CONTRIBUTIONS to SCIENCE
OPEN ACCESS

CONTRIB SCI 11:103-112 (2015)
doi:10.2436/20.7010.01.218

RESEARCH REVIEWS
Institut d'Estudis Catalans, Barcelona, Catalonia
www.cat-science.cat

The theory of stable allocations and the practice of market design. The Nobel Prize in Economics 2012 for Alvin E. Roth and Lloyd S. Shapley

Jordi Massó[1,2]

[1]Department of Economy and of History of Economy, Autonomous University of Barcelona. [2]Barcelona Graduate School of Economics, Barcelona, Catalonia

Correspondence:
Jordi Massó
Dept. d'Economia i d'Història Econòmica
Facultat d'Economia i Empresa
Universitat Autònoma de Barcelona
08193 Bellaterra, Catalonia

E-mail: jordi.masso@uab.es

Summary. The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2012 was awarded jointly to Alvin E. Roth and Lloyd S. Shapley for their contributions to the theory of stable allocations and the practice of market design. The theory of stable allocations consists of a family of models that study assignment problems in which two disjoint sets of agents (or a set of agents and a set of objects) have to be matched. For example, men to women, workers to firms, students to schools, or patients to live donor kidneys. A matching is stable if no subset of agents can improved upon their proposed matches by rematching only among themselves. Stability is an essential property if matching is voluntary. The practice of market design consists of applying those two-sided matching models to specific assignment problems with the aim of proposing improvements on how they are solved. This paper presents a brief description of the academic career of the laureates and describes their contributions by presenting the most basic two-sided matching model and some of its market design applications, including the organization of a centralized system to propose kidney transplantations to use kidneys of live donors that are incompatible with their respective patients, the yearly assignment of North-American medical students to hospital internship programs, and children to schools in cities such as Boston and New York. [**Contrib Sci** 11(1): 103-112 (2015)]

The Laureates

Alvin Roth was born in New York city, NY, on December 18, 1951. He graduated from Columbia University in 1971 (when he was 19 years old!) with a degree in Operations Research. He obtained his Ph.D. in Operations Research from Stanford University in 1974 under the supervision of Robert B. Wilson. His first two jobs were at the Departments of Economics at the University of Illinois (from 1974 to 1982) and at the University of Pittsburgh (from 1982 to 1998). In 1998 he

Keywords: game theory · market design · stable allocations · mathematical economics · kidney transplantation

Fig. 1. Alvin E. Roth (left), and Lloyd S. Shapley (right), awarded with the Nobel Prize in Economics 2012.

moved to Harvard University with a joint appointment from the Economics Department and the Harvard Business School. He stayed there until the beginning of 2013 when he moved to the Economics Department at Stanford University. During this period at Harvard University he supervised a large group of Ph.D. students, most of whom work now at the best universities in the USA and Europe (Fig. 1).

Roth thesis was on von Neumann and Morgenstern stable sets. His research interests have been wide and moved very consistently to include axiomatic bargaining, experimental economics, learning in non-cooperative games, the theory of stable allocations in matching markets, and market design. According to the Royal Swedish Academy of Sciences the prize was awarded to him for his research on the last two areas, although he has made fundamental contributions in the other ones [5,12,17].[1]

Lloyd Shapley was born in Cambridge, MA, on June 2, 1923, and died in Tucson, AZ, on March 12, 2016. After serving in the Army Air Corps in Chengdu, China, during the WWII, he went to Harvard University, where he graduated in 1948 with a degree in mathematics. He obtained his Ph.D. in mathematics from Princeton University in 1953 under the supervision of Albert W. Tucker. He has had only two affiliations: at RAND Corporation (from 1954 to 1981) and at the Departments of Mathematics and Economics at the University of California, Los Angeles, since 1981 (Fig. 1).

Shapley's doctoral thesis was on additive and non-additive set functions. He has made fundamental contributions in all areas of game theory; for instance to the theory of the Core, the Shapley value, repeated and stochastic games, the potential of a game and the theory of stable allocations in matching markets. Many game theorists thought that the fact that Shapley had not been awarded the Nobel Prize in Economics yet was a sad omission. We are now pleased that this was corrected in 2012.

