

2. ARTÍCULOS DE INVESTIGACIÓN OPERATIVA

MECHANISM DESIGN IN SCHOOL CHOICE: SOME LESSONS IN A NUTSHELL

Flip Klijn*

Institute for Economic Analysis (CSIC)
Campus UAB, Bellaterra (Barcelona)

Abstract

This paper deals with school choice as an application of matching theory. Although the use of matching theory for the design and study of school choice mechanisms is rather recent, some of its tools were already successfully employed in several other markets, the most noticeable being the labor market for medical doctors in the US. I first briefly describe the problems that some US school districts had, and why and how economic engineering has contributed a lot to the improvement of school choice programs. Then, I will review and interpret a selection of the most recent developments and results.

Keywords: school choice, matching, stability, efficiency, strategy-proofness, preference revelation, indifferences

1. Introduction

A common feature of many markets is their bilateral structure and the need to match agents from one side of the market to the other side of the market. An important instance is the assignment of students to colleges, or workers to firms.

When David Gale and Lloyd Shapley published their elegant paper back in 1962 they probably did not imagine the stream of literature that would follow.¹ They described a model in which students have to be matched to colleges, and where each student and each college has preferences over the other side of the market. They proposed an algorithm that always produces a matching (assignment) that is stable in the following sense: each agent obtains an acceptable mate, and no pair of agents who are not matched to one another would prefer to be. Their algorithm is referred to as the

“deferred acceptance algorithm,” the reason being that (loosely speaking) its two key elements are the “proposals” from either side of the market and the “deferred acceptance” by the other side of the market.

Gale and Shapley’s deferred acceptance algorithm has turned out to be key in the design of many markets. In fact, even before Gale and Shapley’s paper appeared in the *American Mathematical Monthly* very similar ideas were already incorporated in the design of the first centralized market for medical doctors in the US. More recently, variants of the deferred acceptance algorithms were implemented after the redesign of school choice systems in Boston and New York.² Every now and then variants of the deferred acceptance algorithm are independently rediscovered and implemented to solve market failures.³

In their pioneering paper, Abdulkadiroğlu and

*Corresponding Author. E-mail: flip.klijn@iae.csic.es. I thank Bettina Klaus and Ana Meca for their comments on an earlier version of this paper. Any remaining errors are mine. Financial support through *Plan Nacional I+D+I* (SEJ2005-01690), the *Consolider-Ingenio* 2010 (CSD2006-00016) program, and the Barcelona Economics Program of XREA is gratefully acknowledged.

¹For an overview until 1989, see Gusfield and Irving (1989) and Roth and Sotomayor (1990). Al Roth’s game theory, experimental economics, and market design page has an updated bibliography.

²See Abdulkadiroğlu, Pathak, and Roth (2005) and Abdulkadiroğlu, Pathak, Roth, and Sönmez (2005) for reports on the redesign of the New York and Boston school matches, respectively.

³To mention a few: medical interns in the US and the UK (Roth, 1990), secondary schools in Singapore (Teo et al., 2001), and higher education in Hungary (Biró, 2008), Spain (Romero-Medina, 1998), and Turkey (Balinski and Sönmez, 1999). See also Roth’s (2008a) review on the theory and practice of the deferred acceptance algorithm.

Sönmez (2003) discussed some of the problems that several US school districts were experiencing and proposed two student assignment mechanisms. One mechanism they proposed is directly based on Gale and Shapley's deferred acceptance algorithm. The other mechanism is based on Gale's Top Trading Cycles algorithm, which is first described by Shapley and Scarf (1974) in the context of so-called housing markets. Each of the two mechanisms has a desirable feature that the other mechanism fails to satisfy. The Gale-Shapley mechanism is stable, but not (Pareto-)efficient, i.e., there are situations in which the mechanism yields a matching that can be improved upon for some students without hurting the others. The Top Trading Cycles mechanism on the other hand is efficient, but not stable.

A first question of course is why certain school choice districts were experiencing problems until very recently. This is what I discuss in Section 2. In Section 3, I recall Abdulkadiroğlu and Sönmez's (2003) model of school choice and give a description of the mechanisms involved. Next, in Section 4, I address some important questions and complications for the full application of the theory to real school choice problems. I review and interpret a selection of the most recent developments and results.⁴ Section 5 concludes.

2. Why mechanism design in school choice?

The key issue of school choice is providing the parents of a child the possibility to have a say in the assignment of their child to a public school. In their study of school choice plans in the US, Abdulkadiroğlu and Sönmez (2003) mentioned two basic shortcomings of several school choice programs until recently. First, it turned out that many school districts simply did not have rigorous procedures. (In fact, until a decade ago children were assigned to public schools without taking into account the preferences of the parents at all.) This very often led to evasive action by the students and their parents that took its ultimate form in court cases, which were very often won by the parents, making the procedure even more vulnerable to future "attacks." Second, the school choice programs that did have explicit procedures suffered from serious short-

comings. A representative case is the mechanism that was used in the Boston school district. Since seats in schools are a scarce resource it is generally not possible to assign each student to his most preferred school. Therefore, apart from the students' preferences the Boston (and other) school district authorities took into account the priorities of students for schools. The priority of a student for a particular school typically is determined by factors as living in the walk zone of the school, siblings already attending the school, etc.

