KUNGL.
VETENSKAPS-
AKADEMIEN
THE ROYAL SWEDISH ACADEMY OF SCIENCES

15 OCTOBER 2012

Scientific Background on the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2012

STABLE ALLOCATIONS AND THE PRACTICE OF MARKET DESIGN

compiled by the Economic Sciences Prize Committee of the Royal Swedish Academy of Sciences

BOX 50005 (LILLA FRESCATIVÄGEN 4 A), SE-104 05 STOCKHOLM, SWEDEN
TEL +46 8 673 95 00, FAX +46 8 15 56 70, INFO@KVA.SE ⋆ HTTP://KVA.SE

1 Introduction

Economists study how societies allocate resources. Some allocation problems are solved by the price system: high wages attract workers into a particular occupation, and high energy prices induce consumers to conserve energy. In many instances, however, using the price system would encounter legal and ethical objections. Consider, for instance, the allocation of public-school places to children, or the allocation of human organs to patients who need transplants. Furthermore, there are many markets where the price system operates but the traditional assumption of perfect competition is not even approximately satisfied. In particular, many goods are indivisible and heterogeneous, whereby the market for each type of good becomes very thin. How these thin markets allocate resources depends on the institutions that govern transactions.

This year's prizewinning work encompasses a theoretical framework for analyzing resource allocation, as well as empirical studies and actual redesign of real-world institutions such as labor-market clearinghouses and school admissions procedures. The foundations for the theoretical framework were laid in 1962, when David Gale and Lloyd Shapley published a mathematical inquiry into a certain class of allocation problems. They considered a model with two sets of agents, for example workers and firms, that must be paired with each other. If a particular worker is hired by employer A, but this worker would have preferred employer B, who would also have liked to hire this worker (but did not), then there are unexploited gains from trade. If employer B had hired this worker, both of them would have been better off. Gale and Shapley defined a pairing to be *stable* if no such unexploited gains from trade exist. In an ideal market, where workers and firms have unrestricted time and ability to make deals, the outcome would always be stable. Of course, real-world markets may differ from this ideal in important ways. But Gale and Shapley discovered a "deferred-acceptance" procedure which is easy to understand and always leads to a stable outcome. The procedure specifies how agents on one side of the market (e.g., the employers) make offers to those on the other side, who accept or reject these offers according to certain rules.

The empirical relevance of this theoretical framework was recognized by Alvin Roth. In a study published in 1984, Roth found that the U.S. market for new doctors had historically suffered from a series of market failures, but a centralized clearinghouse had improved the situation by implementing a procedure essentially equivalent to Gale and Shapley's deferred-acceptance process. Roth's 1984 article clarified the tasks that markets perform, and showed how the concept of stability provides an organizing principle which helps us understand why markets sometimes work well, and why they sometimes fail to operate properly.

Subsequently, Roth and his colleagues used this framework, in combination with empirical studies, controlled laboratory experiments and computer simulations, to examine the functioning of other markets. Their research has not only illuminated how these markets operate, but has also proved useful in designing institutions that help markets function better, often by implementing a version or extension of the Gale-Shapley procedure. This has led to the emergence of a new and vigorous branch of economics known as market design. Note that in this context the term "market" does not presuppose the existence of a price system. Indeed, monetary transfers are ruled out in many important applications.

The work that is rewarded this year uses tools from both non-cooperative and cooperative game theory. Non-cooperative game theory was the subject of the 1994 Prize to John Harsanyi, John Nash and Reinhard Selten, and the 2005 Prize to Robert Aumann and Thomas Schelling. The starting point for a non-cooperative analysis is a detailed description of a strategic problem faced by individual decision makers. In contrast, cooperative game theory studies how groups ("coalitions") of individuals can further their own interests by working together. The starting point for a cooperative analysis is therefore a description of what each coalition can achieve. The person chiefly responsible for the development of cooperative game theory is Lloyd Shapley.

In many ways, the cooperative and non-cooperative approaches complement each other. Two properties of key importance for market design are *stability*, which encourages groups to voluntarily participate in the market, and *incentive compatibility*, which discourages strategic manipulation of the market. The notion of stability is derived from cooperative game theory, while incentive compatibility comes from the theory of mechanism design, a branch of non-cooperative game theory which was the subject of the 2007 Prize to Leonid Hurwicz, Eric Maskin and Roger Myerson.

Controlled laboratory experiments are frequently used in the field of market design. Vernon Smith shared the 2002 Prize for his work in experimental economics. Alvin Roth is another major contributor in this area.

The combination of game theory, empirical observations and controlled experiments has led to the development of an empirical science with many important practical applications. Evidence from the actual implementation of newly designed or redesigned institutions creates an important interplay and feedback effect: the discovery of a practical problem in implementation may trigger theoretical elaboration, new experiments, and finally changes in a design. Although these components form an integrated whole, we describe them separately, starting with some basic theoretical concepts. We introduce the idea of *stability* in Section 2. Then we describe some models of *matching markets* in Section 3, with emphasis on the Gale-Shapley deferred-acceptance procedure. In Section 4, we review how Alvin Roth recognized the real-world relevance of the theory. Some real-world cases of market design are outlined in Section 5. In Section 6, we note other important contributions of the two laureates. Section 7 concludes.

2 Theory I: Stability

Gale and Shapley (1962) studied stable allocations in the context of a specific model which will be described in Section 3. But first we will consider the idea of stability from the more general perspective of cooperative game theory.

2.1 Coalitional games with transferable utility

In this section we introduce some basic definitions from cooperative game theory.[1] Consider a set $N = \{1, 2, ..., n\}$ of n individuals (or "players"), for example, traders in a market. A group of individuals who cooperate with each other are said to form a *coalition*. A *game in coalitional form with transferable utility* specifies, for each coalition $S \subseteq N$, its "worth" $v(S)$. The worth is an economic surplus (a sum of money) that coalition S can generate using its own resources. If coalition S forms, then its members can split the surplus $v(S)$ in any way they want, and each member's utility equals her share of the surplus; we call this "transferable utility". The function v is called the *characteristic function*. Two special coalitions are the singleton coalition $\{i\}$ consisting only of player $i \in N$, and the grand coalition N consisting of all players.

Cooperative game theory studies the incentives of individuals to form coalitions, given that any potential conflicts of interest *within* a coalition

[1] The formal apparatus of cooperative game theory was introduced in von Neumann and Morgenstern's (1944) classical work.

can be solved by binding agreements. These agreements induce the coalition members to take actions that maximize the surplus of the coalition, and this maximized surplus is what the coalition is worth. A difficulty arises, however, if the surplus also depends on actions taken by non-members. In this case, the worth of a coalition can be determined in a consistent way by assuming that the non-members try to maximize their own payoffs (Huang and Sjöström, 2003, Kóczy, 2007).

In games with transferable utility, it is assumed that the players can freely transfer utility among themselves, in effect by making side-payments. But in some environments, side-payments are constrained and utility is not (perfectly) transferable. For example, in the National Resident Matching Program discussed below, wages are fixed before the market opens (Roth, 1984a). In other situations, such as donations of human organs, side-payments are considered "repugnant" (Roth, 2007). Cooperative game theory can handle such situations, as it is very well developed for general non-transferable utility games.

2.2 Stability and the core

Let x_i denote individual i's payoff, and let $x = (x_1, x_2, ...x_n)$ denote the payoff vector. If the members of some coalition S can use their own resources to make themselves better off, then we say that coalition S can *improve upon* x, or *block* x. When utility is transferable, coalition S can improve upon x if

$$\sum_{i \in S} x_i < v(S). \tag{1}$$

Indeed, if inequality (1) holds, then S can produce $v(S)$ and distribute this surplus so as to make all its members strictly better off than they are under x. The allocation x is then unstable.

An allocation is said to be *stable* if it cannot be improved upon by any coalition.[2] Thus, with transferable utility, the payoff vector x is stable if

$$\sum_{i \in S} x_i \geq v(S)$$

for every coalition $S \subseteq N$. The set of all stable payoff vectors is called the *core*.

[2] Stability has various definitions in the literature. Throughout this document we refer solely to stability against any possible coalitional deviation.

Although we have introduced stability in the context of transferable utility games, the definition extends in a straightforward way to general non-transferable utility games. In general, an allocation is stable if no coalition can improve upon it. That is, no coalition, by using its own resources, can bring about an outcome that all its members prefer.[3]

The idea of stability in cooperative game theory corresponds to the idea of Nash equilibrium in non-cooperative game theory. In non-cooperative game theory, a Nash equilibrium is a situation such that no individual can deviate and make herself better off. In cooperative game theory, a stable allocation is a situation such that no coalition can deviate and make its members better off. From an economic point of view, stability formalizes an important aspect of idealized frictionless marketplaces. If individuals have unlimited time and ability to strike deals with each other, then the outcome must be stable, or else some coalition would have an incentive to form and make its members better off. This basic idea is due to Edgeworth (1881), and is implicit in von Neumann and Morgenstern's (1944) analysis of stable set solutions. D.B. Gillies (1953a,b, 1959) and Shapley (1953c, 1955) were the first to explicitly consider the core as an independent solution concept. Laboratory experiments, where subjects must reach an agreement without any formalized procedure for making and accepting proposals, have provided support for the prediction that the final agreement will belong to the core (Berl, McKelvey, Ordeshook and Winer, 1976).

The following example shows how stable allocations are identified and that the core (i.e., the set of stable allocations) is sometimes quite large.

Example 1 *A partnership consisting of one senior partner (Mary) and two junior partners (Peter and Paul) generates earnings of 135. If Mary leaves the partnership, she can earn 50 on her own:* $v(\{Mary\}) = 50$*. Any junior partner can earn 10 on his own:* $v(\{Peter\}) = v(\{Paul\}) = 10$*. Mary and one junior partner together can earn 90, so* $v(\{Mary, Peter\}) =$

[3] In non-transferable utility games there can be a distinction between weak and strong improvement. The most common definition states that a coalition can improve upon an allocation if all its members can be made strictly better off. However, the results of Roth and Postlewaite (1977) suggest that it can sometimes be reasonable to use a weaker requirement: a coalition can improve upon an allocation if some members can be made strictly better off while no member is made strictly worse off. If improvement is defined in this weaker sense, then some coalition members may be indifferent with respect to participating in the coalition, but they are still assumed to participate.