The awarding of the 2012 Nobel Prize to both Roth and Shapley may be seen as recognizing two complementary sides of a research: Shapley for his theoretical contributions to the theory of stable allocations in two-sided matching problems [7,22] and Roth for his applications of this theory to improve the functioning of institutions solving two-sided assignment real-life problems [15,18,20].[2] Roth and Shapley did not write jointly, but Roth has been closely following Shapley's research as his fourth paper [11] and his fourth book show [16].

David Gale (1921–2008) made also fundamental contributions to the theory of stable allocations and he might

[1] The first paper is still his most cited paper in SCOPUS.

[2] Roth and Sotomayor (1990) contains a masterful review of all matching literature from 1962 to 1990 and it is still the best gateway to the theory and applications of two-sided matching problems.

have been also been awarded with the prize if he had he been alive in 2012. He was born in New York city, NY, on December 13, 1921 and died in Berkeley (California) on March 7, 2008. He obtained his Ph.D. from Princeton University in 1949 under the supervision of Albert W. Tucker (1905–1995). He had two main affiliations, the first at Brown University (from 1954 to 1981) and the second at the University of California, Berkeley, (from 1965 to 2008). He made also relevant contributions to mathematical economics and game theory, and his work is still a very useful reference on the applications of linear programming to economics [6].

The theory of stable allocations and the practice of market design

Participants in some markets cannot be divided a priori between buyers and sellers. If the price of a good changes sufficiently, a participant can be a seller and a buyer in a few minutes of difference. Stocks are clear examples of goods exchanged in such markets. However, there are many other markets without this property: participants are either buyers or sellers, independently of the price of the good. Physical or legal characteristics of the participants make them to be in one, and only one side of the market. For instance, a university professor cannot become a university, even after a dramatic decline of the professors' wage, nor the university can become a university professor after its increase.

There are many two-sided assignment problems, not necessarily solved through markets, in which participants are divided a priori between two disjoint sets, for instance, men and women, workers and firms, and students and colleges. The assignment problem is precisely to match each participant in one of the two sets with a participant in the other set (or to remain unmatched) taking into account the preferences that each participant in one set has on the participants on the other set (plus the prospect of remaining unmatched). But the matching has to be bilateral: if a is matched with b, b is matched with a. Moreover, those problems have often two additional properties that distinguish them from conventional markets. First, the matching between two participants requires mutual agreement: if a chooses to be matched with b, a has to be chosen by b. Second, prices do not play any role to facilitate the matching and to resolve the potential disequilibrium of the mutual wills.

Two-sided matching models formalize the main characteristics of these assignment problems. Shapley contributed to the development of the earlier stages of this theory, specifically by proposing the notion of stability of an allocation as the relevant property whenever the assignment has to be voluntary [7]. An assignment (or a matching) between the two sets of participants is stable at a preference profile if: (a) all participants are either unmatched or matched with a participant that is strictly preferred to remaining unmatched and (b) there is no pair of participants that are not matched with each other but they would prefer to be so rather than staying with the partner proposed by the assignment.

Although Roth also has fundamental theoretical contributions to two-sided matching models he has been the founder and main contributor to market design. This area uses two-sided matching models and other tools to analyze practical assignment problems. It restricts the attention to specific situations by modifying the general and abstract model to incorporate the specific details of the particular problem under consideration. Hence, it obtains conclusions that do not have general validity (of course) but that, by taking into account the institutional details of the problem at hand, allows the researcher to perform a deeper analysis and recommend possible changes to improve the way that specific assignment problems are solved in practice. For instance, Roth and his collaborators have proposed substantial modifications on the solutions of problems like the yearly assignment of North-American medical interns to hospital internship programs, children to schools in cities like Boston and New York, or the organization of a centralized system to propose kidney transplantations of live donors that are incompatible with their respective and loved patients.