The major problem with the Boston mechanism was that students possibly could benefit from misrepresenting their preferences over schools. In other words, the Boston mechanism is not strategy-proof. A more detailed description of the mechanism is given in the next section, but to see why it is not strategy-proof it is enough to know that it first tries to assign as many students as possible to their first choice. More precisely, seats of a school are filled by students that put it as a first choice, one by one, following the priority order for that school. It is only when all possible first choices are assigned when the mechanism considers the second choices of the remaining (still unassigned) students. The problem is that a student that remains unassigned in the first round now may fail to get its second choice because this school has exhausted its capacity with students that put it as a first choice. This may even be true for a school that the student would have been assigned to had he put it as a first choice. In other words, even if the student has a higher priority than the students that were assigned to the school in the first round, the school remains out of reach. Hence, the Boston mechanism forces participants to play a complicated admissions game. Another cost of the Boston mechanism not being strategy-proof is that it does not provide the information the authorities would like to have — if parents do not reveal their true preferences it is hard to set the right policies or make appropriate changes. Abdulkadiroğlu and Sönmez (2003) therefore argued replacing the Boston mechanism by the Gale-Shapley or Top Trading Cycles mechanisms: both mechanisms are strategy-proof and each has additional desirable properties. An experimental study of Chen and Sönmez (2006) confirmed the superiority of the

⁴By no means do I claim this review to be exhaustive. The number of studies on school choice using matching theory is quickly growing.

latter two mechanisms over the Boston mechanism. Their data shows that the Boston mechanism induces massive preference manipulation, and, as a consequence, significant welfare loss.

Abdulkadiroğlu, Pathak, Roth, and Sönmez (2006) provided empirical evidence that under the Boston mechanism there are different levels of sophistication of play. On the one hand, there are parents groups that discuss how to submit preferences strategically. On the other hand, there is also a large number of parents that do not strategize (well). As a consequence, the first groups of parents systematically obtain “better” results than the latter. This evidence was one of the reasons why the Boston school district authorities replaced the Boston mechanism by the Gale-Shapley mechanism. Abdulkadiroğlu, Pathak, Roth, and Sönmez (2005) report on some further design considerations.

Around the same time, Atila Abdulkadiroğlu, Parag Pathak, Alvin Roth, and Tayfun Sönmez assisted the New York City department in the design of a new student assignment mechanism. New York City has the largest public school system in the US and each year there are about 90,000 entering students. Abdulkadiroğlu, Pathak, and Roth (2005) pointed out that the matching system that was in use suffered from three problems. First, there were not enough rounds to allocate all students, and as a consequence, about 30,000 students were assigned to a school that was not on their list. Second, the parents had to strategize (very much for the same reason as in Boston). Third, schools acted as strategic agents by withholding some of their capacity. This last point already hints at what was concluded after the initial discussions in the redesign process: the New York City schools are a two-sided market. The experience and the success of the redesign of a very similar two-sided market, where American physicians are assigned to hospitals through the Gale-Shapley mechanism, became very useful (see Roth and Peranson, 1999 and Roth, 2002 for details). It was decided to adapt the Gale-Shapley mechanism to the particular needs and characteristics of the New York City school match. Abdulkadiroğlu, Pathak, and Roth (2005) report on the first year of operation of the new mechanism. In contrast to the previous mechanism, only about 3,000 were assigned to schools they did not list. One (but not the only) reason is that in the new mechanism

students are allowed to rank 12 schools (instead of the previous maximum of 5). In Section 4.2, I will discuss this restriction on the length of submittable preference lists in more detail.

Further theoretical, experimental, and empirical studies have disentangled some of the open problems and have provided additional insights in the design of student assignment mechanisms. In the next session I will first recall the basic model and give a description of the three mechanisms. Subsequently I will focus on some concerns that mostly deal with the gap between the simplifications inherent to the modeling and real-life school choice procedures.

3. The model

Following Abdulkadiroğlu and Sönmez (2003), a *school choice problem* is defined by a set of schools and a set of students, each of which has to be assigned a seat at not more than one of the schools. Each student is assumed to have strict preferences over the schools and the option of remaining unassigned. Each school is endowed with a strict priority ordering over the students and a fixed capacity of seats. Formally, a *school choice problem* is a 5-tuple (I, S, q, P, f) that consists of

1. a set of *students* $I = \{i_1, \dots, i_n\}$,
2. a set of *schools* $S = \{s_1, \dots, s_m\}$,
3. a *capacity* vector $q = (q_{s_1}, \dots, q_{s_m})$,
4. a profile of strict *student preferences* $P = (P_{i_1}, \dots, P_{i_n})$, and
5. a strict *priority structure* of the schools over the students $f = (f_{s_1}, \dots, f_{s_m})$.

We denote by i and s a generic student and a generic school, respectively.

The preference relation P_i of student i is a linear order over $S \cup \{i\}$, where i denotes his outside option (e.g., enrolling in a private school). Student i prefers school s to school s' if $sP_i s'$. School s is *acceptable* to i if $sP_i i$. Let R_i denote the weak preference relation associated with the preference relation P_i .

The *priority ordering* f_s of school s assigns ranks to students according to their priority for school s . The *rank* of student i for school s is $f_s(i)$. Then, $f_s(i) < f_s(j)$ means that *student i has higher priority (or lower rank) for school s than student j .*

Throughout the paper the set of students I and the set of schools S do not vary. Hence, a *school choice problem* is given by a triple (P, f, q) , or simply by P when no confusion is possible.

School choice is closely related to the college admissions model (Gale and Shapley, 1962). The only but key difference between the two models is that in school choice schools are mere “objects” to be consumed by students, whereas in the college admissions model (or more generally, in two-sided matching) both sides of the market are agents with preferences over the other side. In other words, a college admissions problem is given by 1–4 above and 5’ below:

$$5'. \text{ a profile of strict school preferences } P_S = (P_{s_1}, \dots, P_{s_m}),$$

where P_s denotes the strict preference relation of school $s \in S$ over the students.

Priority orderings in school choice can be reinterpreted as school preferences in the college admissions model. Therefore, many results or concepts for the college admissions model have their natural counterpart for school choice.⁵ In particular, an outcome of a school choice or college admissions problem is a *matching* $\mu : I \cup S \rightarrow 2^I \cup S$ such that for any $i \in I$ and any $s \in S$,

- $\mu(i) \in S \cup \{i\}$,
- $\mu(s) \in 2^I$,
- $\mu(i) = s$ if and only if $i \in \mu(s)$, and
- $|\mu(s)| \leq q_s$.