$v(\{Mary, Paul\}) = 90$. *The two juniors together can earn 25, so that* $v(\{Peter, Paul\}) = 25$. *The grand coalition is worth 135 and utility is transferable, so they are free to divide up the 135 in any way they want. What is the maximum and minimum payoff Mary can get in a stable allocation?*

Mary must get at least 50, and each junior partner must get at least 10, to induce them to participate. Thus, stability requires

$$x_{Mary} \geq 50, \quad x_{Peter} \geq 10, \quad x_{Paul} \geq 10.$$

Two-player coalitions must also be taken into account: the two junior partners can improve on the allocation if they together get strictly less than 25, while a coalition of Mary and one junior partner can improve if they together get strictly less than 90. Thus, stability also requires

$$x_{Peter} + x_{Paul} \geq 25, \quad x_{Mary} + x_{Peter} \geq 90, \quad x_{Mary} + x_{Paul} \geq 90. \tag{2}$$

These inequalities, together with the partnership's budget restriction $x_{Mary} + x_{Peter} + x_{Paul} = 135$, *imply that Mary's minimum payoff is 50, and the maximum is 110.*[4]

2.3 Do stable allocations always exist?

Example 1 shows that the core may be quite large. In other instances, the situation may be quite the opposite and the core may even be empty. To illustrate this, suppose Example 1 is modified so that the surplus generated by the grand coalition is only 101. This yields the budget restriction $x_{Mary} + x_{Peter} + x_{Paul} = 101$. But if we add the three inequalities in (2), which must still be satisfied, we find that stability requires $2\left(x_{Mary} + x_{Peter} + x_{Paul}\right) \geq 205$. Thus, the surplus of 101 is too small to allow a stable allocation. In general, if there is not enough surplus available, it may be impossible to divide it up in a stable way. Bondareva (1963) and Shapley (1967) independently derived an exact formula for how much surplus must be available in order for the core to be non-empty in games with transferable utility. Their result was extended to games without transferable utility by Scarf (1967) and Billera

[4] The most important *single-valued* solution concept in cooperative game theory is the *Shapley value* (Shapley 1953a). Mary's Shapley value is 80, the midpoint of the interval $[50, 110]$ and arguably a "reasonable compromise".

(1970). Shapley (1971) showed that the core is always non-empty if the game is *convex* (in the sense that the value of a player's marginal contribution to a coalition is increased if other players join the coalition).[5]

2.4 Core and competitive equilibrium

Edgeworth (1881) was the first to argue that if some traders are not satisfied with what they receive on the market, then they can recontract, i.e., withdraw from the market and trade among themselves (not necessarily at prevailing market prices). The *contract curve* is the set of outcomes that cannot be destabilized by recontracting. As Shubik (1959) noted, Edgeworth's contract curve corresponds to the core of the economy. Edgeworth conjectured that in markets with sufficiently many traders, the contract curve would approximately equal the competitive equilibrium, and he verified this conjecture for the special case of two goods and two types of traders. Debreu and Scarf (1963) verified Edgeworth's conjecture under more general assumptions: if the economy is replicated so the number of traders of each type becomes very large, then the core approximately equals the set of competitive equilibria (see also Shubik, 1959). Thus, without having to specify the precise rules that govern trade, the core provides a key theoretical foundation for competitive equilibrium.

But many environments differ considerably from the perfectly competitive benchmark. Examples include collective-choice problems, such as choosing the level of a public good, and the matching markets which will be described in the next section. In the non-cooperative approach to such problems, institutions are analyzed in detail, and a solution concept such as Nash equilibrium is applied. The cooperative approach, on the other hand, can make predictions which are independent of the fine details of institutions. Specifically, if agents have unrestricted contracting ability, then the final outcome must be stable, for any unstable outcome will be overturned by some coalition that can improve upon it. We will now describe how Shapley and his colleagues applied this idea to various models of matching.

[5] Example 1 is an example of a convex game.

3 Theory II: Matching Markets

In many markets, goods are private but indivisible and heterogeneous, and the traditional assumption of perfect competition cannot be maintained. Important examples include markets for skilled labor. Since no two workers have exactly the same characteristics, the market for each particular bundle of labor services can be quite thin. In such markets, the participants must be appropriately *matched* in order to trade with each other.

3.1 Two-sided matching

Consider a market with two disjoint sets of agents — such as buyers and sellers, workers and firms, or students and schools — that must be matched with each other in order to carry out transactions. Such *two-sided matching markets* were studied by Gale and Shapley (1962). They ruled out side-payments: wages (and other match characteristics) are not subject to negotiation.

Stable matchings To be specific, suppose one side of the market consists of medical students and the other of medical departments. Each department needs one intern and each medical student wants one internship. A matching is an assignment of internships to applicants. Naturally, the students have preferences over departments, and the departments have preferences over students. We assume for convenience that preferences are strict (i.e., no ties). A matching is said to be *unacceptable* to an agent if it is worse than remaining unmatched.

In general, a matching is stable if no coalition can improve upon it. In this particular model, a stable matching must satisfy the following two conditions: (i) no agent finds the matching unacceptable, and (ii) no department-student pair would prefer to be matched with each other, rather than staying with their current matches. Condition (i) is an *individual rationality* condition and condition (ii) is *pairwise stability.* The two conditions imply that neither any singleton coalition, nor any department-student pair, can improve on the matching. (These are the only coalitions we need to consider in this model.)[6]

[6] Gale and Shapley (1962) defined a matching to be stable if no coalition consisting of one agent from each side of the market could improve on it (i.e., pairwise stability). Given the special structure of their model, this was equivalent to finding a matching in the core.

The Gale-Shapley algorithm Gale and Shapley (1962) devised a *deferred-acceptance algorithm* for finding a stable matching. Agents on one side of the market, say the medical departments, make offers to agents on the other side, the medical students. Each student reviews the proposals she receives, holds on to the one she prefers (assuming it is acceptable), and rejects the rest. A crucial aspect of this algorithm is that desirable offers are not immediately accepted, but simply held on to: deferred acceptance. Any department whose offer is rejected can make a new offer to a different student. The procedure continues until no department wishes to make another offer, at which time the students finally accept the proposals they hold.

In this process, each department starts by making its first offer to its top-ranked applicant, i.e., the medical student it would most like to have as an intern. If the offer is rejected, it then makes an offer to the applicant it ranks as number two, etc. Thus, during the operation of the algorithm, the department's expectations are lowered as it makes offers to students further and further down its preference ordering. (Of course, no offers are made to unacceptable applicants.) Conversely, since students always hold on to the most desirable offer they have received, and as offers cannot be withdrawn, each student's satisfaction is monotonically increasing during the operation of the algorithm. When the departments' decreased expectations have become consistent with the students' increased aspirations, the algorithm stops.

Example 2 *Four medical students (1, 2, 3 and 4) apply for internships in four medical departments: surgery (S), oncology (O), dermatology (D) and pediatrics (P). All matches are considered acceptable (i.e., better than remaining unmatched). The students have the following preference orderings over internships:*

$$\begin{array}{ll} 1: & S \succ O \succ D \succ P \\ 2: & S \succ D \succ O \succ P \\ 3: & S \succ O \succ P \succ D \\ 4: & D \succ P \succ O \succ S \end{array}$$

Thus, surgery is the most desirable internship, ranked first by three of the students. Each medical department needs one intern. They have the following preference orderings over students:

$$\begin{array}{ll} S: & 4 \succ 3 \succ 2 \succ 1 \\ O: & 4 \succ 1 \succ 3 \succ 2 \\ D: & 1 \succ 2 \succ 4 \succ 3 \\ P: & 2 \succ 1 \succ 4 \succ 3 \end{array}$$

Internships are allocated using the Gale-Shapley algorithm, with the departments making proposals to the students. Each department makes its first offer to its top-ranked applicant: student 1 gets an internship offer from D, student 2 gets one from P, and student 4 gets offers from S and O. Student 4 prefers O to S, so she holds on to the offer from O and rejects the offer from S. In the second round, S offers an internship to student 3. Now each student holds an internship, and the algorithm stops. The final assignment is:

$$1 \to D, \quad 2 \to P, \quad 3 \to S, \quad 4 \to O.$$

Gale and Shapley (1962) proved that the deferred-acceptance algorithm is stable, i.e., it always produces a stable matching. To see this, note that in Example 2, the algorithm allocates student 2 to her least preferred department, P, the only one to make her an offer. Now note that departments D, S and O have been assigned interns they think are preferable to student 2 – this must be the case, otherwise they would have offered 2 an internship before making offers to their assigned interns. Thus, even if they could replace their assigned interns with student 2, they would not want to do so. By this argument, any department which a student prefers to her assignment will not prefer her to its assigned intern, so the match is pairwise stable. Individual rationality holds trivially in Example 2, since there are no unacceptable matches, but it would hold in general as well, because students would reject all unacceptable offers, and departments would never make offers to unacceptable applicants.

The case where each department wants one intern corresponds to Gale and Shapley's (1962) "marriage" model. The case where departments may want more than one intern is their "college admissions" model. Gale and Shapley (1962) showed how the results for the marriage model generalize to the college admissions model. In particular, a version of the deferred-acceptance algorithm produces stable matchings even if departments want to hire more than one intern.[7]

The algorithm provides an existence proof for this type of two-sided matching problem: since it always terminates at a stable matching, a stable matching exists. In fact, more than one stable matching typically exists. Gale and Shapley (1962) showed that interests are polarized in the sense that

[7] In the college admissions model, Gale and Shapley (1962) did not formally specify the employers' preferences over different sets of employees, but this was done in later work (see Roth, 1985).

different stable outcomes favor one or the other side of the market.[8] This leads to a delicate issue: for whose benefit is the algorithm operated?

Who gains the most? In Example 2, the final assignment favors the medical departments more than the students.[9] In general, the *employer-proposing* version of the algorithm, where employers propose matches as in Example 2, produces an *employer-optimal* stable matching: all employers agree it is the best of all possible stable matchings, but all applicants agree it is the worst. The symmetric *applicant-proposing* version of the algorithm instead leads to an *applicant-optimal* stable matching (which all applicants agree is the best but all employers agree is the worst). This illustrates how the applicants' interests are opposed to those of the employers, and how stable institutions can be designed to systematically favor one side of the market.