In the remaining of the paper, instead of presenting different models of two-sided assignment problems and their applications as practices of market design, I will restrict myself to present some examples. I will start with an application, describing (in my view) one of the most interesting practices of market design: the kidney exchange problem. To do so, I will present the adaptation of the Gale's top trading cycle algorithm as the best solution to solve kidney exchange problems [20], with some references to the Spanish case. I will also present the notion of stable allocations in the basic marriage model [7] and the main results in terms of the strategic incentives faced by participants in centralized two-sided matching markets. I will mention two other applications of this theory: the yearly assignment of medical students to hospital internship programs in North-America and the yearly assignment of students to schools in Boston and New York cities.

Kidney exchange

There are two treatments for patients with renal disease: dialysis and transplantation. Since dialysis requires a strong dependence and has many side effects (physical as well as psychological), transplantation is considered the best treatment. Kidneys for transplantation come from either deceased or living donors. The first successful kidney transplantation took place on December 23, 1954 in Boston. It was done between two identical twins (to eliminate the immune reaction) and performed, among others, by Joseph E. Murray (1919–2012), J. Hartwell Harrison (1909–1984), and John P. Merrill (1917–1984).[3] The patient survived eight years after the transplantation. At the end of the last century, and after the improvement in immunosuppressive therapies, the majority of transplanted kidneys in many countries were from deceased donors; for instance, in 1999 in Spain less than 1% of all kidney transplants were from living donors (only 17 among 2023). However, there is a unanimous agreement that the quality and success-rate of kidney transplants from living donors are greater than those from deceased ones. In particular, the likelihood that the transplanted kidney survives 5 years is 0.87 if it comes from live donors and 0.80 if it comes from deceased donors, and the likelihood that the recipient will survive 5 years is 0.93 and 0.86, respectively. Furthermore, promoting the donation of kidneys from living donors may help solve the shortage of kidneys for transplantation. Indeed, all countries with active transplantation programs suffer from shortage of kidneys.

Almost everywhere the average time that a patient has to stay in the waiting list for a kidney transplant is well above two years. In addition, increasing life expectancy as well as the decrease in mortality due to car and motorcycle accidents has made the shortage even more severe. For all these reasons, in the last ten years, many countries are promoting living donation; for instance, in 2011 in Spain already more than 12% of all kidney transplants were from living donors (312 among 2498).

In the direct donation, the patient receives, if compatible, one of the two kidneys from a relative or friend (usually, the spouse and siblings of the patient). The most basic incompatibilities are blood and tissue type (the latter is related to genetics that produce immune reaction), although the age of the kidney is also relevant for the graft kidney survival. But if the kidney is not compatible, the transplant is not possible and the donor's kidney is removed from the system. It is estimated that approximately one third of patients with a friend or family donor are excluded from the system due to different incompatibilities.

Until very recently this was the only live donation that was taking place, and there was no system to take advantage of rejected donors, which were simply sent home. In 1986, Felix T. Rapaport (1929–2001) was the first to propose kidney exchanges from living donors. The idea is simple: suppose that one day a nephrologist receives a patient accompanied by a relative who is willing to donate a kidney. Unfortunately, the analysis shows that they are incompatible. The next day, the same doctor receives another patient-donor pair who are also incompatible. But each patient is compatible with the donor of the other pair, and hence, a kidney exchange is possible (in this case, by satisfying a cycle of length two). Or even longer cycles involving three or more incompatible patient-donor pairs could be undertaken.

A kidney exchange problem consists of a set of incompatible patient-donor pairs together with a profile of ordered lists of all donors' kidneys, one list for each patient. Formally, let $N = \{1,...,n\}$ be the set of patients and let $K = \{k_1,...,k_n\}$ be the set of live donors kidneys. Each patient $i \in N$ has a donor whose kidney k_i is not compatible with i. Thus, $\{(1,k),...,(n,k)\}$ is the set of n incompatible patient-donor pairs. Each patient $i \in N$ has a preference order (a strict) ranking P_i of all donors kidneys. For instance, with $n = 4$,

$$P_3: k_2,\ k_4,\ \boxed{k_3},\ k_1$$

indicates that, for patient 3, k_2 and k_4 are two compatible kidneys, k_2 is better than k_4, and k_1 is not compatible (the ordering between incompatible kidneys is irrelevant). A patient's ordered list of all kidneys (from the best to the worst) reflects, according to the patient's nephrologist, the *ex-ante* ordinal quality of the match between each kidney and the patient.