For $i \in I$, if $\mu(i) = s \in S$ then student i is *assigned* a seat at school s under μ . If $\mu(i) = i$ then student i is *unassigned* under μ .

A key property of matchings in the two-sided matching literature that does not lose its importance in school choice is stability. Informally, a matching is stable if, for any student, all the schools he prefers to the one he is assigned to have exhausted their capacity with students that have higher priority. Formally, let P be a school choice problem. A matching μ is stable if

- it is *individually rational*, i.e., for all $i \in I$, $\mu(i) R_i i$,
- it is *non wasteful* (Balinski and Sönmez, 1999), i.e., for all $i \in I$ and all $s \in S$, $s P_i \mu(i)$ implies $|\mu(s)| = q_s$, and

- there is no *justified envy*, i.e., for all $i, j \in I$ with $\mu(j) = s \in S$, $s P_i \mu(i)$ implies $f_s(j) < f_s(i)$.

The set of stable matchings is denoted by $S(P)$.

Another desirable property for a matching is Pareto-efficiency. In the context of school choice, the schools are mere “objects.” Therefore, to determine whether a matching is Pareto-efficient we only take into account students’ welfare. A matching μ' Pareto dominates a matching μ if all students prefer μ' to μ and there is at least one student that strictly prefers μ' to μ . Formally, μ' *Pareto dominates* μ if $\mu'(i) R_i \mu(i)$ for all $i \in I$, and $\mu'(i') P_{i'} \mu(i')$ for some $i' \in I$. A matching is *Pareto-efficient* if it is not Pareto dominated by any other matching. We denote the set of Pareto-efficient matchings by $PE(P)$.

A (student assignment) *mechanism* systematically selects a matching for each school choice problem. A mechanism is stable (or Pareto-efficient) if it always selects a stable (or Pareto-efficient) matching. A mechanism φ is manipulable by student i at problem P if there exists a preference relation P'_i such that $\varphi(P_{-i}, P'_i)(i) P_i \varphi(P)(i)$. A mechanism is *strategy-proof* if no student can ever manipulate it. In other words, a mechanism is strategy-proof if no student can ever benefit by unilaterally misrepresenting his preferences.⁶

3.1. The three mechanisms

The following concise description, which is taken from Calsamiglia et al. (2007), integrates the three student assignment mechanisms that I discussed earlier. It distinguishes between the students’ and the schools’ point of view. The reason is that in the eventual computations the three mechanisms only differ in the way a student is “rejected” by a school.

The Boston (β , BOS), Gale-Shapley (γ , GS), and Top Trading Cycles (τ , TTC) mechanisms:

Step 1: For each school, a priority ordering of students is determined (based on state and local laws/policies, etc.).

Step 2: Each student submits a preference ranking of the schools.

⁵See, for instance, Balinski and Sönmez (1999).

⁶In game theoretic terms, a mechanism is strategy-proof if truthful preference revelation is a weakly dominant strategy.

Step 3: The assignment of seats is obtained through an algorithm, as follows.

ROUND k , $k \geq 1$ [STUDENTS]: Each student that has not been removed yet but is rejected in the previous round⁷ points to the next highest ranked school in his submitted list that has not been removed yet (if there is no such school then the student points to himself).

ROUND k , $k \geq 1$ [SCHOOLS, BOS]: Each school assigns seats one at a time to the students that point to it following its priority order. If the school capacity is or was attained, the school rejects any remaining students that point to it. If a student points to himself, he is assigned to himself. Any student that is assigned is removed.

The Boston algorithm terminates when all students have been removed.

ROUND k , $k \geq 1$ [SCHOOLS, GS]: Each school tentatively assigns seats one at a time to the students that point to it following its priority order. When the school capacity is attained the school rejects any remaining students that point to it. If a student points to himself, he is tentatively assigned to himself.

The Gale-Shapley algorithm terminates when no student is rejected. The tentative matching becomes final.

ROUND k , $k \geq 1$ [SCHOOLS, TTC]: Each school that has not been removed yet points to the student with highest priority among the students that have not been removed yet. There is at least one cycle. If a student is in a cycle he is assigned a seat at the school he points to (or to himself if he is in a self-cycle). Students that are assigned a seat are removed. If a school is in a cycle then its number of vacant seats is decreased by 1. If a school has no longer vacant seats then it is removed.

The Top Trading Cycles algorithm terminates when all students or schools have been removed.

The Gale-Shapley and Top Trading Cycles mechanisms are strategy-proof, but the Boston mechanism is not. The Gale-Shapley mechanism is also stable. In fact, it always generates the best stable matching for the students. However, it is not Pareto

efficient. In contrast, the Top Trading Cycles mechanism is Pareto efficient but not stable. In fact, it is well-known that stability and efficiency are incompatible. That is, there is no mechanism that is both stable and efficient. The Boston mechanism is Pareto efficient, but since it is not strategy-proof, Pareto efficiency is likely to be distorted in practice.

4. Issues in the design of student assignment mechanisms

Below I will discuss, in no particular order, some of the key issues that played a role in the design of student assignment mechanisms or that possibly will be relevant in future redesigns.

4.1. Indifference classes in the priority structure

An important question in the application of theory of school choice to real-life situations is the presence of indifferences in the priority structure f . In practical situations there is only a limited number of relevant criteria that determine the priority of each student for a given school. Thus, it may happen that some students have exactly the same priority, i.e., there are distinct students i, j and a school s such that $f_s(i) = f_s(j)$. Then, instead of having a strict priority list, a school's priority list typically contains several indifference classes. From the school's point of view, all students in the same indifference class are indistinguishable. Clearly, this complication is relevant if (and only if) some students of the same indifference class compete for the same school seat.