Example 3 *The preferences are as in Example 2, but now the students have the initiative and make the proposals. Students 1, 2 and 3 start by making proposals to S, while student 4 makes a proposal to D. Since S prefers student 3, it rejects 1 and 2. In the second round, 1 makes a proposal to O and 2 makes a proposal to D. Since D prefers 2 to 4, it rejects 4. In the third round, 4 proposes to P. Now each department has an intern, and the algorithm stops. The final assignment is:*

$$1 \to O, \quad 2 \to D, \quad 3 \to S, \quad 4 \to P.$$

It can be checked that the students strictly prefer this assignment to the assignment in Example 2, except for student 3 who is indifferent (she is assigned to S in both cases). The departments are strictly worse off than in Example 2, except for S which gets student 3 in either case.

A "social planner" could conceivably reject both the applicant-optimal and employer-optimal stable matchings in favor of a stable matching that satisfies some fairness criterion, or perhaps some version of majority rule

[8] This insight follows from the more general fact that the set of stable matchings has the mathematical structure of a lattice (Knuth, 1976).

[9] Departments P, D and O get their most preferred intern, while S gets the intern it ranks second. The only student who is assigned to her favorite department is student 3. Students 1 and 4 are allocated to departments they rank third, while student 2 is assigned to the department she considers the worst.

(Gärdenfors, 1975). In practice, however, designs have tended to favor the applicants. In the context of college admissions, Gale and Shapley (1962) argued in favor of applicant-optimality based on the philosophy that colleges exist for the sake of the students, not the other way around.

Incentive compatibility Can the Gale-Shapley algorithm help participants in real-world markets find stable matchings? An answer to this question requires a non-cooperative analysis, that is, a detailed analysis of the rules that govern the matching process and the incentives for strategic behavior, to which we now turn.

Above, the deferred-acceptance algorithm was explained as a decentralized procedure of applications, offers, rejections and acceptances. But in practice, the algorithm is run by a clearinghouse in a centralized fashion. Each applicant submits her preference ordering, i.e., her personal ranking of the employers from most to least preferred. The employers submit their preferences over the applicants. Based on these submitted preferences, the clearinghouse goes through the steps of the algorithm. In the language of mechanism design theory, the clearinghouse runs a *revelation mechanism,* a kind of virtual market which does not suffer from the problems experienced by some real-world markets (as discussed later, these include unraveling and congestion). The revelation mechanism induces a simultaneous-move game, where all participants submit their preference rankings, given a full understanding of how the algorithm maps the set of submitted rankings into an allocation. This simultaneous-move game can be analyzed using non-cooperative game theory.

A revelation mechanism is (dominant strategy) *incentive compatible* if truth-telling is a dominant strategy, so that the participants always find it optimal to submit their *true* preference orderings. The employer-proposing algorithm – viewed as a revelation mechanism – is incentive compatible for the employers: no employer, or even coalition of employers, can benefit by misrepresenting their preferences (Dubins and Freedman, 1981, Roth, 1982a).[10] However, the mechanism is not incentive compatible for the applicants. To see this, consider the employer-proposing algorithm of Example 2. Suppose all participants are truthful except student 4, who submits $D \succ P \succ S \succ O$,

[10] More precisely, no coalition of employers can improve in the strong sense that every member is made strictly better off (see Roth and Sotomayor, 1990, Chapter 4, for a discussion of the robustness of this result).

which is a manipulation or strategic misrepresentation of her true preferences $D \succ P \succ O \succ S$. The final matching will be the one that the applicant-proposing algorithm produced in Example 3, so student 4's manipulation makes her strictly better off.[11] This proves that truth-telling is not a dominant strategy for the applicants. Indeed, Roth (1982a) proved that *no* stable matching mechanism exists for which stating the true preferences is a dominant strategy for every agent. However, notice that despite student 4's manipulation, the final matching is stable under the true preferences. Moreover, it is an undominated Nash equilibrium outcome. This illustrates a general fact about the Gale-Shapley algorithm, proved by Roth (1984b): all undominated Nash equilibrium outcomes of the preference manipulation game are stable for the true preferences.[12]

The usefulness of Roth's (1984b) result is limited by the fact that it may be difficult for applicants to identify their best responses, as required by the definition of Nash equilibrium. For example, if student 4 knows that the other applicants are truthful but not what their true preferences are, then student 4 will not be able to foresee the events outlined in Footnote 11. Therefore, she cannot be sure that this particular manipulation is profitable. This argument suggests that in large and diverse markets, where participants have very limited information about the preferences of others, the scope for strategic manipulation may be quite limited. Roth and Rothblum (1999) verify that when an applicant's information is sufficiently limited, she cannot gain by submitting a preference ordering which reverses her true ordering of two employers. However, it may be profitable for her to pretend that some acceptable employers are unacceptable.

[11] Consider how Example 2 would be different if student 4's preferences were $D \succ P \succ S \succ O$. Then, when student 4 receives simultaneous offers from O and S, she rejects O. In the second round, O would make an offer to student 1 who holds an offer from D but prefers O and therefore rejects D. In the third round, D makes an offer to 2 who holds an offer from P but prefers D and therefore rejects P. In the fourth round, P makes an offer to 1 who holds an offer from O and rejects P. In the fifth round, P makes an offer to 4 who holds an offer from S and now rejects S. In the sixth round, S makes an offer to 3. The algorithm then stops; the final assignment is

$$1 \to O, \quad 2 \to D, \quad 3 \to S, \quad 4 \to P.$$

[12] If we believe coalitions can jointly manipulate their reports, we can restrict attention to undominated Nash equilibria that are strong or rematching proof (Ma, 1995). Roth's (1984b) result applies to these refinements as well.

Roth and Peranson (1999) used computer simulations with randomly generated data, as well as data from the National Resident Matching Program, to study the gains from strategic manipulation of a deferred-acceptance algorithm. Their results suggested that in large markets, very few agents on either side of the market could benefit by manipulating the algorithm. Subsequent literature has clarified exactly how the gains from strategic manipulation vanish in large markets (Immorlica and Mahdian, 2005, Kojima and Pathak, 2009).

A related question is the following: if participants have incomplete information, does there exist a Bayesian-Nash equilibrium (not necessarily truthtelling) such that the outcome is always stable for the true preferences? Roth (1989) proved that this cannot be true for *any* mechanism, assuming both sides of the market behave strategically.[13] However, the applicant-proposing deferred-acceptance mechanism is incentive compatible for the applicants, so if the employers do not behave strategically, then truthtelling is a Bayesian-Nash equilibrium which produces a stable matching.[14] Even if both sides of the market are strategic, the lack of incentive-compatibility is less serious in large markets where, as Roth and Peranson (1999) discovered, the potential gains from strategic manipulation are limited. Under certain conditions, truthful reporting by both sides of the market is an approximate equilibrium for the applicant-proposing deferred-acceptance mechanism in a sufficiently large market (Kojima and Pathak, 2009).

Adjustable prices and wages Shapley and Shubik (1971) considered a transferable-utility version of the Gale-Shapley model called the *assignment game*. When employers are matched with workers, transferable utility means that match-specific wages are endogenously adjusted to clear the market.

Shapley and Shubik (1971) showed that the core of the assignment game is non-empty, and that competition for matches puts strict limits on the set of core allocations. With transferable utility, any core allocation must involve a matching which maximizes total surplus. Generically, this matching is

[13] This negative result applies to all mechanisms, not just revelation mechanisms. However, Roth's (1989) proof relies on the revelation principle, which states that without loss of generality, we can restrict attention to incentive compatible revelation mechanisms.

[14] The revelation mechanism which selects the applicant-optimal stable matching is (dominant strategy) incentive compatible for the applicants also in the "college admissions" model, where employers make multiple hires (Roth, 1985a).

unique. However, wages are not in general uniquely determined, thereby creating a polarization of interests similar to the Gale-Shapley model. Employer-optimal and applicant-optimal stable allocations exist and are characterized by the lowest and highest possible market-clearing wages. The core of the assignment game captures a notion of free competition reminiscent of traditional competitive analysis. In fact, in this model there is an exact correspondence between core and competitive equilibria.

Shapley and Shubik (1971) did not provide an algorithm for reaching stable allocations when utility is transferable, but Crawford and Knoer (1981) showed that a generalized Gale-Shapley algorithm accomplishes this task (see also Demange, Gale and Sotomayor, 1986). In the employer-proposing version, each employer starts by making a low salary offer to its favorite applicant. Any applicant who receives more than one offer holds on to the most desirable offer and rejects the rest. Employers whose offers are rejected continue to make offers, either by raising the salary offer to the same applicant, or by making an offer to a new applicant. This process always leads to the employer-optimal stable allocation. Kelso and Crawford (1982) and Roth (1984c, 1985b) generalized these results still further. Specifically, Kelso and Crawford (1982) introduced the assumption that employers, who have preferences over sets of workers, consider workers to be substitutes. Under this assumption, an employer-proposing deferred-acceptance algorithm still produces the employer-optimal stable allocation, while an applicant-proposing version produces the applicant-optimal stable allocation (Kelso and Crawford, 1982, Roth, 1984c).[15]

When side-payments are available, the deferred-acceptance algorithm can be regarded as a simultaneous multi-object English auction, where no object is allocated until bidding stops on all objects. As long as the objects for sale are substitutes, this process yields the bidder-optimal core allocation. Roth and Sotomayor (1990, Part III) discuss the link between matching and auctions, a link which was further strengthened by Hatfield and Milgrom (2005). Varian (2007) and Edelman, Ostrovsky and Schwarz (2007) showed that the assignment game provides a natural framework for analyzing auctions used by Internet search engines to sell space for advertisements.

[15] If employers want to hire more than one worker, the employer-proposing algorithm lets them make multiple offers at each stage. To see why substitutability is required for this algorithm to work as intended, note that this assumption guarantees that if an offer is rejected by a worker, the employer does not want to withdraw any previous offers made to other workers.

3.2 One-sided matching

Shapley and Scarf (1974) studied a *one-sided* market, where a set of traders exchange indivisible objects (such as plots of land) without the ability to use side-payments. Each agent initially owns one object. Abdulkadiroğlu and Sönmez (1999) later generalized the model to allow for the possibility that some agents do not initially own any objects, while some objects have no initial owner.

