Matching and Economic Design

7 November 2011

by Georgy Artemov[*], Sven Feldmann[§], and Simon Loertscher[*]

1. Introduction

As an economics student at an Australian university you may have applied to your course via your state's Tertiary Admissions Centre. Like you, many students wanted to study economics, but there were not enough places to accommodate them all. At the same time other courses may not have had enough quality applicants to fill all the places they had. How should students be assigned to courses? This is the kind of problem addressed by *matching theory*. However, the questions matching theory addresses go well beyond tertiary admission problems as there are many situations in which scarce resources need to be allocated to individuals with heterogeneous preferences without relying on monetary transfers or prices to balance demand and supply.

This article provides a brief survey of the literature on matching. This literature started with a seminal article by Gale and Shapley (1962), in which a general matching problem—of matching men and women in a marriage market—was introduced and analysed. Over the last twenty years or so, the matching literature has led to important real-world applications, beginning with the successful re-design of the national residency program for medical students in the US (Roth and Peranson 1999). In the New York City public school district a new process based on matching theory to assign high school students to schools reduced the number of unmatched students from about 30,000 to about 3,000 per year (Abdulkadiroglu, Pathak and Roth 2005, Klein 2011). Applications of matching theory have also helped to significantly increase the number of life-saving organ transplants (Roth 2005). Together with the economics of auctions, matching theory has become part of a new field known as economic design (Roth 2002; Loertscher and Wilkening 2011).

The article is structured as follows. Section 2 introduces the two-sided matching model. Sections 3 and 4 provide, respectively, results on stable matchings and strategy-proof matching mechanisms. Section 5 discusses the school choice model, and Section 6 concludes.

2. Two-sided Matching

Two-sided matching is concerned with allocating—or matching—agents from two separate groups (such as students and colleges, workers and firms, or men and women) to one another. We describe the theory in the context of the *college admission*

[*] *University of Melbourne* [§] *Melbourne Business School*

model, which is concerned with matching students and colleges. [1] The two disjoint groups are a set of n students, denoted by $S = \{s_1, s_2, ..., s_n\}$, and a set of m colleges, denoted by $C = \{c_1, c_2, \ldots, c_m\}$. Students can enrol in at most one college, and each college c_j can accept up to q_j students, the college's *capacity*.[2] Each student s has a *strict* preference over colleges so that they rank the colleges from most to least preferred with no indifference between any two colleges.[3]

We write every student's preference ranking as an ordered list of colleges in decreasing order of desirability. Some colleges may be so undesirable that they become unacceptable. We indicate unacceptable colleges by placing a student herself in her preference list with the interpretation that colleges that are listed below (or after) the student are unacceptable for the student. For example, $P_s = c_1, c_3, c_4, s$ means that student s prefers college c_1 to college c_3, college c_3 to college c_4, and beyond that prefers not to go to college.

A college's preferences are more complicated than those of a student. Since a college c can admit up to q_c students, it might care about the composition of the entering class (including its cultural or gender diversity, for instance). A college's preferences are thus defined over *sets* of students. A natural restriction of these preferences to make the problem more tractable is to assume that a college can rank, without ties, all students and prefers an entering class with a given (acceptable) student to a class in which this student is replaced by a lower-ranked student or an empty place. College preferences with this property are said to be *responsive* to their ranking of individual students, we assume this property throughout.

A college may wish to leave some places unfilled rather than admitting unacceptable students. We insert the name of college c in the ranking of students at the place where any student ranked after c is deemed unacceptable to the college. For example, $P_c = s_1, s_5, s_3, c$ means that college c ranks student s_1 above student s_5, and s_5 above s_3, finds all three acceptable but beyond that finds any other student unacceptable.

A *matching* is an assignment of every student to at most one college and of every college c to at most q_c students in such a way that if student s is assigned to college c, then college c is also assigned to student s. A student who is not assigned to any college is notionally assigned to him- or herself. Similarly, if the capacity of a college is not exhausted, then the empty slots are notionally filled by the college itself.

---

[1] Roth and Sotomayor (1990) provide a comprehensive survey of two-sided matching theory. All proofs of the theorems in this article can be found there.

[2] The term "college" corresponds to a *course*, or field of study at a given university, in the Australian tertiary admissions context. We use the term college here as it is the one used in the matching literature.

[3] For simplicity, we write s for a student (and c for a college) when we do not refer to a specific student or college.