A first idea is to break the ties in order to get rid of the indifference classes and obtain a strict priority structure. Next, one can apply either the Gale-Shapley or Top Trading Cycles mechanism. Regarding the Top Trading Cycles mechanism, Pathak (2007) shows that it is insensitive to how ties are broken. The Gale-Shapley mechanism, however, is very much affected by the way the ties are broken. Erdil and Ergin (2008) show by means of an example that if we break ties in an arbitrary way, then it is possible that the stable matching induced by the Gale-Shapley mechanism is Pareto dominated by another stable matching. Gale and Shapley (1962) showed that this is never the case in their original setting without indifferences.

⁷In order to incorporate the initial step correctly, we use the convention that at "Round 0" all students are rejected.

Erdil and Ergin (2008) propose the following procedure to obtain a stable matching that is not Pareto dominated by any other stable matching. They first introduce the concept of a stable improvement cycle.

Definition 4.1. (Erdil and Ergin, 2008)

A matching μ admits a *stable improvement cycle* if there is a cycle of distinct students $\langle i_1, \dots, i_n =: i_0 \rangle$ ($n \geq 2$) such that for any $l = 0, \dots, n-1$, $\mu(i_{l+1})P_{i_l}\mu(i_l) \in S$ and $f_{\mu(i_{l+1})}(i_l) = \min \{f_{\mu(i_{l+1})}(j) : j \in I \text{ and } \mu(i_{l+1})P_j\mu(j)\}$. \triangle

Loosely speaking, each student i in a stable improvement cycle desires the school s to which the next student in the cycle is assigned at μ , and in addition student i has the highest priority (i.e., lowest rank order) among all students that desire school s .

Given a stable improvement cycle $\langle i_1, \dots, i_n =: i_0 \rangle$ for a matching μ , one can construct a new matching μ' as follows,

$$\mu'(j) := \begin{cases} \mu(j) & \text{if } j \notin \{i_1, \dots, i_n\}, \\ \mu(i_{l+1}) & \text{if } j = i_l \text{ for} \\ & \text{some } l = 0, \dots, n-1. \end{cases}$$

One easily verifies that if μ is stable, then the matching μ' that results from “satisfying” the stable improvement cycle is again a stable matching.

Theorem 4.1. (Erdil and Ergin, 2008)

Let μ be a stable matching. If μ is Pareto dominated by another stable matching, then μ admits a stable improvement cycle.

This result induces a simple algorithm to find a matching μ^* that is stable and constrained efficient. Here, *constrained efficient* means that there is no other *stable* matching that Pareto dominates μ^* . (Of course, this does not discard the existence of an unstable matching that Pareto dominates μ^* .) First, break the ties of the priority structure in an arbitrary way and apply the deferred acceptance algorithm to obtain a stable matching. Next, satisfy iteratively stable improvement cycles. Clearly, this algorithm terminates in a finite number of steps and by Theorem 4.1 the resulting matching is stable and constrained efficient. In fact, Erdil and Ergin (2008) show that their algorithm has a remarkably small computational complexity. More-

over, Abdulkadiroğlu, Pathak, and Roth (2008) report that if the algorithm would have been applied to the preference data of the 2003-2004 New York City school match, then more than 10% of the about 64,000 students involved would have been assigned to a strictly preferred school, without hurting the other students.

Given that in practice the priority structure is not strict, why not immediately replace the Gale-Shapley mechanism by the algorithm proposed by Erdil and Ergin (2008)? A problem of the latter algorithm is that it cannot induce a strategy-proof mechanism. (Note that there is range of possibilities for such mechanism since the final matching crucially depends on the tie-breaking.) In other words, unlike the Gale-Shapley mechanism, it does not make it a (weakly) dominant strategy to state one’s true preferences. In fact, Erdil and Ergin (2008) proved the following impossibility result.

Theorem 4.2. (Erdil and Ergin, 2008)

There is no mechanism that is both strategy-proof and constrained efficient.

Abdulkadiroğlu, Pathak, and Roth (2008) further analyzed the impact of the presence of ties in the priority structure. A *tie-breaker* is a bijection $r : I \rightarrow \{1, \dots, n\}$ and can be used to break ties at a school s by replacing its priority f_s by f'_s as follows: $f'_s(i) < f'_s(j)$ if and only if [$f_s(i) < f_s(j)$ or [$f_s(i) = f_s(j)$ and $r_s(i) < r_s(j)$]]. A *tie-breaking rule* is a collection of tie-breakers, one for each school. A single-tie breaking rule uses the same tie-breaker for all schools, a multiple-tie breaking rule may use different tie-breakers for different schools. Let γ^r denote the mechanism that yields the matching from the deferred acceptance algorithm, after breaking possible ties according to r .

Using data from the New York City school match they compared different single and multiple-tie breaking rules obtained from independent draws from uniform distributions. On the one hand, single-tie breaking does not stochastically dominate multiple-tie breaking. On the other hand, the mean and the number of students that receive their top choices are higher under single-tie breaking than multiple-tie breaking. The next theoretical result further supports the use of a single-tie breaking rule.

Theorem 4.3. (Abdulkadiroğlu, Pathak, and Roth, 2008)

For any priority structure f and any school choice problem P , if μ is constrained efficient then there is a single-tie breaking rule r such that $\gamma^r(P) = \mu$.

As has been pointed out before, if we use a tie-breaking rule and apply the deferred acceptance algorithm then the resulting matchings need not be constrained efficient. Theorem 4.2 implies that Pareto improving upon these matchings comes at the price of losing strategy-proofness. Abdulkadiroğlu, Pathak, and Roth (2008) further strengthen this result as follows. A mechanism dominates another mechanism if the first mechanism always gives a weakly better match to all students than the second mechanism, and for some school choice problem and for some student the first mechanism gives a strictly better match than the second mechanism.

Theorem 4.4. (Abdulkadiroğlu, Pathak, and Roth, 2008)

For any tie-breaking rule r , there is no strategy-proof mechanism that dominates γ^r .