Shapley and Scarf (1974) proved that the *top-trading cycle* algorithm, which they attributed to David Gale, always produces a stable allocation. The algorithm works as follows. Starting from the initial endowment, each agent indicates her most preferred object. This can be described in a "directed graph" indicating, for each agent, whose object this agent would prefer. There must exist at least one "cycle" in the directed graph, i.e., a set of agents who could all obtain their preferred choices by swapping among themselves. These swaps occur, and the corresponding agents and objects are removed from the market. The process is repeated with the remaining agents and objects, until all objects have been allocated. The algorithm is illustrated in the following example.

Example 4 *There are four agents, 1, 2, 3, and 4, and four objects, A, B, C, and D. The agents have the following preferences:*

$$\begin{array}{ll} 1: & A \succ B \succ C \succ D \\ 2: & B \succ A \succ D \succ C \\ 3: & A \succ B \succ D \succ C \\ 4: & D \succ C \succ A \succ B. \end{array}$$

Given two alternative initial endowment structures, Figure 1 indicates the implied preferences with arrows.

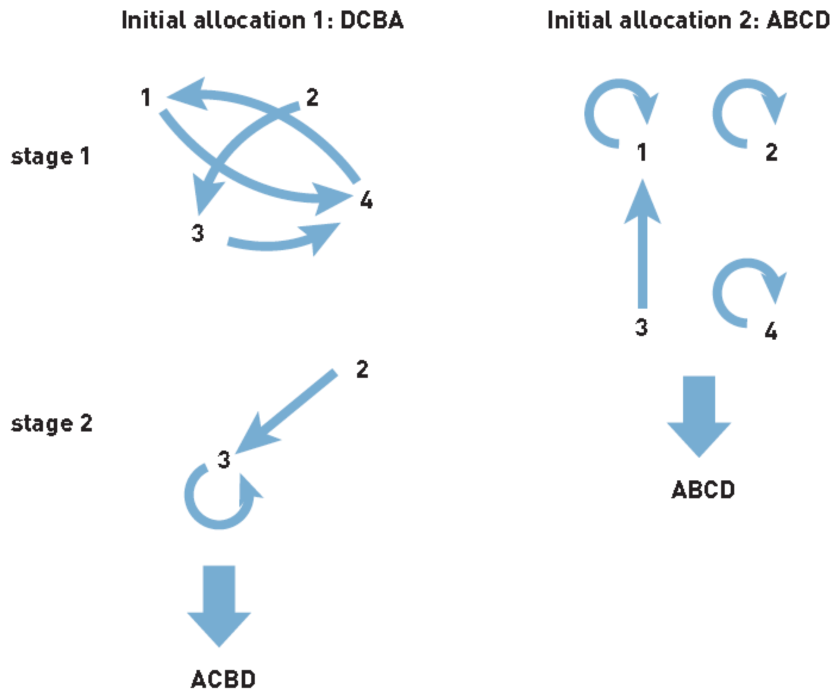

Figure 1: Top Trading Cycles for different endowment structures.

On the left-hand side of Figure 1 the initial allocation (endowment) is $DCBA$, that is, agent 1 owns object D, agent 2 owns object C, etc. In stage 1, agent 1 indicates that her favorite object is owned by 4, while 4 indicates that her favorite object is owned by 1. Thus, agents 1 and 4 form a cycle. They swap objects and are removed together with their objects D and A. Now agents 2 and 3 remain, with their endowments C and B. In the second stage, both 2 and 3 indicate that 3's object is their favorite among the two objects that remain. Therefore, the process terminates with the final allocation $ACBD$. This allocation is stable: no coalition of traders can reallocate their initial endowments to make all members better off. The right-hand side of the figure shows that, had the initial endowment been $ABCD$, then no trade would have occurred.

Roth and Postlewaite (1977) show that if preferences over objects are strict, and if stability is defined in terms of weak improvements (see Footnote 3), then for any given initial endowment there is a *unique* stable allocation. For example, if the initial endowment in Example 4 is $DCBA$, then the coalition consisting of agents 1 and 4 can obtain their favorite objects simply by swapping their endowments. Hence, any stable outcome must give A to 1 and D to 4. Object C cannot be given to agent 3, because she can block this by refusing to trade. Hence the unique stable allocation is $ACBD$. In contrast, when the initial endowment is $ABCD$, agents 1, 2 and 4 would block any outcome where they do not obtain their favorite objects (which they already own), so $ABCD$ is the only stable allocation.

The revelation mechanism that chooses the unique stable allocation, computed by the top-trading cycle algorithm from submitted preference orderings and given endowments, is dominant strategy incentive compatible for all participants (Roth, 1982b). In fact, this is the only revelation mechanism which is Pareto efficient, individually rational and incentive compatible (Ma, 1994).

Important real-world allocation problems have been formalized using Shapley and Scarf's (1974) model. One such problem concerns the allocation of human organs, which will be discussed in Section 5.3. Another such problem concerns the allocation of public-school places to children. In the school-choice problem, no "initial endowments" exist, although some students may be given priority at certain schools. Abdulkadiroğlu and Sönmez (2003) adapted the top-trading cycle to the school choice problem, but another approach is to incorporate the schools' preferences over students via the Gale-Shapley algorithm. This will be discussed in Section 5.2.

4 Evidence: Markets for Doctors

The work on stable allocations and stable algorithms was recognized as an important theoretical contribution in the 1960s and 1970s, but it was not until the early 1980s that its practical relevance was discovered. The key contribution is Roth (1984a), which documents the evolution of the market for new doctors in the U.S. and argues convincingly that a stable algorithm improved the functioning of the market. This work opened the door to Roth's participation in actual design, which began in the 1990s. Roth also conducted empirical studies of other medical markets, documenting and analyzing how several regions in the U.K. had adopted different algorithms (Roth, 1991a). These further strengthened the case for stable algorithms. The overall evidence provided by Roth was pivotal.

Centralized matching mechanisms, such as the one in the U.S. market for new doctors, have well-defined "rules of the game" known to both the participants themselves and the economists who study the market. Knowledge of these rules makes it possible to test game-theoretic predictions, in the field as well as in laboratory experiments. Moreover, the rules can be redesigned to improve the market functioning (see Section 5). Accordingly, these types of matching mechanisms have been studied in depth and are by now well understood. Other markets with clearly defined rules have also been the subject of intensive studies; the leading example is auction markets. In fact, matching and auction theory are closely linked, as mentioned above.

We begin this section by describing the U.S. market for new doctors, and then turn to the U.K. regional medical markets. We also consider how important evidence regarding the performance of matching algorithms have been generated using laboratory experiments.

4.1 The U.S. market for new doctors

Roth (1984a) studied the evolution of the U.S. market for new doctors. Students who graduate from medical schools in the U.S. are typically employed as residents (interns) at hospitals, where they comprise a significant part of the labor force. In the early twentieth century, the market for new doctors was largely decentralized. During the 1940s, competition for medical students forced hospitals to offer residencies (internships) increasingly early, sometimes several years before a student would graduate. This so-called *unraveling* had many negative consequences. Matches were made before students could produce evidence of how qualified they might become, and even before they knew what kind of medicine they would like to practice. The market also suffered from *congestion*: when an offer was rejected, it was often too late to make other offers. A congested market fails to clear, as not enough offers can be made in time to ensure mutually beneficial trades. To be able to make more offers, hospitals imposed strict deadlines which forced students to make decisions without knowing what other opportunities would later become available.

Following Roth's (1984a) study, similar problems of congestion and unraveling were found to plague many markets, including entry-level legal, business school and medical labor markets in the U.S., Canada and the U.K., the market for clinical psychology internships, dental and optometry residencies in the U.S., and the market for Japanese university graduates (Roth and Xing, 1994). When indivisible and heterogeneous goods are traded, as in these markets for skilled labor, offers must be made to specific individuals rather than "to the market". The problem of coordinating the timing of offers can cause a purely decentralized market to become congested and unravel, and the outcome is unlikely to be stable. Roth and Xing (1994) described how market institutions have been shaped by such failures, and explained their findings in a theoretical model (see also Roth and Xing, 1997).

In response to the failures of the U.S. market for new doctors, a centralized clearinghouse was introduced in the early 1950s. This institution is now called the National Resident Matching Program (NRMP). The NRMP matched doctors with hospitals using an algorithm which Roth (1984a) found to be essentially equivalent to Gale and Shapley's employer-proposing deferred-acceptance algorithm. Although participation was voluntary, essentially all residencies were allocated using this algorithm for several decades. Roth (1984a) argued that the success of the NRMP was due to the fact that its algorithm produced stable matchings. If the algorithm had produced unstable matchings, doctors and hospitals would have had an incentive to bypass the algorithm by forming preferred matches on the side (a doctor could simply contact her favorite hospitals to inquire whether they would be interested in hiring her).[16]

When a market is successfully designed, many agents are persuaded to participate, thereby creating a "thick" market with many trading opportunities. The way in which a lack of stability can create dissatisfaction and reduce participation rates is illustrated by Example 2. An unstable algorithm might assign student 1 to pediatrics. But if the dermatology department finds out that their top-ranked applicant has been assigned to a department she likes less than dermatology, they would have a legitimate reason for dissatisfaction. A stable algorithm would not allow this kind of situation. It is thus more likely to induce a high participation rate, thereby creating many opportunities for good matches which, in turn, induces an even higher participation rate. Roth and his colleagues have identified this virtuous cycle in a number of real-world markets, as well as in controlled laboratory experiments.

[16] As in the original Gale and Shapley (1962) model, the only issue considered by the NRMP is to find a matching. Salaries are determined by employers before residencies are allocated, so they are treated as exogenous to the matching process. Crawford (2008) argues that it would be quite feasible to introduce salary flexibility into the matching process by using a generalized deferred-acceptance algorithm of the type considered by Crawford and Knoer (1981). See also Bulow and Levin (2006).

4.2 Regional medical markets in the U.K.

Roth (1990, 1991a) observed that British regional medical markets suffered from the same kinds of problems in the 1960s that had afflicted the U.S. medical market in the 1940s. Each region introduced its own matching algorithm. Some were stable, others were not (see Table 1). Specifically, the clearinghouses in Edinburgh and Cardiff implemented algorithms which were essentially equivalent to the deferred-acceptance algorithm, and these operated successfully for decades. On the other hand, Birmingham, Newcastle and Sheffield quickly abandoned their unstable algorithms.