A matching can come about via a decentralised process where students apply individually to colleges, colleges send out offers, and students respond with acceptances. Or, and this is the assumption employed in much of the matching literature, the agents submit their intentions to a central planner, or *clearinghouse*. The clearinghouse employs what is called a *matching mechanism:* a procedure that asks each student and every college to report their preferences (over colleges and students, respectively), and generates a matching based on the submitted preferences. While finding an appropriate matching is a non-trivial problem even if the agents' preferences are publicly known, in practice matching mechanisms rely on the preferences students and colleges report rather than on their true preferences.

Example

It is useful to illustrate some of the main findings of the literature with a simple example. Assume that there are four students, which for simplicity we call 1, 2, 3, and 4, and three colleges, called *A*, *B* and *C*. Colleges *A* and *B* have capacities of one each while *C* has a capacity of two. Preferences are as follows:

$P_1 = A, B, C$	$P_2 = B, A, C$	$P_A = 2, 1, 3, 4$	$P_B = 4, 1, 3, 2$
$P_3 = B, C, A$	$P_4 = C, B, A$	$P_C = 1, 2, 3, 4$	

Suppose that the matching mechanism is as follows: each student and each college first writes down their preferences. (For now we assume that they do so truthfully.) Based on the submitted preferences each student is proposed to her first-listed choice. If a college receives fewer proposals than its capacity, all proposed students are accepted; if the number of proposals exceeds the capacity, the students highest on the college's ranking are accepted up to capacity, the remainder is rejected. Students rejected in the first round are then proposed to their second choice and are accepted if there is capacity, otherwise they are proposed to their third choice, etc., until all students are assigned or all places available at the colleges are filled. This procedure has been used to assign students to public schools and is known in the literature as the *Boston mechanism*. Which matching results from this mechanism given the above preferences?

The allocation procedure is illustrated in Table 1, where each column represents a place at one of the colleges (hence, two for *C*). All new assignments in each round are indicated in bold and students unassigned in that round are in parentheses. In the first round all students are tentatively assigned to their first-choice college: student 1 to college *A*, 2 and 3 to *B*, and 4 to *C*. Yet, college *B* exceeds its capacity, and since it prefers student 3 to 2, 2 is rejected. In the second round, 2 is tentatively assigned to his next-preferred college, *A*, but is rejected because *A*'s capacity is already exhausted. In the third round, 2 is assigned to college *C* and is accepted. This leads to the final matching of 1 to *A*, 3 to *B*, and 2 and 4 to *C*.

Colleges	<u>*A*</u>	<u>*B*</u>	<u>*C*</u>	<u>*C*</u>
Round 1	**1**	**(2),3**	**4**	

Round 2	1,**(2)**	3	4	
Round 3	1	3	4	**2**

Table 1: The Boston Mechanism under truthful preferences

Three of the four students (1, 3, and 4) are matched to their top choice. Since improving the welfare of either college or student 2 would assign at least one of these three students to a college she likes less, the resulting matching under truthful applications is *Pareto efficient*.[4] Two important shortcomings of this mechanism are that (i) the outcome is not *stable*, and (ii) submitting untruthful preferences may lead the student to be matched to a more preferred college. We consider these two important properties in turn.

1. Stability

Observe that under the outcome of the Boston mechanism above, student 2 is matched to *C*, his least preferred college. Since 2 is the top-ranked student at college *A*, he could approach college *A* and suggest accepting him instead of student 1 to whom the college is matched by the mechanism. Both student 2 and college *A* would benefit from this trade. A mechanism is called *unstable* if such pairwise improvements over the final allocation are possible. In practice and in laboratory experiments it has been observed that participants learn to circumvent central mechanisms if its matchings are unstable, resulting in a collapse of the central allocation mechanism.[5]

A matching is (pairwise) *stable* if there is no pair consisting of a student and a college who are not assigned to each other but who would be made better off if they were matched to each other.

2. Non-manipulability

In applying the Boston mechanism above we have assumed that students and colleges submit their preferences truthfully. It turns out they do not necessarily have an inventive to do so. To see this, notice that student 2 could, instead of submitting his true preference list, put his second choice *A* at the top of his list. Since student 2 is ranked first at college *A*, he is accepted in round 1 with this list, a much better outcome for him than being matched with *C* when he submits his preferences truthfully. By submitting a less ambitious top choice the student can affect the resulting allocation in his favour.

Beneficial manipulations of this kind are often possible under mechanisms that, like the Boston mechanism, are not *strategy-proof*:

---

[4] An allocation is called Pareto efficient if there is no other feasible allocation that makes at least one individual strictly better off without making any other individual worse off.

[5] The resident matching mechanism used by the American Medical Association collapsed after the emergence of married couples seeking joint positions rendered it unstable (Roth 2008).