4.2. Constrained school choice

Real-life school choice situations typically involve a large number of participants and a relatively small number of school programs. Parents are often asked to submit a preference list containing only a limited number of schools.⁸ In other words, there is a constraint or quota on the number of schools that can be listed. This restriction is reason for concern: true preference revelation is (typically) no longer an option. As a consequence, the Gale-Shapley and Top Trading Cycles mechanisms are no longer strategy-proof. Since the (desirable) properties of the two mechanisms are relative to the revealed preferences, it is far from clear what these properties mean in case the revealed preferences are not the true preferences. In the setting of “constrained school choice” (i.e., school choice with a restriction on the length of submittable preference lists), it is likely that participants adopt strategic behavior. For instance, if a participant fears rejection by his most preferred programs, it can be advantageous not to apply to these programs and

use instead its allowed application slots for less preferred programs.

Manipulability

Pathak and Sönmez (2008b) develop a notion to compare mechanisms that are not strategy-proof based on the degree to which they encourage manipulation. Subsequently, they use their notion to compare several well-known mechanisms in the matching and auction literature. One specific case is the setting of constrained school choice where the Gale-Shapley and Top Trading Cycles mechanisms are not strategy-proof.

Pathak and Sönmez (2008b) call a mechanism ψ weakly more manipulable than mechanism φ if whenever φ is manipulable, ψ is also manipulable, even though the converse does not hold. Equivalently, ψ is weakly more manipulable than φ if whenever truth-telling is a Nash equilibrium under ψ truth-telling is a Nash equilibrium under φ as well.

Theorem 4.5. (Pathak and Sönmez, 2008b)

The stronger the restriction on the length of submittable preferences lists (i.e., the smaller the quota) the weakly more manipulable becomes the Gale-Shapley mechanism.

Theorem 4.6. (Pathak and Sönmez, 2008b)

Let $1 \leq k < m$. If students can list at most k schools, then the Boston mechanism is weakly more manipulable than the Gale-Shapley mechanism.

Pathak and Sönmez (2008b) consider also a second notion to compare manipulability of different mechanisms: a mechanism ψ is strongly more manipulable than mechanism φ if at any profile R , any student that can manipulate φ can also manipulate ψ , even though the converse does not hold. Next, they show by means of an example that Theorem 4.6 cannot be strengthened through replacement of the first notion by the second notion.

An interesting open problem is how the Top Trading Cycles mechanisms compares to the other two mechanisms and whether it becomes more manipulable as the restriction on the length of submittable preferences becomes more stringent.

Stability and efficiency

Haeringer and Klijn (2007) study the impact of the constraint on stability and efficiency by intro-

⁸For instance, in the school district of New York City each year more than 90,000 students are assigned to about 500 school programs, and parents are asked to submit a preference list containing at most 12 school programs (Abdulkadiroğlu, Pathak, and Roth, 2005, 2008).

ducing a preference revelation game where students can only declare up to a fixed number (the *quota*) of schools to be acceptable. Each possible quota, from 1 up to the total number of schools, together with a student assignment mechanism induces a strategic “quota-game.” Since the presence of the quota eliminates the existence of a dominant strategy when the mechanism at hand is the Gale-Shapley and Top Trading Cycles mechanisms, we focus on the Nash equilibria of the quota-games.

Fix the priority ordering f and the capacities q . Given a profile of preferences $P = (P_{i_1}, \dots, P_{i_n})$, a quota k with $1 \leq k \leq m$, and a mechanism φ , the induced *quota-game* $\Gamma^\varphi(P, k)$ is a triple $\langle I, \mathcal{Q}(k)^I, P \rangle$. The strategy set of each player (student) is the set of preference lists with at most k acceptable schools which is denoted by $\mathcal{Q}(k)$. Outcomes of the game are evaluated through the true preferences P . Let $\mathcal{E}^\varphi(P, k)$ and $\mathcal{O}^\varphi(P, k)$ denote the set of Nash equilibria and Nash equilibrium outcomes, respectively. Haeringer and Klijn (2007) show that for all three mechanisms discussed above there are Nash equilibria in pure strategies.

Below I summarize the results regarding the stability of the equilibrium outcomes, and subsequently comment on efficiency. The following benchmark is, given the negative empirical evidence regarding the Boston mechanism, quite surprising.

Theorem 4.7. (Ergin and Sönmez, 2006 and Haeringer and Klijn, 2007)⁹

For any quota k and any school choice problem P , the game $\Gamma^\beta(P, k)$ implements $S(P)$ in Nash equilibria, i.e., $\mathcal{O}^\beta(P, k) = S(P)$.

When $k = 1$ the deferred acceptance algorithm consists of only one step, which moreover coincides with the (then also unique step) of the Boston algorithm, i.e., $\Gamma^\gamma(P, 1) = \Gamma^\beta(P, 1) = S(P)$. In case $k \neq 1$, it is easy to see that under the Gale-Shapley mechanism each stable matching can still be obtained as an equilibrium outcome. However, as Example 3 in Sotomayor (1998) already showed, not all equilibrium outcomes need to be stable. Therefore, an important question is to find out whether and when we can guarantee a stable outcome. One possible interpretation of “when” is to find necessary and sufficient conditions on the priority structure.

⁹Ergin and Sönmez (2006) deal with $k = m$. Kojima (2008) extends their result to the case with general priority structures.

For $s \in S$ and $i \in I$, we denote by $U_s^f(i)$ the set of students that have higher priority than student i for school s , i.e., $U_s^f(i) = \{j \in I : f_s(j) < f_s(i)\}$.

Definition 4.2. Ergin-Acyclicity (Ergin, 2002)

Given a priority structure f , an *Ergin-cycle* is constituted of distinct $s, s' \in S$ and $i, j, l \in I$ such that the following two conditions are satisfied:

Ergin-cycle condition: $f_s(i) < f_s(j) < f_s(l)$ and $f_{s'}(l) < f_{s'}(i)$ and

ec-scarcity condition: there exist (possibly empty and) disjoint sets $I_s, I_{s'} \subseteq I \setminus \{i, j, l\}$ such that $I_s \subseteq U_s^f(j)$, $I_{s'} \subseteq U_{s'}^f(i)$, $|I_s| = q_s - 1$, and $|I_{s'}| = q_{s'} - 1$. A priority structure is *Ergin-acyclic* if no Ergin-cycles exist. \triangle

Theorem 4.8. (Haeringer and Klijn, 2007)

Let $k \neq 1$. Then, f is an Ergin-acyclic priority structure if and only if for any school choice problem P , $\mathcal{O}^\gamma(P, k) = S(P)$.