TABLE I

STABLE AND UNSTABLE (CENTRALIZED) MECHANISMS

Market	Stable	Still in use (halted unraveling)
American medical markets		
NRMP	yes	yes (new design in '98)
Medical Specialties	yes	yes (about 30 markets)
British Regional Medical Markets		
Edinburgh ('69)	yes	yes
Cardiff	yes	yes
Birmingham	no	no
Edinburgh ('67)	no	no
Newcastle	no	no
Sheffield	no	no
Cambridge	no	yes
London Hospital	no	yes
Other healthcare markets		
Dental Residencies	yes	yes
Osteopaths (<'94)	no	no
Osteopaths (≥'94)	yes	yes
Pharmacists	yes	yes
Other markets and matching processes		
Canadian Lawyers	yes	yes (except in British Columbia since 1996)
Sororities	yes (at equilibrium)	yes

Table 1. Reproduced from Roth (2002, Table 1).

4.3 Experimental evidence

The empirical evidence seems to support the hypothesis that stable matching algorithms can prevent market failure (Roth and Xing, 1994, Roth, 2002, Niederle, Roth, and Sönmez, 2008). However, many conditions influence the success or failure of market institutions. The objective of market designers is to isolate the role of the mechanism itself, and compare the performance of different mechanisms under the same conditions. But this is difficult to accomplish in the real world. For example, British regional medical markets might differ in numerous ways that cannot be controlled by an economist.

Accordingly, market designers have turned to controlled laboratory experiments to evaluate and compare the performance of mechanisms.

Kagel and Roth (2000) compared the stable (deferred-acceptance) algorithm used in Edinburgh and Cardiff with the unstable "priority-matching" algorithm used in Newcastle.[17] In their experiment, a centralized matching mechanism was made available to the subjects, but they could choose to match in a decentralized way, without using the mechanism. When the mechanism used priority matching, the experimental market tended to unravel, and many matches were made outside the mechanism. The deferred-acceptance mechanism did not suffer from the same kind of unraveling. This provided experimental evidence in favor of Roth's hypothesis that the matching algorithm itself and, in particular, its stability, contribute importantly to the functioning of the market.

Two regions, Cambridge and London Hospital, presented an anomaly for Roth's hypothesis. In these regions, the matching algorithms solved a linear programming problem which did not produce stable outcomes. Yet, these markets did not appear to unravel, and the unstable mechanisms remained in use (see Table 1). In experiments, the linear programming mechanisms seem to perform no better than priority matching, which suggests that conditions specific to Cambridge and London Hospital, rather than the intrinsic properties of their matching algorithms, may have prevented unraveling there (Ünver, 2005). Roth (1991a) argued that these markets are in fact so small that social pressures may prevent unraveling.

In one U.S. medical labor market (for gastroenterology), a stable algorithm was abandoned after a shock to the demand and supply of positions. McKinney, Niederle and Roth's (2005) laboratory experiments suggested that this market failed mainly because, while employers knew about the exogenous shock, the applicants did not. Shocks that both sides of the market knew about did not seem to cause the same problems. This suggested that the algorithm would fail only under very special conditions. Roth and M. Niederle

[17] The unstable algorithms in Birmingham and Sheffield used a similar method as the Newcastle algorithm. In priority-matching, an applicant's ranking of an employer and the employer's ranking of the applicant jointly determine the applicant's "priority" at that employer. Thus, the highest priority matches are those where the two parties rank each other first. Apart from being unstable, such methods are far from incentive-compatible; deciding whom to rank first is a difficult strategic problem. A similar problem is discussed below with regard to the "Boston mechanism" (which itself is a kind of priority-matching algorithm).

helped the American Gastroenterology Association reintroduce a deferred-acceptance matching algorithm in 2006. Niederle, Proctor and Roth (2008) describe early evidence in favor of the reintroduced matching mechanism.

5 Market Design

The theory outlined in Sections 2 and 3 and the empirical evidence discussed in Section 4 allow us to understand the functions that markets perform, the conditions required for them to be performed successfully, and what can go wrong if these conditions fail to hold. We now consider how these insights have been used to improve market functioning. Of course, real-world markets experience idiosyncratic complications that are absent in theoretical models. Real-world institutions have to be robust to agents who make mistakes, do not understand the rules, have different prior beliefs, etc. They should also be appropriate to the historical and social context and, needless to say, respect legal and ethical constraints on how transactions may be organized. Given the constraints of history and prevailing social norms, small-scale incremental changes to existing institutions might be preferred to complete reorganizations.

This section deals with three sets of real-world applications: first, the market for doctors in the U.S.; second, the design of school-admission procedures; and third, a case of one-sided matching (kidney exchange).

5.1 Redesigning the market for new doctors

As described in Section 4.1, Roth's work illuminated why the older, and more decentralized, system had failed, and why the new (deferred-acceptance) algorithm adopted by the NRMP performed so much better. However, as described by Roth and Peranson (1999), the changing structure of the medical labor market caused new complexities to arise which led the NRMP to modify its algorithm. By the 1960s a growing number of married couples graduated from medical school, and they often tried to bypass the algorithm by contacting hospitals directly. A couple can be regarded as a composite agent who wants two jobs in the same geographic location, and whose preferences therefore violate the assumption of substitutability. Roth (1984a) proved that in a market where some agents are couples, there may not exist any stable matching. The design of matching and auction mechanisms in the presence of complementarities is an important topic in the recent literature.

A need for reform: the Roth-Peranson algorithm In the 1990s, the very legitimacy of the NRMP algorithm was challenged. Specifically, it was argued that what was primarily an employer-proposing algorithm favored hospitals at the expense of students.

> Medical-school personnel responsible for advising students about the job market began to report that many students believed the NRMP did not function in the best interest of students, and that students were discussing the possibility of different kinds of strategic behavior (Roth and Peranson, 1999, p. 749).

The basic theory of two-sided matching, outlined in Section 3.1, shows that the employer-proposing algorithm is not incentive compatible for the applicants, i.e., it is theoretically possible for them to benefit by strategically manipulating or "gaming" it. However, the applicant-proposing version is incentive compatible for the applicants. The complexity of the medical labor market, with complementarities involving both applicants and positions, means that the basic theory cannot be applied directly. However, computational experiments show that the theory can provide useful advice even in this complex environment (Roth and Peranson, 1999). Overall, there seemed to be strong reasons to switch to an applicant-proposing algorithm.

In 1995, Alvin Roth was hired by the Board of Directors of the NRMP to direct the design of a new algorithm. The goal of the design was "to construct an algorithm that would produce stable matchings as favorable as possible to applicants, while meeting the specific constraints of the medical market" (Roth and Peranson, 1999, p. 751). The new algorithm, designed by Roth and Elliott Peranson, is an applicant-proposing algorithm modified to accommodate couples: potential instabilities caused by the presence of couples are resolved sequentially, following the instability-chaining algorithm of Roth and Vande Vate (1990). Computer simulations suggested that the Roth-Peranson algorithm would turn out to be somewhat better for the applicants than the old NRMP employer-proposing algorithm (Roth and Peranson, 1999). The simulations also revealed that, in practice, it would essentially be impossible to gain by strategic manipulation of the new algorithm (Roth, 2002).

Since the NRMP adopted the new algorithm in 1997, over 20,000 doctors per year have been matched by it (Roth and Peranson, 1999, Roth, 2002). The same design has also been adopted by entry-level labor markets in other professions (see Table 2). The empirical evidence suggests that the outcome is stable despite the presence of couples.

Table 1: Labor markets that adopted the Roth-Peranson clearinghouse design after 1998 (and date of first use of a centralized clearinghouse of some sort):

- Postdoctoral Dental Residencies in the United States
    - Oral and Maxillofacial Surgery (1985)
    - General Practice Residency (1986)
    - Advanced Education in General Dentistry (1986)
    - Pediatric Dentistry (1989)
    - Orthodontics (1996)
- Psychology Internships in the United States and Canada (1999)
- Neuropsychology Residencies in the U.S. and Canada (2001)
- Osteopathic Internships in the United States (before 1995)
- Pharmacy Practice Residencies in the United States (before 1994)
- Articling Positions with Law Firms in Alberta, Canada (1993)
- Medical Residencies in the United States (NRMP) (1952)
- Medical Residencies in Canada (CaRMS) (before 1970)
- Specialty Matching Services (SMS/NRMP):
    - Abdominal Transplant Surgery (2005)
    - Child & Adolescent Psychiatry (1995)
    - Colon & Rectal Surgery (1984)
    - Combined Musculoskeletal Matching Program (CMMP)
        - Hand Surgery (1990)
    - Medical Specialties Matching Program (MSMP)
        - Cardiovascular Disease (1986)
        - Gastroenterology (1986-1999; rejoined in 2006)
        - Hematology (2006)
        - Hematology/Oncology (2006)
        - Infectious Disease (1986-1990; rejoined in 1994)
        - Oncology (2006)
        - Pulmonary and Critical Medicine (1986)
        - Rheumatology (2005)
    - Minimally Invasive and Gastrointestinal Surgery (2003)
    - Obstetrics/Gynecology
        - Reproductive Endocrinology (1991)
        - Gynecologic Oncology (1993)
        - Maternal-Fetal Medicine (1994)
        - Female Pelvic Medicine & Reconstructive Surgery (2001)
    - Ophthalmic Plastic & Reconstructive Surgery (1991)
    - Pediatric Cardiology (1999)
    - Pediatric Critical Care Medicine (2000)
    - Pediatric Emergency Medicine (1994)
    - Pediatric Hematology/Oncology (2001)
    - Pediatric Rheumatology (2004)
    - Pediatric Surgery (1992)
    - Primary Care Sports Medicine (1994)
    - Radiology
        - Interventional Radiology (2002)
        - Neuroradiology (2001)
        - Pediatric Radiology (2003)
    - Surgical Critical Care (2004)
    - Thoracic Surgery (1988)
    - Vascular Surgery (1988)

Table 2. Reproduced from Roth (2008a, Table 1).

5.2 School admission

Many students simply attend the single school where they live. Sometimes, however, students have potential access to many schools. A matching of students with schools should take into account the preferences of the students and their parents, as well as other important concerns (about segregation, for example). Should schools also be considered strategic agents with preferences over students? Some schools might prefer students with great attendance records, others might be mainly concerned about grades, etc. If the schools, as well as the applicants, are regarded as strategic agents, then a two-sided matching problem ensues.