The market design question in this case is to determine a systematic way of selecting, for each kidney exchange prob-

[3] Joseph E. Murray received the Nobel Prize in Physiology or Medicine in 1990, jointly with E. Donnall Thomas, for their discoveries concerning organ and cell transplantation in the treatment of human disease.

lem, a set of compatible transplants with some desirable properties. A set of compatible transplants can be represented by a matching $\alpha : N \to K$, where $\alpha(i) = k_j$ means that if $i \neq j$, i receives kidney k_j and if $i = j$, i does not receive any kidney (and stays under dialysis waiting for a new run of the match). Note that the set of incompatible patient-donor pairs can be represented by the matching μ, where $\mu(i) = k_i$ for all $i \in N$. An instance of a kidney exchange problem is thus a tuple (N, K, μ, P), where N is the set of patients, K is the set of kidneys, μ represents the set of incompatible patient-donor pairs and $P = (P_1, ..., P_n)$ is the profile of agents preferences on K.[4] Roth, Sönmez, and Ünver [20] study the kidney exchange problem and propose an adaptation of the general model presented by Shapley and Scarf [22] as well as of an already known algorithm in matching theory to solve all kidney exchange problems. The algorithm is known as the Gale's top trading cycle algorithm, and I will refer to it as the TTC algorithm.

Given a kidney exchange problem (remember, a set of incompatible patient-donor pairs and a profile of patients' lists, each list ordering all donors' kidneys) the TTC algorithm solves the problem (i.e., proposes a set of compatible transplants) in stages. At each stage, the TTC algorithm roughly works as follows: (a) It constructs a graph whose nodes are the patient-donor pairs that have not yet been matched in the previous stages; (b) it directs the graph (a single arrow leaves from each node pointing to a node) by making that each patient points to the best kidney (according to his ordered list of kidneys) among those still present in the stage; (c) it identifies the nodes on the cycles of the directed graph; and (d) it satisfies the cycles, matching each patient of the nodes of the cycles to his pointed kidney. The TTC algorithm keeps identifying and satisfying successively the cycles along the stages.

Note that in each stage, there is always at least one cycle, if there are several cycles they do not intersect each other and a cycle may have a single node whose patient points to the kidney of his donor (obviously, since they are not compatible, the patient in this case will not be transplanted and the patient will remain under dialysis). Thus, the input of the TTC algorithm is an instance of a kidney exchange problem and its output is a solution of the problem (i.e., a matching) that consists of a proposal of transplants based on the cycles identified along its stages. I denote by η the matching representing the transplants proposed by the output of the TTC algorithm applied to the kidney exchange problem at hand. Example 1 below illustrates how the TTC algorithm works.[5]

Example 1. Let (N, K, μ, P) be a kidney exchange problem with eight incompatible patient-donor pairs, $\mu(i) = k_i$ for each $i = 1, ..., 8$, and the profile P represented in Table 1 below, where the kidneys inside an square in each agent' preference list indicates the initial assignment μ of agents to kidneys (Table 1).

Figure 2 represents the three steps of the TTC algorithm applied to profile P to obtain the assignment η.

The TTC algorithm has many desirable properties. First, it is *individually rational*: every patient that receives a kidney from another donor at the outcome of the TTC algorithm prefers this situation rather than not receiving a kidney and remaining under dialysis. Second, it is *efficient*: all patients cannot improve simultaneously; that is, if there is another set of transplants where one patient receives a strictly better kidney, then there must exist another patient that receives a strictly worse kidney. Third, the output of the TTC algorithm is a stable assignment (in game-theoretic terms, it belongs to the *core* of the kidney exchange problem): there is no subset of patient-donor pairs that, by reassigning only the kidneys of the donors of the patients in the subset, can obtain better kidneys; i.e., no subgroup of the patient-donor pairs (for example those from a hospital, a city or a region) can object unanimously to the output of the TTC algorithm [22]. Moreover, the core of each kidney exchange problem is unique