A matching mechanism is *strategy-proof* (or *non-manipulable*) if no agent (student or college) can ever benefit from reporting preferences other than their true ones.[6]

Mechanisms that are not strategy-proof impose a heavy burden on participants as they essentially require participants, for example students, to guess what their chances of acceptance at each college are, and then to make the difficult choice of how to rank a very desirable college with a low probability of being admitted and a less desirable, but 'safer' college.[7] With strategy-proof mechanisms, students do not need to make such calculations.

3. Finding a stable matching

A natural and important question is therefore whether there always exists some stable matching for any configuration of student and college preferences and, if so, whether a strategy-proof mechanism exists that implements a stable matching. The answer to the first question is affirmative:

Theorem 1. *In the college admission model a stable matching exists.*

Gale and Shapley (1962) showed that a stable matching exists by providing a procedure, called the *deferred-acceptance algorithm (DAA)*, which produces such a matching. The DAA occurs in two closely related versions. We first describe the *student proposing DAA*. It works as follows:

Students and colleges submit preference rankings to the clearinghouse. The clearinghouse then performs the following "rounds" of the algorithm. In the first round, the clearinghouse tentatively assigns each student to a "waitlist" at the student's most-preferred college. The clearinghouse then sorts the students on each college *c*'s waitlist according to that college preference over students and keeps the top q_c acceptable students on the list. Excess students and unacceptable students are crossed off the list.

In each subsequent round every student who was crossed off a waitlist in the previous round is now assigned to the waitlist at his or her next-preferred college. If there is no such college, the student remains unmatched. The clearinghouse then updates the waitlist for each college by keeping the top q_c students on the list and crossing off excess and unacceptable students and repeats the process.

The process ends when no more students are crossed off a list and all students who are not on a waitlist prefer remaining unmatched to being matched to any college to which they have not been proposed yet. The waitlists are then the final matching.

---

[6] Put differently, a strategy-proof matching mechanism is defined as a mechanism in which reporting preferences truthfully is a *weakly dominant strategy* for each agent.

[7] In the Australian tertiary match, students are often explicitly advised by the tertiary admission centres to be 'realistic' in their choices (Artemov et al., 2011). In the UK the use of Boston-style mechanisms is prohibited for being unfair (Department of Education 2007, 2010).

The key feature of the DAA is that the matching is not finalised and a student is not accepted by a college until *all* students have been allocated (possibly to themselves if no acceptable college has a vacant place). This assures that a student who is rejected in some round has an equal chance of being accepted by a college that ranks lower in his preference list as when he or she had listed the college at the top of the list. This is exactly why the Boston mechanism fails to deliver a stable matching: Despite the fact that both student 2 and college *A* would prefer being matched to each other, student 2 is rejected at college *A* in round 2 because *A* is already matched to 1 from round 1.

The *college-proposing DAA* is exactly the same except that offers are made on behalf of colleges. As colleges may have multiple seats, an offer will be issued in every round for every remaining seat of a college. Since the college-proposing DAA is just the “mirror image” of the student-proposing DDA, it also results in a stable matching, but not necessarily the same one.

Let us illustrate how the DAA works using the same preferences and student rankings as in the example. Table 2a presents the student-proposing DAA and Table 2b the college proposing DAA. Since in the college-proposing DAA college places are tentatively assigned to students in each round, the columns in Table 2b represent students and the entries are the places. New proposals in each round are indicated in bold, and rejections in that round are in parentheses as before. The last round represents the final matching.

Colleges	*A*	*B*	*C*	*C*
Round 1	**1**	**(2),3**	**4**	
Round 2	(1),**2**	3	4	
Round 3	2	**1**,(3)	4	
Round 4	2	1	4	**3**

Table 2a: The student-proposing deferred-acceptance algorithm

Students	1	2	3	4
Round 1	***C***	***A*,(*C*)**		***B***
Round 2	*C*	*A*	***C***	*B*

Table 2b: The college-proposing deferred-acceptance algorithm

The two mechanisms result in different, stable matchings. These final matchings are marked—with a plus sign (student-proposing) and circle (college-proposing)—within the student and college preference orderings below.

$P_1 = A, B^{+}, C^{\circ}$	$P_2 = B, A^{+\circ}, C$	$P_A = 2^{+\circ}, 1, 3, 4$	$P_B = 4^{\circ}, 1^{+}, 3, 2$
$P_3 = B, C^{+\circ}, A$	$P_4 = C^{+}, B^{\circ}, A$	$P_C = 1^{\circ}, 2, 3^{+\circ}, 4^{+}$	

Table 3: Matchings under the student-proposing DAA, indicated by $^{+}$*,*
and the college-proposing DAA, indicated by °.