In the theoretical models of Balinski and Sönmez (1999) and Abdulkadiroğlu and Sönmez (2003), classroom slots are allocated among a set of applicants, but the schools are not considered strategic agents. Insights from two-sided matching models are still helpful, however. An applicant may be given high priority at some particular school (for example, if she lives close to the school, has a sibling who attends the school, or has a high score on a centralized exam). In this case, the school can be said to have preferences over students, in the sense that higher-priority students are more preferred. Stability then captures the idea that if student 1 has higher priority than student 2 at school S, and student 2 attends school S, then student 1 must attend a school that she likes at least as well as S (perhaps S itself).

An important difference between this model and the two-sided model of Section 3.1 is that the priority ranking of students can be based on objectively verifiable criteria. In such instances, the problem of incentive-compatibility does not necessarily arise on the part of the schools. Moreover, the priority orderings do not have the same welfare implications as preference orderings usually have. These arguments suggest using the applicant-proposing deferred-acceptance algorithm, which is not only fully incentive compatible for the applicants, but also applicant optimal (i.e., every applicant prefers it to any other stable match).[18] The New York City public high schools started using a version of the deferred-acceptance algorithm in 2003, and the Boston public school system started using a different version in 2005 (Roth, 2008b).[19]

[18] For experimental evidence on school-choice mechanisms, see Chen and Sönmez (2006), Featherstone and Niederle (2011) and Pais and Pintér (2008).

[19] The problems the market designers faced in these two markets were somewhat different. In New York, the school-choice system is in effect a two-sided market where the

Prior to 2003, applicants to New York City public high schools were asked to rank their five most preferred schools and these preference lists were sent to the schools. The schools then decided which students to admit, reject, or wait-list. The process was repeated in two more rounds. Students who had not been assigned to any school after the third round were assigned via an administrative process. This process suffered from congestion, as the applicants did not have sufficient opportunities to express their preferences, and the schools did not have enough opportunities to make offers. The market failed to clear: about 30,000 students per year ended up, via the administrative process, at a school for which they had not expressed any preference (Abdulkadiroğlu, Pathak and Roth, 2005).

Moreover, the process was not incentive-compatible. Schools were more likely to admit students who ranked them as number one. Therefore, if a student was unlikely to be admitted to her favorite school, her best strategy would be to list a more realistic option as her "first choice". In 2003, Roth and his colleagues A. Abdulkadiroğlu and P.A. Pathak helped re-design this admissions process. The new system uses an applicant-proposing deferred-acceptance algorithm, modified to accommodate regulations and customs of New York City. This algorithm is incentive compatible for the applicants, i.e., it is optimal for them to report their preferences truthfully, and congestion is eliminated. During the first year of the new system, only about 3,000 students had to be matched with schools for which they had not expressed a preference, a 90 percent reduction compared to previous years.

Prior to 2005, the Boston Public School system (BPS) used a clearing-house algorithm known as the "Boston mechanism". This type of algorithm first tries to match as many applicants as possible with their first-choice school, then tries to match the remaining applicants with their second-choice school and so on (Abdulkadiroğlu, Pathak, Roth and Sönmez, 2005). Evidently, if an applicant's favorite school is very difficult to get accepted at, with this type of mechanism it is best to list a less popular school as the first choice. This presented the applicants with a vexing strategic situation: to game the system optimally, they had to identify which schools were realistic options for them. Applicants who simply reported their true preferences suffered unnecessarily poor outcomes. Even if, miraculously, everyone had found a best-response strategy, every Nash equilibrium would have been

schools are active players. In Boston, the schools are passive and priorities are determined centrally.

Pareto dominated by the truthful equilibrium of the applicant-proposing deferred-acceptance mechanism (Ergin and Sönmez, 2006). Roth and his colleagues, A. Abdulkadiroğlu, P.A. Pathak and T. Sönmez, were asked to provide advice on the design of a new BPS clearinghouse algorithm. In 2005, an applicant-proposing deferred-acceptance algorithm was adopted. Since it is incentive-compatible for the applicants, the need for strategizing is eliminated. Other school systems in the U.S. have followed New York and Boston by adopting similar algorithms; a recent example is the Denver public school system.[20]

5.3 Kidney exchange

In important real-world situations, side-payments are ruled out on legal and ethical grounds. For example, in most countries it is illegal to exchange human organs, such as kidneys, for money. Organs have to be assigned to patients who need transplants by some other method. Some patients may have a willing kidney donor. For example, a husband may be willing to donate a kidney to his wife. A direct donation is sometimes ruled out for medical reasons, such as incompatibility of blood types. Still, if patient A has a willing (but incompatible) donor A', and patient B has a willing (but incompatible) donor B', then if A is compatible with B' and B with A', an exchange is possible: A' donates to B and B' to A. Such bilateral kidney exchanges were performed in the 1990s, although they were rare.

Roth, Sönmez and Ünver (2004) noted the similarity between kidney exchange and the Shapley-Scarf one-sided matching model described in Section 3.2, especially the version due to Abdulkadiroğlu and Sönmez (1999). One important difference is that, while all objects in the Shapley-Scarf model can be assigned simultaneously, some kidney patients must be assigned to a waiting list, in the hope that suitable kidneys become available in the

[20] Many different kinds of matching procedures are used in various parts of the world. Braun, Dwenger, Kübler and Westkamp (2012) describe the two-part procedure that the German central clearinghouse uses to allocate admission to university medical studies and related subjects. In the first part, 20 percent of all available university seats are reserved for applicants with very good grades, and 20 percent for those with the longest waiting time since completing high school. These seats are allocated using the Boston mechanism. In the second part, all remaining seats are allocated using a university-proposing deferred-acceptance algorithm. The authors use laboratory experiments to study the incentives to strategically manipulate this two-part procedure. More information on matching algorithms in Europe can be found at http://www.matching-in-practice.eu/.

future.[21] Roth, Sönmez and Ünver (2004) adapted the top-trading cycle algorithm to allow for waiting-list options. The doctors indicate the most preferred kidney, or the waiting-list option, for each patient. If there is a cycle, kidneys are exchanged accordingly. For example, three patient-donor pairs (A, A'), (B, B') and (C, C') may form a cycle, resulting in a three-way exchange (A gets a kidney from B', B from C', and C from A'). The rules allow for "chains" where, for example, A gets a kidney from B' while B is assigned a high priority in the waiting list (and another patient can receive a kidney from A'). Roth, Sönmez and Ünver (2004) constructed efficient and incentive-compatible chain selection rules.

A bilateral exchange between (A, A') and (B, B') requires a "double coincidence of wants": A' must have what B needs while B' must have what A needs. A clearinghouse with a database of patient-donor pairs that implements more complex multilateral exchanges can increase market thickness, i.e., raise the number of possible transplants. This is especially important if many highly sensitized patients are compatible with only a small number of donors (Ashlagi and Roth, 2012). However, complex multilateral exchanges may not be feasible due to logistical constraints. Roth, Sönmez and Ünver (2005b) showed how efficient outcomes with good incentive properties can be found in computationally efficient ways when only bilateral exchanges are feasible. But significant gains can be achieved with exchanges involving three patient-donor pairs (Saidman et al., 2006, Roth, Sönmez, and Ünver, 2007).

A number of regional kidney exchange programs in the U.S. have in fact moved towards more complex exchanges. The New England Program for Kidney Exchange, founded by Roth, Sönmez and Ünver, in collaboration with Drs. Frank Delmonico and Susan Saidman, was among the early pioneers (Roth, Sönmez and Ünver, 2005a). Recently, interest has focused on long chains involving "altruistic donors", who want to donate a kidney but have no particular patient in mind. Such chains suffer less from logistical constraints, because the transplants do not need to be conducted simultaneously (Roth et al., 2006).

This work on kidney exchange highlights an important aspect of market design. Specific applications often uncover novel problems, such as the NRMP's couples problem, the priorities of school choice, or the waiting-list

[21] The problem of kidney exchange is inherently more dynamic than the applications discussed in Sections 5.1 and 5.2. Whereas residencies and public-school places can be allocated once per year, there is no such obvious timing of kidney exchanges. This has led to theoretical work on the optimal timing of transactions (Ünver, 2010).

and logistical problems of kidney exchange. These new problems stimulate new theoretical research, which in turn leads to new applications, etc. Alvin Roth has made significant contributions to all parts of this iterative process.

6 Other Contributions

6.1 Lloyd Shapley

In non-cooperative game theory, Shapley's contributions include a number of innovative studies of dynamic games. Aumann and Shapley's (1976) *perfect folk theorem* shows that any feasible payoff vector (where each player gets at least the minimum amount he can guarantee for himself) can be supported as a strategic equilibrium payoff of a repeated game involving very patient players. The theory of repeated games was generalized by Shapley (1953b), who introduced the important notion of a *stochastic game,* where the actions chosen in one period may change the game to be played in the future. This has led to an extensive literature (e.g., Mertens and Neyman, 1981). Shapley (1964) showed that a certain class of learning dynamics may not converge to an equilibrium point, a result which has stimulated research on learning in games. Shapley and Shubik (1977) is an important study of strategic market games.

Lloyd Shapley is the most important researcher in the field of cooperative game theory. Shapley and Shubik (1969) characterized the class of transferable-utility market games, and showed that such games have non-empty cores. Shapley (1953a) introduced, and axiomatically characterized, the main single-valued solution concept for coalitional games with transferable utility, nowadays called the *Shapley value.* Shapley (1971) proved that for convex games, the Shapley value occupies a central position in the core. Harsanyi (1963) and Shapley (1969) extended the Shapley value to games without transferable utility.

The Shapley value has played a major role in the development of cooperative game theory, with a large variety of applications. Although originally intended as a prediction of what a player could expect to receive from a game, it is often given a normative interpretation as an equitable outcome, for example, when costs are allocated by some administrative procedure (e.g., Young, 1994). The book by Aumann and Shapley (1974) contains extensions of the major justifications, interpretations and computations of the Shapley value to games with infinitely many players. This work has important applications to problems of cost allocation (e.g., Billera, Heath and Raanan, 1978). The book also contains a version of Edgeworth's conjecture: in certain large markets, the Shapley value and the core both coincide with the competitive equilibrium allocation. The Shapley value for coalitional political games is known as the *Shapley-Shubik power index* (Shapley and Shubik, 1954). It has been used, in particular, to evaluate power shifts caused by changes in voting systems (e.g., Hosli, 1993).