Definition 4.3. Kesten-Acyclicity (Kesten, 2006)

Given a priority structure f , a *Kesten-cycle* is constituted of distinct $s, s' \in S$ and $i, j, l \in I$ such that the following two conditions are satisfied:

Kesten-cycle condition $f_s(i) < f_s(j) < f_s(l)$ and $f_{s'}(l) < f_{s'}(i), f_{s'}(j)$ and

kc-scarcity condition there exists a (possibly empty) set $I_s \subseteq I \setminus \{i, j, l\}$ with $I_s \subseteq U_s^f(i) \cup [U_s^f(j) \setminus U_{s'}^f(l)]$ and $|I_s| = q_s - 1$.

A priority structure is *Kesten-acyclic* if no Kesten-cycles exist. \triangle

Kesten-acyclicity implies Ergin-acyclicity (Lemma 1, Kesten, 2006). It is easy to check that the reverse holds if all schools have one seat.

Theorem 4.9. (Haeringer and Klijn, 2007)

Let $1 \leq k \leq m$. Then, f is a Kesten-acyclic priority structure if and only if for any school choice problem P , $\mathcal{O}^\tau(P, k) = S(P)$.

Theorems 4.7–4.9 show that in all three school choice procedures stability can be guaranteed by strategic interaction in spite of possible constraints on preference revelation. While no particular assumption is needed for the Boston mechanism, stringent conditions are required for the Gale-Shapley and the Top Trading Cycles mechanisms. Since real-life priority structures typically do not satisfy these conditions, the transition to either of

the supposedly superior mechanisms may yield “unfair” assignments in the sense that there are students that prefer a seat that is occupied by a lower priority student.

Regarding efficiency, Haeringer and Klijn (2007) identify similar (but new) acyclicity conditions and show that they are necessary and sufficient to guarantee the efficiency of the three mechanisms. Qualitatively, the two (main) differences are that (i) the Boston and Gale-Shapley mechanisms perform equally well (or bad), and (ii) the Top Trading Cycles mechanism performs better than the other two mechanisms.

As a closing remark, the acyclicity conditions that guarantee stability or Pareto-efficiency are very limiting, and unlikely to hold in practice. This is a clear call for unrestricting preference revelation, i.e., setting $k = m$. Of course the negative results above still apply for $k = m$ but it is likely that in this case students use their (weakly) dominant strategy, namely submitting their true preferences. Recall that in the unconstrained setting the Gale-Shapley and Top Trading Cycles mechanisms yield a stable and efficient matching, respectively. The choice for either of the two mechanisms then depends on whatever has the highest priority for the policy makers: stability or efficiency.

4.3. Manipulation by schools

The formal model of school choice only allows the students to act strategically. More precisely, the priorities and capacities of the schools are determined by law and there is no room for strategizing. Nevertheless, Abdulkadiroğlu, Pathak, and Roth (2005) noted that in the New York City school match schools acted as strategic agents by withholding some of their capacity (Sönmez, 1997). Another possibility of manipulation is through pre-arranged matches (Sönmez, 1999): a student-school pair commit to a mutually beneficial agreement prior to the centralized procedure, according to which (i) the student does not participate in the procedure and (ii) he is rewarded with a seat at the school. Here, mutually beneficial means that at least the student or the school obtains a strictly better match (and the other agent does not get hurt). Sönmez (1997, 1999) showed that any stable mechanism can be manipulated through capacity withholding and also through pre-arranged matches.

In many school districts, each school is required to admit a minimum number of students. In other words, in this extended model each school s cannot declare less than say q_s seats. To formalize the first type of manipulation we need the following piece of notation. Let P_s denotes the strict preference relation of school $s \in S$ over the (individual) students (see item 5’ in Section 3). More precisely, $f_s(i) < f_s(j)$ if and only if $iP_s j$ for all $i, j \in I$. With a slight abuse of notation, let P_s also denote the preferences of school s over sets of students. A usual assumption in the literature is that for each school $s \in S$ the preferences P_s are responsive (Roth, 1985), i.e.,

- if $i \notin I'$ and $|I'| < q_s$, then $(I' \cup \{i\})P_s I'$ if and only if $iP_s \emptyset$, and
- if $i \notin I'$ and $j \in I'$, then $((I' \setminus \{j\}) \cup \{i\})P_s I'$ if and only if $iP_s j$.

Let $P_S = (P_{s_1}, \dots, P_{s_m})$.

Definition 4.4. (Sönmez, 1997)

A mechanism φ is *non-manipulable via capacities* if for all (P_S, q) , all $s \in S$, and all $q_s < q'_s < q_s$, $\varphi(P_S, q)(s)R_s \varphi(P_S, q'_s, q_{-s})(s)$. \triangle

Theorem 4.10. (Kesten, 2008b)

The Boston and Top Trading Cycles mechanisms are non-manipulable via capacities. The Gale-Shapley mechanism is non-manipulable via capacities if and only if f is an Ergin-acyclic priority structure.

The second type of manipulation by schools is formalized as follows.

Definition 4.5. (Sönmez, 1999)

A mechanism φ is *non-manipulable via pre-arranged matches* at (I, P_I, P_S, q) if there is a student $i \in I$ and a school $s \in S$ such that $sR_i \varphi(I, P_I, P_S, q)(i)$ and $\left(\{i\} \cup \varphi(I \setminus \{i\}, P_{I \setminus \{i\}}, P_S^{I \setminus \{i\}}, q_{-s}, q_s - 1)(s) \right) R_s \varphi(I, P_I, P_S, q)(s)$ with at least one of the relations holding strictly. \triangle

Here, $P_S^{I \setminus \{i\}}$ denotes the preferences of the schools over the set of students $I \setminus \{i\}$. As the following result shows, it is virtually impossible to avoid manipulation via pre-arranged matches.