Table 3 shows that *all* students weakly prefer their assignment under the student-proposing DAA to their assignment under the college-proposing DAA, and the converse is true for the colleges. As a matter of fact, this is always the case:

Theorem 2*: Using truthful preferences as input, the outcome under the student-proposing (college-proposing) DAA is weakly preferred by all students (colleges) to any other stable matching.*

Theorem 2 states that there is agreement between all students (all colleges) as to which stable matching is the best for students (colleges), and further, that the DAA produces this outcome. A further result says that students and colleges have exactly opposite rankings for different stable matchings, so that the matching that is best for students (the *student-optimal stable matching*) is the least-preferred by colleges among all the stable matchings, and vice versa, the *college-optimal stable matching* is the worst stable matching for students.

4. Strategy-proof Mechanisms

To implement the DAA in practice, a clearinghouse must rely on the *submitted* preferences as inputs. The results above then hold with respect to these submitted preferences. A critically important question is therefore whether students and colleges have an interest to reveal their true preferences to the mechanism. The answer to this question is a mixed bag: truthful reporting on the side of the students is possible, but no mechanism exists where both students and colleges have an incentive to submit truthful reports.[8]

Theorem 3: *In the student proposing DAA all students have a dominant strategy of reporting their true preferences.*

Theorem 4: *In the college admissions model no strategy-proof, stable matching mechanism exists.*

Let us confirm theorem 4 for a particular mechanism that we have studied: the student-proposing DAA. Assume college *C truncates* its preference list by stating its ranking as P_C = 1, 2, 3 (i.e., listing student 4 as unacceptable), and that all other participants report truthfully. With these misrepresented preferences, the student proposing DAA proceeds according to Table 4. In this example, this simple and

[8] Note that theorem 4 requires truthful reporting to be a dominant strategy for any possible matching problem. If one restricts attention to a particular subclass of problems, a positive answer may be obtained. An important example is the problems with large number of participants, where strategy-proof mechanisms are often possible (see Roth and Peranson 1999, Pathak and Kojima 2009 and Azevedo and Leshno 2011).

arguably realistic manipulation by college *C* not only benefits *C* but is actually sufficient to induce the college-optimal stable matching.[9]

	Colleges	<u>*A*</u>	<u>*B*</u>	<u>*C*</u>	<u>*C*</u>
Round 1		**1**	**(2),3**	**(4)**	
Round 2		(1),**2**	(3),**4**		
Round 3		2	4	**1**	**3**

Table 4: Student-proposing DAA under manipulation by college C

The reader may verify that even as college *C* manipulates the mechanism, it remains optimal for every student to report their preference truthfully, in accordance with Theorem 3.

5. Matching and School Choice

As a result of Theorem 4 it is in general impossible to devise a mechanism that results in stable matchings *and* is strategy proof for all agents. There is, however, a hopeful application of the theory when one side of the match does not behave strategically.

The *school choice model* is identical to the college admission model except that it treats (public) schools as non-strategic players. Schools are assumed to be endowed with *priorities* over students, reflecting pertinent legislation or regulations. For example, the schools may offer priority to disabled or disadvantaged students, or to students who live within walking distance. These priorities are formally treated by the mechanism as preferences but are, unlike preferences in the college admissions model, are regulated and known to the clearinghouse and therefore cannot be mis-represented by the school.

If schools are non-strategic actors, then the student-proposing DAA is a strategy-proof, stable mechanism.

In this model schools are not treated as having preferences; the priorities accorded to students can even be viewed as a *right* that a student can, in principle, forfeit or trade, if it makes her better off. Therefore, a matching should be evaluated by the extent to which it satisfies students' preferences—irrespective of whether school priorities are met.

Consider for example the stable matching resulting from the student-proposing DAA (displayed in the last line of Table 2). This matching is stable by construction but is not Pareto optimal for students: students 1 and 2 would be better off if they swapped

[9] A possible manipulation under the college-proposing DAA is for student 1 to delete college *C* from his preference list. This unilateral deviation not only benefits student 1 but induces the student optimal stable matching.

schools, leaving the other students unaffected.[10] In this environment an alternative mechanism, the *Top Trading Cycle* mechanism (TTC), may be most advisable. In the TTC, starting from some initial allocation, students swap places if it is in their mutual interest to do so, up to the point where no beneficial exchanges are possible. Thus, TTC generates a Pareto optimal matching, and it is also strategy-proof for students. However, TTC does not generate stable outcomes, making it problematic to use in the college admissions context where colleges have preferences over the entering class.

