $(n_{g-1}, n_g) \in D(m)$, then (since $\sum_i \sum_o m_{io} = |I|$) there is an $o \in O \setminus \{n_{g-1}||$ such that $(n_g, o) \in C(m)$. Since N_E is finite, continuing the construction in this fashion leads eventually to $n_{g+1} \in \{n_1, \ldots, n_{g-2}\}$, so this process eventually constructs an allowed cycle. □

To generate a random integral allocation whose expectation is the given m we repeatedly execute the computation described in this argument, passing to m^0 with probability $\frac{\beta}{\alpha+\beta}$ and passing to m^1 with probability $\frac{\alpha}{\alpha+\beta}$. It is easy to show that this is a polynomial time algorithm, but in fact it has been implemented (Online Appendix D) and in practice its running time is insignificant..

D GCPS MCC Schools

For the application to school choice, versions of the algorithm described in Sections 5–7 and Online Appendix C have been encoded, using the `C` programming language, in the software package `GCPS MCC Schools`, which can be downloaded[13]. This package provides the six executables `makex`, `gcps`, `lpgcps`, `mcc`, `emcc`, and `purify`.

The executable `makex` constructs random examples of the sort of problem that might occur in a large school district. The schools and students are spaced evenly around a circle. Each student's safe school is the school that is closest to her home.

[13] Open the url `https://github.com/Coup3z-pixel/SchoolOfChoice/` in a web browser. Detailed instructions for download and installation are given in *GCPS_Schools_User_Guide.pdf*.

Each school has a random *valence*, which is normally distributed, for each student-school pair there is a normally distributed *idiosyncratic shock*, and the student's utility for a seat in the school is the sum of these two quantities minus the distance between her home and the school. The schools that the student is eligible for are those that provide at least as much utility as the safe school, and the student's preference over such schools is the one induced by these utilities.

Each student has a normally distributed *test score*. Each school's *raw priority* for a student is her test score minus the distance from her home to the school. There is a positive integer number of priority classes. The students receiving top priority at a school are those students for whom the school is their safe school. The other students are divided, as evenly as possible, into the remaining priority classes, with students with higher raw priorities at the school receiving higher priority. The students' preferences over eligible schools and the schools' priorities constitute a *school choice problem*.

The executables `gcps`, `mcc`, and `emcc` each take a school choice problem as input and output a feasible allocation. Both `gcps` and `emcc` have been extensively tested, on sample problems generated by `makex` with up to 500 schools and up to 100,000 students. They are reliable, with running times that vary only slightly with the particular problem, and which scale with roughly the square of the number of students, with `emcc` consistently taking roughly one third of the time consumed by `gcps`. (For a given number of students, `gcps` is somewhat slower if there are more schools and fewer students per school.) The running times are acceptable for practical application: for example, for a problem with 500 schools and 99,500 students, the running time of `gcps` on a MacBook Pro is 23 hours and 47 minutes, and the running time of `emcc` is 4 hours and 32 minutes.

Using the algorithm described by BCKM and in Online Appendix C, the executable `purify` takes a feasible allocation as an input and outputs a random deterministic assignment whose distribution averages to the given allocation. At the scale of problems considered here its running time is insignificant. The computations passing from the `mcc` allocation to the `emcc` allocation have a similar character, so there is no significant difference between the running times of `mcc` and `emcc`.

Overall, we can conclude that there are no computational barriers impairing `gcps` and `emcc`, so these algorithms are ready for initial practical application.