6.2 Alvin Roth

The book by Roth and Sotomayor (1990) documents the state of two-sided matching theory three decades ago, including many key results due to Roth and coauthors. Among his other theoretical contributions, Roth (1977) characterized the Shapley value as a risk-neutral utility function defined on the space of coalitional games with transferable utility.

Roth (1991b) describes how laboratory experiments and field observations can interact with game theory, thereby establishing economics as a more satisfactory empirical science. Through his own laboratory experiments, Alvin Roth has greatly contributed to this research program. In a series of experiments, Roth and his coauthors tested the predictions of cooperative bargaining theory (Roth and Malouf, 1979, Roth, Malouf and Murnighan, 1981, Roth and Murnighan, 1982, Murnighan, Roth and Schoumaker, 1988). Cooperative bargaining models were found to correctly predict the qualitative effects of changes in risk aversion. These tests were facilitated by a device of Roth and Malouf (1979), who controlled for the subject's inherent risk-aversion by using lottery tickets as rewards. By varying the information given to a subject about another subject's payoffs, the experiments revealed the importance of focal-point effects and fairness concerns. A series of experiments by Ochs and Roth (1989) tested the predictions of non-cooperative bargaining models. This was followed by the important cross-cultural study of Roth, Prasnikar, Okuno-Fujiwara and Zamir (1991) which investigated bargaining behavior in four different countries.

Laboratory experiments often reveal that subjects change their behavior over time. Roth and Erev (1995) developed a reinforcement learning model, where players tend to repeat choices that produce good outcomes. This model turned out to be consistent with actual behavior in a number of experimental games. Slonim and Roth (1998) used this type of model to explain behavior in a simple non-cooperative bargaining game, while Erev and Roth (1998) showed that a reinforcement learning model can predict behavior ex ante (rather than merely explaining it ex post). This influential series of articles has shown that the explanatory and predictive power of game theory can be increased if realistic cognitive limitations are taken into account.

7 Conclusion

Lloyd Shapley has led the development of cooperative game theory. His work has not only strengthened its theoretical foundations, but also enhanced the theory's usefulness for applied work and policy making. In collaboration with D. Gale, H. Scarf and M. Shubik, he created the theory of matching markets. Launching the theory, Gale and Shapley (1962) expressed the hope that one day it would have practical applications. This hope has been fulfilled by the emerging literature on market design.

The work by Alvin Roth has enhanced our understanding of how markets work. Using empirical, experimental and theoretical methods, Roth and his coauthors, including A. Abdulkadiroğlu, P.A. Pathak, T. Sönmez and M.U. Ünver, have studied the institutions that improve market performance, thereby illuminating the need for stability and incentive compatibility. These contributions led directly to the successful redesign of a number of important real-world markets.

For further reading An elementary introduction to cooperative game theory can be found in Moulin (1995), while Shubik (1984) offers a more advanced treatment. Serrano's (2009) survey emphasizes the core and the Shapley value, while Maschler (1992) discusses alternative cooperative solution concepts. For introductions to matching theory, see Roth and Sotomayor (1990) or the original article by Gale and Shapley (1962). For general aspects of market design, see Roth (2002) and (2008b). Roth (2008a) discusses the history, theory and practical aspects of deferred-acceptance algorithms. Sönmez and Ünver (2011) provide a detailed technical survey of the design of matching markets. For the most recent developments in market design, see Alvin Roth's blog, http://marketdesigner.blogspot.com/.

References

[1] Abdulkadiroğlu, A. and T. Sönmez. 1999. House allocation with existing tenants. *Journal of Economic Theory* **88**: 233–260

[2] Abdulkadiroğlu, A. and T. Sönmez. 2003. School choice: A mechanism design approach. *American Economic Review* **93**: 729–47.

[3] Abdulkadiroğlu, A., P.A. Pathak and A.E. Roth. 2005. The New York City high school match. *American Economic Review* **95**: 364–67.

[4] Abdulkadiroğlu, A., P.A. Pathak, A.E. Roth and T. Sönmez. 2005. The Boston public school match. *American Economic Review* **95**: 368–71.

[5] Ashlagi, I. and A.E. Roth 2012. New challenges in multi-hospital kidney exchange. *American Economic Review* **102**: 354-359.

[6] Aumann, R.J. and L.S. Shapley. 1974. *Values of Non-atomic Games.* Princeton: Princeton University Press.

[7] Aumann, R.J. and L.S. Shapley. 1976. Long term competition - a game theoretic analysis. Mimeo.

[8] Balinski, M. and T. Sönmez. 1999. A tale of two mechanisms: student placement. *Journal of Economic Theory* **84**: 73-94.

[9] Berl, J.E., R.D. McKelvey, P.C. Ordeshook and M.D. Winer. 1976. An experimental test of the core in a simple n-person cooperative nonsidepayment game. *Journal of Conflict Resolution* **20:** 453-479.

[10] Billera, L. 1970. Some theorems on the core of an n-person game without side-payments. *SIAM Journal on Applied Mathematics* **18**: 567-579.

[11] Billera, L., D. Heath and J. Raanan. 1978. Internal telephone billing rates - A novel application of non-atomic game theory. *Operations Research* **26**: 956-965.

[12] Bondareva, O. N. 1963. Some applications of linear programming to the theory of cooperative games. *Problemy Kibernetiki* **10**: 119–39 [in Russian]. English translation in *Selected Russian Papers in Game Theory 1959–1965.* Princeton: Princeton University Press, 1968.

[13] Braun, S., N. Dwenger, D. Kübler and A. Westkamp. 2012. Implementing quotas in university admissions: An experimental analysis. Working paper, Max Planck Institute for Tax Law and Public Finance, Munich.

[14] Bulow, J. and J. Levin. 2006. Matching and price competition. *American Economic Review* **96**: 652-668.

[15] Chen, Y. and T. Sönmez. 2006. School choice: An experimental study. *Journal of Economic Theory* **127**: 2002-2031.

[16] Crawford V.P. 2008. The flexible-salary match: A proposal to increase the salary flexibility of the National Resident Matching Program. *Journal of Economic Behavior and Organization* **66**: 149–160.

[17] Crawford V.P. and E.M. Knoer. 1981. Job matching with heterogeneous firms and workers. *Econometrica* **49**: 437-450.

[18] Debreu, G. and H. Scarf 1963. A limit theorem on the core of an economy. *International Economic Review* **4**: 234–246.

[19] Demange, G., D. Gale and M. Sotomayor. 1986. Multi-item auctions. *Journal of Political Economy* **94**: 863-872.

[20] Dubins, L.E. and D.A. Freedman. 1981. Machiavelli and the Gale-Shapley algorithm. *American Mathematical Monthly* **88**: 485-494.

[21] Edelman, B., M. Ostrovsky and M. Schwarz. 2007. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *American Economic Review* **97**: 242–259.

[22] Edgeworth, F. Y. 1881. *Mathematical Psychics*. London: C. Kegan Paul & Co.

[23] Erev, I. and A.E. Roth. 1998. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review* **88**: 848-881.

[24] Ergin, H. and T. Sönmez. 2006. Games of school choice under the Boston mechanism. *Journal of Public Economics* **90**: 215-237.

[25] Featherstone, C. and M. Niederle. 2011. School choice mechanisms under incomplete information: An experimental investigation. Working Paper, Harvard Business School, Cambridge, MA.

[26] Gale, D. and L.S. Shapley. 1962. College admissions and the stability of marriage. *American Mathematical Monthly* **69**: 9–15.

[27] Gärdenfors, P. 1975. Match making: assignments based on bilateral preferences. *Behavioral Science* **20**: 166–173.

[28] Gillies, D.B. 1953a. Location of solutions. In *Report of an Informal Conference on the Theory of n-Person Games held at Princeton University March 20-21, 1953*: 11–12. Mimeo, Department of Mathematics, Princeton University.

[29] Gillies, D.B. 1953b. *Some Theorems on n-Person Games*. Ph. D. Thesis, Department of Mathematics, Princeton University (June 1953).

[30] Gillies, D.B. 1959. Solutions to general non-zero sum games. In *Contributions to the Theory of Games IV*. Edited by A.W. Tucker and R.D. Luce. Princeton, N.J.: Princeton University Press.

[31] Harsanyi, J.C. 1963. A simplified bargaining model for the n-person cooperative game. *International Economic Review* **4**: 194–220.

[32] Hatfield, J. and P. Milgrom. 2005. Matching with contracts. *American Economic Review* **95**: 913–35.

[33] Hosli, M. 1993. Admission of European Free Trade Association states to the European Community: effects on voting power in the European Community Council of Ministers. *International Organization* **47**: 629-643.

[34] Huang, C.Y. and T. Sjöström. 2003. Consistent solutions for cooperative games with externalities. *Games and Economic Behavior* **43**:196-213.

[35] Immorlica, N. and M. Mahdian. 2005. Marriage, honesty, and stability. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*: 53-62. Philadelphia: Society for Industrial and Applied Mathematics.

[36] Kagel, J.H. and A.E. Roth. 2000. The dynamics of reorganization in matching markets: A laboratory experiment motivated by a natural experiment. *Quarterly Journal of Economics* **115**: 201–35.

[37] Kelso, A.S. and V.P. Crawford. 1982. Job matching, coalition formation, and gross substitutes. *Econometrica* **50**: 1483–504.

[38] Knuth, D. 1976. *Mariages Stables*. Montréal: Les Presses de l'Université de Montréal.

[39] Kóczy, L. 2007. A recursive core for partition function form games. *Theory and Decision* **63**: 41-51.

[40] Kojima, F. and P.A. Pathak. 2009. Incentives and stability in large two-sided matching markets. *American Economic Review* **99**: 608–27.

[41] Ma, J. 1994. Strategy-proofness and the strict core in a market with indivisibilities. *International Journal of Game Theory* **23**: 75-83.

[42] Ma, J. 1995. Stable matchings and rematching-proof equilibria in a two-sided matching market. *Journal of Economic Theory* **66**: 352–369.

[43] Maschler, M. 1992. The bargaining set, kernel, and nucleolus. In *Handbook of Game Theory with Economic Applications*, Vol. 1. Edited by R. J. Aumann and S. Hart. Elsevier Science.