6. Conclusions

Matching theory has led to many important real-life applications, which, in turn, have influenced the development of the theory. Apart from the applications discussed in the Introduction, these include assigning students to courses at a given university (Sönmez and Ünver 2010), allocating housing to new immigrants and dorms to college students (Abdulkadiroglu and Sönmez 1999, Chen and Sönmez 2002), and the creation of paired kidney exchanges in the U.S. (Roth, Sönmez, and Ünver 2005, 2007) and in Australia, where a paired kidney exchange program informed by this research has been in place since 2010.[11] There are many other settings, where matching theory could be advantageously applied. In Australia, these include improved admission procedures for childcare centres, kindergarten, public schools and universities.

Current research in the matching literature focuses on how the expected welfare of the participants depends on the different matching mechanisms being used, on the informational constraints the participants face, and on what are appropriate solution concepts given these constraints.

References

Abdulkadiroglu, Atila, Pathak, Parag, and Alvin E. Roth, (2005). “The New York City high school match,” *American Economic Review, Papers and Proceedings* 95(2): 364–7.

Abdulkadiroglu, Atila and Tayfun Sönmez (1999): “House Allocation with Existing Tenants,” *Journal of Economic Theory* 88: 233–60.

Abdulkadiroglu, Atila and Tayfun Sönmez (2003): “School Choice: A Mechanism Design Approach,” *American Economic Review* 93(3), 729–47.

[10] Recall that the matching resulting from the student-proposing DAA is student-optimal among all *stable* matchings, not among all matchings.

[11] In a paired kidney exchange a living donor whose kidneys are incompatible with the intended recipient is matched with other incompatible donor–patient pair(s) in order to achieve compatible transplantations. Cadaveric exchange programs between jurisdictions have been in place for many years.

Artemov, Georgy, Bland, James, Feldmann, Sven, Halstead, Imogen, and Simon Loertscher (2011): “The Australian Tertiary Match: Stylized Facts and Open Issues,” mimeo, University of Melbourne.

Azevedo, Eduardo and Jacob Leshno (2011): “The college admissions problem with a continuum of students,” mimeo, Harvard University.

Chen, Yan and Tayfun Sönmez (2002): “Improving Efficiency of On-Campus Housing: An Experimental Study,” *American Economic Review* 92(5): 1669–86.

Department of Education (UK), School Admission Code, 2007 and 2010.

Gale David and Lloyd Shapley (1962): “College admissions and the stability of marriage,” *The American Mathematical Monthly* 69: 9–15.

Klein, Joel (2011): “The Failure of American Schools,” *The Atlantic Monthly*, June. (http://www.theatlantic.com/magazine/print/2011/06/the-failure-of-american-schools/8497, accessed Aug. 1 2011)

Loertscher, Simon and Tom Wilkening (2011): “Auctions and Economic Design,” *Australian Economic Review* 44(3): 347–354.

McKinney, C. Nicholas, Niederle, Muriel and Alvin E. Roth (2005): “The collapse of a medical labor clearinghouse (and why such failures are rare),” *American Economic Review* 95(3): 878–89.

Pathak, Parag and Fuhito Kojima (2009): “Incentives and Stability in Large Two-Sided Matching Markets,” *American Economic Review* 99: 608–27.

Roth, Alvin E. (1985): “The college admissions problem is not equivalent to the marriage problem,” *Journal of Economic Theory* 36: 277–88.

Roth, Alvin E. (2002): “The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics,” *Econometrica* 70(4): 1341–78.

Roth, Alvin E. (2008): “What Have We Learned From Market Design?” *Economic Journal*, 118(527): 285–310.

Roth, Alvin E. and Elliott Peranson (1999) “The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design,” *American Economic Review* 89(4): 748–80.

Roth, Alvin E., Sönmez, Tayfun and Utku Ünver (2005): “Kidney Exchange,” *Quarterly Journal of Economics* 119(2): 457–88.

Roth, Alvin E., Sönmez, Tayfun and Utku Ünver (2007): “Efficient Kidney Exchange: Coincidence of Wants in Markets with Compatibility-Based Preferences”, *American Economic Review* 97(3): 828–51.

Roth, Alvin E. and Marilda Sotomayor (1990): *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*, Econometric Society Monographs.

Sönmez, Tayfun and Utku Ünver (2010): “Course Bidding at Business Schools”, *International Economic Review* 51(1): 99–123.