Theorem 4.11. (Kesten, 2008b)

Suppose that at (I, P_I, P_S, q) for some school $s \in S$, $q_s < n$. Then, any mechanism is manipulable via pre-arranged matches at some problem.

Kojima (2007) studies the same two kinds of manipulation. As Kojima (2007) points out, there are two differences between his and Kesten's (2008b) paper. First, Kojima (2007) follows the literature on school choice by assuming that the schools' priorities (or preferences) are publicly known. Kesten (2008b) assumes that the priorities are private information. Second, Kojima (2007) obtains conditions in terms of preferences of an individual school under which that particular school cannot manipulate. Clearly, Kesten's results (Theorems 4.10 and 4.11) deal with conditions on the entire priority structure such that no school can manipulate.

Definition 4.6. (Konishi and Ünver, 2006)
Preference relation P_s is *strongly monotone in population* if, if $q_s > |I'| > |I''|$ and each student $i \in I'$ is acceptable (i.e., $iP_s\emptyset$), then $I'P_sI''$. \triangle

The following two results show that the class of strongly monotone preferences in population is a maximal domain for non-manipulability via capacities of the Gale-Shapley mechanism.

Theorem 4.12. (Konishi and Ünver, 2006)
If P_s is *strongly monotone in population*, then s cannot manipulate the Gale-Shapley mechanism via capacities.

Theorem 4.13. (Kojima, 2007)
Fix $I, S, s \in S, P_s$, and q_s . If P_s is not *strongly monotone in population*, then there exist preferences of students and other schools (P_I, P_{-s}) such that s can manipulate the Gale-Shapley mechanism via capacities. The preferences of the other schools can be taken as *strongly monotone in population*.

Kojima (2007) also identifies a maximal domain for the second type of manipulation.

Definition 4.7. (Kojima, 2007)
Preference relation P_s is *weakly maximin* if $[q_s = |I'| = |I''|]$, each student $i \in I'$ is acceptable, and the least preferred student in I' is strictly preferred to the least preferred student in I'' implies $I'P_sI''$. \triangle

Theorem 4.14. (Kojima, 2007)

(i) If P_s is *weakly maximin*, then s cannot manipulate the Gale-Shapley mechanism via pre-arranged matches.

(ii) Fix $I, S, s \in S, P_s$, and q_s . If P_s is not *weakly maximin*, then there exist preferences of students and other schools (P_I, P_{-s}) such that s can manipulate the Gale-Shapley mechanism via pre-arranged matches. The preferences of the other schools can be taken as *weakly maximin*.

4.4. Further issues

There are many other important issues and new developments in the field. I just mention three. Abdulkadiroğlu and Ehlers (2006) study how in practice one can assign students to school while maintaining racial and ethnic balance. They introduce a notion of fairness and show that there is always a constrained non-wasteful matching that satisfies it. On the other hand, they prove that there is no such mechanism that is also strategy-proof.

Kesten (2008a) provides theoretical and computational evidence that the Gale-Shapley mechanism may suffer large welfare losses. He proposes an adjustment such that a student's waiving his priority for a particular school leads to a Pareto improvement. He further shows that the adjustment practically does not disrupt strategy-proofness.

Abdulkadiroğlu, Che, and Yasuda (2008) expand the set of strategies by allowing students to "signal" a school (in addition to a preference list). They introduce a new mechanism and show that it improves upon the Gale-Shapley mechanism.

5. Concluding remarks

It has become clear that matching theory and mechanism design can provide a better understanding of several real-life markets and help to improve their working.¹⁰ The case of school choice is a clear example. The assistance of economists in the design of school choice programs has led to student assignment mechanisms that are considered fair or efficient. Additionally, the new mechanisms have also taken away concerns of the parents since they no longer have to strategize and can simply submit their true preferences. As an illustration, the Institute for Innovation in Public School Choice, which

¹⁰See, for instance, Roth, 2008b.

¹¹For further information and developments see <http://iipsc.org/index.htm> or Al Roth's page.

is a nonprofit organization in the US, makes use of the know-how of economists in the field of market design.¹¹

On the other hand, recent literature shows there is still room for further improvements in student assignment mechanisms. I mentioned for instance the presence of a cap on the length of submittable preference lists is harmful, and there is no clear reason to not remove it. It is also necessary to take care of the ties in the priority structure. All of this could lead to additional social welfare. An important but difficult task for “economic engineers” is to convince authorities how one should deal with these and other market failures. One difficulty is that authorities or clearinghouses may not always be eager to share information about the exact procedures or cannot reveal the (submitted) preferences of the participants. Finally, groups of interest may be resistant against redesign. For the case of school choice, Pathak and Sönmez (2008a) identified groups of parents that invested in learning about the Boston mechanism. These parents obtained a clear advantage over the other participants, and opposed changing the mechanism.

The interaction between practical problems and theory leads also to new interesting theoretical problems. One challenge is to model, study, and compare different informational environments (that correspond to more realistic settings). Another challenge is to find out how well the mechanisms perform. For instance, if a mechanism is manipulable, inefficient, or unstable, how manipulable, inefficient, or unstable is it really?