[44] McKinney, C.N., M. Niederle and A.E. Roth. 2005. The collapse of a medical labor clearinghouse (and why such failures are rare). *American Economic Review* **95**: 878-889.

[45] Mertens, J.-F. and A. Neyman. 1981. Stochastic games. *International Journal of Game Theory* **10**: 53-66.

[46] Moulin, H. 1995. *Cooperative Microeconomics*. Princeton University Press.

[47] Murnighan, J.K., A.E. Roth and F. Schoumaker. 1988. Risk aversion in bargaining: An experimental study. *Journal of Risk and Uncertainty* **1**: 101-124.

[48] Niederle, M., D. Proctor and A.E. Roth. 2008. The gastroenterology fellowship match: The first two years. *Gastroenterology* **135**: 344-346.

[49] Niederle, M., A.E. Roth and T. Sönmez. 2008. Matching and Market Design. In *New Palgrave Dictionary of Economics*, second edition. Edited by S. N. Durlauf and L. E. Blume. Basingstoke: Palgrave Macmillan.

[50] Ochs, J. and A.E. Roth. 1989. An experimental study of sequential bargaining. *American Economic Review* **79**: 355-384.

[51] Pais, J. and Á. Pintér. 2008. School choice and information: An experimental study on matching mechanisms. *Games and Economic Behavior* **64**: 303–328.

[52] Roth, A.E. 1977. The Shapley value as a von Neumann- Morgenstern utility. *Econometrica* **45**: 657-664.

[53] Roth, A.E. 1982a. The economics of matching: Stability and incentives. *Mathematics of Operations Research* **7**: 617-628.

[54] Roth, A.E. 1982b. Incentive compatibility in a market with indivisibilities. *Economics Letters* **9**: 127-132.

[55] Roth, A.E. 1984a. The evolution of the labor market for medical interns and residents: A case study in game theory. *Journal of Political Economy* **92**: 991–1016.

[56] Roth, A.E. 1984b. Misrepresentation and stability in the marriage problem. *Journal of Economic Theory* **34**: 383–387.

[57] Roth, A.E. 1984c. Stability and polarization of interests in job matching. *Econometrica* **52**: 47-57.

[58] Roth, A.E. 1985a. The college admissions problem is not equivalent to the marriage problem. *Journal of Economic Theory* **36**: 277–288

[59] Roth, A.E. 1985b. Conflict and coincidence of interest in job matching: Some new results and open questions. *Mathematics of Operations Research* **10**: 379-389

[60] Roth, A.E. 1989. Two-sided matching with incomplete information about others' preferences. *Games and Economic Behavior* **1**: 191–209.

[61] Roth, A.E. 1990. New physicians: A natural experiment in market organization. *Science* **250**: 1524–28.

[62] Roth, A.E. 1991a. A natural experiment in the organization of entry level labor markets: Regional markets for new physicians and surgeons in the U.K. *American Economic Review* **81**: 415–40.

[63] Roth, A.E. 1991b. Game theory as a part of empirical economics. *Economic Journal* **101**: 107-114

[64] Roth, A.E. 2002. The economist as engineer: Game theory, experimental economics and computation as tools of design economics. *Econometrica* **70**: 1341–78.

[65] Roth, A.E. 2007. Repugnance as a constraint on markets. *Journal of Economic Perspectives* **21**: 37-58.

[66] Roth, A.E. 2008a. Deferred acceptance algorithms: History, theory, practice, and open questions. *International Journal of Game Theory* **36**: 537–569.

[67] Roth, A.E. 2008b. What have we learned from market design? *Economic Journal* **118**: 285–310.

[68] Roth, A.E. and I. Erev. 1995. Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior* **8**: 164-212.

[69] Roth, A.E. and M. Malouf. 1979. Game-theoretic models and the role of information in bargaining. *Psychological Review* **86**: 574-594

[70] Roth, A.E. and J.K. Murnighan. 1982. The role of information in bargaining: An experimental study. *Econometrica* **50**: 1123-1142.

[71] 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**: 748–79.

[72] Roth, A.E. and A. Postlewaite. 1977. Weak versus strong domination in a market with indivisible goods. *Journal of Mathematical Economics* **4**: 131-137.

[73] Roth, A.E. and U. Rothblum. 1999. Truncation strategies in matching markets: In search of advice for participants. *Econometrica* **67**: 21-43.

[74] Roth, A.E. and M. Sotomayor. 1990. *Two-sided Matching: A Study in Game-theoretic Modeling and Analysis*. Econometric Society Monograph Series, Cambridge: Cambridge University Press.

[75] Roth, A.E. and J.H. Vande Vate. 1990. Random paths to stability in two-sided matching. *Econometrica* **58**: 1475-1480.

[76] Roth, A.E. and X. Xing. 1994. Jumping the gun: Imperfections and institutions related to the timing of market transactions. *American Economic Review* **84**: 992-1044.

[77] Roth, A.E. and X. Xing. 1997. Turnaround time and bottlenecks in market clearing: Decentralized matching in the market for clinical psychologists. *Journal of Political Economy* **105**: 284–329.

[78] Roth, A.E., M. Malouf and J.K. Murnighan. 1981. Sociological versus strategic factors in bargaining. *Journal of Economic Behavior and Organization* **2**: 153-177.

[79] Roth, A.E., T. Sönmez and M.U. Ünver. 2004. Kidney exchange. *Quarterly Journal of Economics* **119**: 457–88.

[80] Roth, A.E., T. Sönmez and M.U. Ünver. 2005a. A kidney exchange clearinghouse in New England. *American Economic Review* **95**: 376–80.

[81] Roth, A.E., T. Sönmez and M.U. Ünver. 2005b. Pairwise kidney exchange. *Journal of Economic Theory* **125**: 151–88.

[82] Roth, A.E., T. Sönmez and M.U. Ünver. 2007. Efficient kidney exchange: Coincidence of wants in markets with compatibility-based preferences. *American Economic Review* **97**: 828–51.

[83] Roth, A.E., V. Prasnikar, M. Okuno-Fujiwara and S. Zamir. 1991. Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: an experimental study. *American Economic Review* **81**: 1068-1095.

[84] Roth, A.E., T. Sönmez, M.U. Ünver, F.L. Delmonico and S.L. Saidman. 2006. Utilizing list exchange and undirected Good Samaritan donation through 'chain' paired kidney donations. *American Journal of Transplantation* **6**: 2694–705.

[85] Saidman, S.L., A.E. Roth, T. Sönmez, M.U. Ünver, and F.L. Delmonico. 2006. Increasing the opportunity of live kidney donation by matching for two and three way exchanges. *Transplantation* **81**: 773–82.

[86] Scarf, H. 1967. The core of an n-person game. *Econometrica* **35**: 50-69.

[87] Serrano, R. 2009. Cooperative games: core and Shapley value. In *Encyclopedia of Complexity and Systems Science*. Edited by R. Meyers. New York: Springer.

[88] Shapley, L.S. 1953a. A value for n-person games. In *Contributions to the Theory of Games*, Vol. 2. Edited by H. W. Kuhn and A. W. Tucker. Princeton, NJ: Princeton University Press.

[89] Shapley, L.S. 1953b. Stochastic games. *Proceedings of the National Academy of Sciences of the USA* **39**: 1095–1100.

[90] Shapley, L.S. 1953c. Open questions. In *Report of an Informal Conference on the Theory of n-Person Games held at Princeton University March 20-21, 1953*: 15. Mimeo, Department of Mathematics, Princeton University.

[91] Shapley, L.S. 1955. Markets as cooperative games. Rand Corporation Memorandum P-629, Santa Monica, CA.

[92] Shapley, L.S. 1964. Some topics in two-person games. In *Advances in Game Theory*. Edited by M. Dresher, L. S. Shapley, and A. W. Tucker. Princeton, NJ: Princeton University Press.

[93] Shapley, L.S. 1967. On balanced sets and cores. *Navel Research Logistics Quarterly* **9**: 45–8.

[94] Shapley, L.S. 1969. Utility comparison and the theory of games. In *La décision: Agrégation et dynamique des ordres de préférence*. Paris: Editions du CNRS.

[95] Shapley, L.S. 1971. Cores of convex games. *International Journal of Game Theory* **1**: 11-26.

[96] Shapley, L.S. and H. Scarf. 1974. On cores and indivisibility. *Journal of Mathematical Economics* **1**: 23–37.

[97] Shapley, L.S. and M. Shubik. 1954. A method of evaluating the distribution of power in a committee system. *American Political Science Review* **48**: 787-792

[98] Shapley, L.S. and M. Shubik. 1969. On market games. *Journal of Economic Theory* **1**: 9–25.

[99] Shapley, L.S. and M. Shubik. 1971. The assignment game I: The core. *International Journal of Game Theory* **1**: 111–30.

[100] Shapley, L.S. and M. Shubik. 1977. Trade using one commodity as a means of payment. *Journal of Political Economy* **85**: 937-968.

[101] Shubik, M. 1959. Edgeworth market games. In *Contributions to the Theory of Games IV*. Edited by R.D. Luce and A.W. Tucker. Princeton, N.J.: Princeton University Press.

[102] Shubik, M. 1984. *Game Theory in the Social Sciences*. MIT Press.

[103] Slonim, R. and A.E. Roth. 1998. Learning in high stakes ultimatum games: an experiment in the Slovak Republic. *Econometrica* **66**: 569-596.

[104] Sönmez, T. and M.U. Ünver. 2011. Matching, allocation, and exchange of discrete resources. In *Handbook of Social Economics*. Edited by J. Benhabib, A. Bisin, and M. Jackson. The Netherlands: North-Holland.

[105] Ünver, M.U. 2005. On the survival of some unstable two-sided matching mechanisms. *International Journal of Game Theory* **33**: 239–254.

[106] Ünver, M.U. 2010. Dynamic kidney exchange. *Review of Economic Studies* **77**: 372–414.

[107] Varian, H. 2007. Position auctions. *International Journal of Industrial Organization* **25**:1163–1178.

[108] von Neumann, J. and O. Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton, N.J.: Princeton University Press.

[109] Young, H.P., 1994. Cost allocation. In *Handbook of Game Theory with Economic Applications*, Vol. 2. Edited by R.J. Aumann and S. Hart. Amsterdam: Elsevier Science.