References

- [1] Abdulkadiroğlu, A. Y.-K. Che, and Y. Yasuda (2008) “Expanding Choice in School Choice,” Mimeo, Duke University.
- [2] Abdulkadiroğlu, A. and L. Ehlers (2006) “Controlled School Choice,” Mimeo, CIREQ, Université de Montréal.
- [3] Abdulkadiroğlu, A., P.A. Pathak, and A.E. Roth (2005) “The New York City High School Match,” *American Economic Review, Papers and Proceedings*, **95**, 364–367.
- [4] Abdulkadiroğlu, A., P.A. Pathak, and A.E. Roth (2008) “Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match,” Mimeo, Harvard University.
- [5] Abdulkadiroğlu, A., P.A. Pathak, A.E. Roth, and T. Sönmez (2005) “The Boston Public Schools Match,” *American Economic Review, Papers and Proceedings*, **95**, 368–371.
- [6] Abdulkadiroğlu, A., P.A. Pathak, A.E. Roth, and T. Sönmez (2006) “Changing the Boston School Choice Mechanism: Strategy-Proofness as Equal Access,” Mimeo, Boston College and Harvard University.
- [7] Abdulkadiroğlu, A. and T. Sönmez (2003) “School Choice: A Mechanism Design Approach,” *American Economic Review*, **93**, 729–747.
- [8] Balinski, M. and T. Sönmez (1999) “A Tale of Two Mechanisms: Student Placement,” *Journal of Economic Theory*, **84**, 73–94.
- [9] Biró, P. (2008) “Higher Education Admission in Hungary by a Score-limit Algorithm,” Mimeo, University of Glasgow.
- [10] Calsamiglia, C. G. Haeringer, and F. Klijn (2007) “Constrained School Choice: An Experimental Study,” Mimeo, Universitat Autònoma de Barcelona.
- [11] Chen, Y. and T. Sönmez (2006) “School Choice: An Experimental Study,” *Journal of Economic Theory*, **127**, 202–231.
- [12] Erdil, A. and H. İ. Ergin (2008) “What’s the Matter with Tie-Breaking? (Improving Efficiency in School Choice),” *American Economic Review*, **98**, 669–689.
- [13] Ergin, H. İ. (2002) “Efficient Resource Allocation on the Basis of Priorities,” *Econometrica*, **70**, 2489–2497.
- [14] Ergin, H. İ. and T. Sönmez (2006) “Games of School Choice under the Boston Mechanism,” *Journal of Public Economics*, **90**, 215–237.
- [15] Gale, D. and L.S. Shapley (1962) “College Admissions and the Stability of Marriage,” *American Mathematical Monthly*, **69**, 9–15.

- [16] Gusfield, D. and R.W. Irving (1989) *The Stable Marriage Problem: Structure and Algorithms*. MIT Press, Cambridge.
- [17] Haeringer, G. and F. Klijn (2007) "Constrained School Choice," *Barcelona Economics Working Paper* 294.
- [18] Kesten, O. (2006) "On Two Competing Mechanisms for Priority-Based Allocation Problems," *Journal of Economic Theory*, **127**, 155–171.
- [19] Kesten, O. (2008a) "An Alternative Mechanism Design Approach to School Choice in the US," Mimeo, Tepper School of Business, Carnegie Mellon University.
- [20] Kesten, O. (2008b) "On Two Kinds of Manipulation for School Choice Problems," Mimeo, Tepper School of Business, Carnegie Mellon University.
- [21] Kojima, F. (2007) "When can Manipulations be Avoided in Two-Sided Matching Markets? Maximal Domain Results," *The B.E. Journal of Theoretical Economics* (contribution), Article 32.
- [22] Kojima, F. (2008) "Games of School Choice under the Boston Mechanism with General Priority Structures," *Social Choice and Welfare*, **31**, 357–365.
- [23] Konishi, H. and U. Ünver (2006) "Games of Capacity Manipulation in Hospital-Intern Markets," *Social Choice and Welfare*, **27**, 3–24.
- [24] Pathak, P.A. (2007) "Lotteries in Student Assignment," Mimeo, MIT.
- [25] Pathak, P.A. and T. Sönmez (2008a) "Leveling the Playing Field: Sincere and Strategic Players in the Boston Mechanism," *American Economic Review*, **98**, 1636–1652.
- [26] Pathak, P.A. and T. Sönmez (2008b) "Comparing Mechanisms by their Vulnerability to Manipulation," Mimeo, MIT.
- [27] Romero-Medina, A. (1998) "Implementation of Stable Solutions in a Restricted Matching Market," *Review of Economic Design*, **3**, 137–147.
- [28] Roth, A.E. (1985) "The College Admissions Problem is not Equivalent to the Marriage Problem," *Journal of Economic Theory*, **36**, 277–288.
- [29] Roth, A.E. (1990) "New Physicians: A Natural Experiment in Market Organization," *Science*, **250**, 1524–1528.
- [30] Roth, A.E. (2002) "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics," *Econometrica*, **70**, 1341–1378.
- [31] Roth, A.E. (2008a) "Deferred Acceptance Algorithms: History, Theory, Practice, and Open Questions," *International Journal of Game Theory*, **36**, 537–569.
- [32] Roth, A.E. (2008b) "What have We Learned from Market Design?," (Hahn Lecture) *Economic Journal*, **118**, 285–310.
- [33] Roth, A.E. and E. Peranson (1999) "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design," *American Economic Review*, **89**, 748–780.
- [34] Roth, A.E. and M.A.O. Sotomayor (1990) *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Econometric Society Monograph Series, Cambridge University Press, New York.
- [35] Shapley, L.S. and H. Scarf (1974) "On Cores and Indivisibility," *Journal of Mathematical Economics*, **1**, 23–37.
- [36] Sönmez, T. (1997) "Manipulation via Capacities in Two-Sided Matching Markets," *Journal of Economic Theory*, **77**, 197–204.
- [37] Sönmez, T. (1999) "Can Pre-arranged Matches be avoided in Two-Sided Matching Markets," *Journal of Economic Theory*, **86**, 148–156.
- [38] Sotomayor, M.A.O. (1998) "The Strategy Structure of the College Admissions Stable Mechanisms," Mimeo, Universidade de São Paulo.
- [39] Teo, C.-P., J. Sethuraman, and W.-P. Tan (2001) "Gale-Shapley Stable Marriage Problem Revisited: Strategic Issues and Applications," *Management Science*, **47**, 1252–1267.