Table 1. Eight incompatible patient-donor pairs

P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
k_2	k_3	k_1	k_8	k_4	k_8	k_4	k_6
k_3	k_1	k_3	k_7	k_7	k_1	k_8	k_8
k_5	k_2	k_7	k_1	k_3	k_6	k_3	k_1
k_6	k_8	k_2	k_1	k_6	k_5	k_6	k_2
k_8	k_6	k_5	k_2	k_1	k_4	k_1	k_3
k_1	k_4	k_8	k_3	k_8	k_3	k_5	k_7
k_7	k_7	k_6	k_5	k_2	k_2	k_2	k_5
k_4	k_5	k_4	k_6	k_5	k_7	k_7	k_4

[4] Given $i, j, t \in N$ and P_i, I will write $k_j R_i k_t$ to denote that either $k_j = k_t$ or else $k_j P_i k_t$.

[5] Example 1, as well as Example 2, can be found in Massó [10].

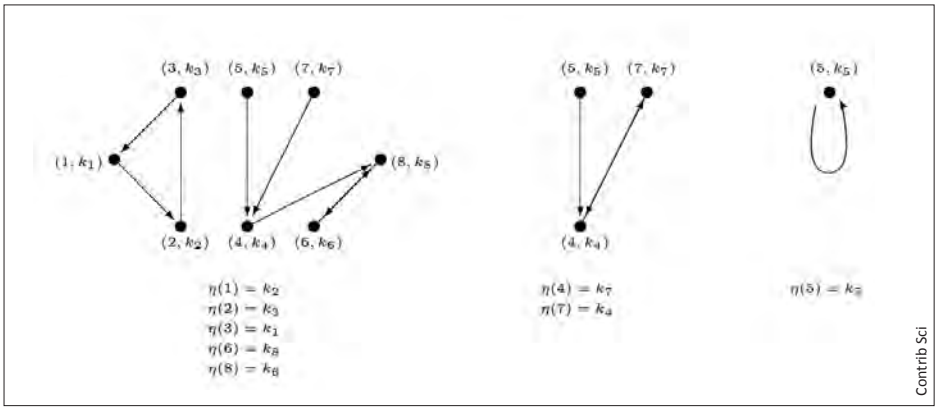

Fig. 2. The three steps of the TTC algorithm applied to obtain the assignment η.

and coincides with the output of the TTC algorithm [19].[6] Fourth, the mechanism associated to the TTC algorithm is strategy-proof: no patient could obtain a strictly better kidney by reporting (in fact, his nephrologist) a false ordered list of kidneys [14]. Furthermore, the mechanism that, for each kidney exchange problem, selects the output of the TTC algorithm is the unique individually rational, efficient, and strategy-proof mechanism [9]. Finally, the quality of the kidney received by each patient in the output of the TTC algorithm depends positively on the quality of his kidney's donor [19].

Roth, Sönmez, and Ünver have also reported some simulations suggesting that the TTC algorithm performs well and that it can be applied to real kidney exchange problems, and indeed it is now used in most countries with kidney exchange programs [20].[7] In addition, the paper has also triggered an already long list of papers studying different issues related to the specific nature of the kidney exchange problem that may require to adapt the TTC algorithm. For instance, (a) to deal with the increasing number of altruistic donors (called "good Samaritans"), whose kidney can be used to initiate chains (instead of cycles) of transplants. On February 18, 2012, *The New York Times* published an article entitled "*60 Lives, 30 Kidneys, All Linked*" describing a chain of 30 transplants initiated one year earlier by a good Samaritan. (b) The consequences of requiring alternative incentive properties (weaker than strategy-proofness) when patients (their nephrologists) submit the ranked list of all donors' kidneys. (c) The presence of patients with several potential donors. (d) Ethical issues related with the *ex-ante* worse situation faced by O blood-type patients since they can only receive kidneys from O blood-type donors. (e) The effects of considering explicitly the dynamic feature of the problem, where the database of pairs keeps changing by the entrance and exit of patient-donor pairs. In any case, kidney exchange has become a natural and successful market design application of the theory of stable allocations to help human beings to live longer and better. Roth and Shapley's contributions have made it possible.

The Theory of Stable Allocations

Following Gale and Shapley's metaphor [7] we will use the marriages between men and women as the reference example to describe a basic matching problem.[8] Let $M = \{m_1, \ldots, m_n\}$ be the set of *men* and let $W = \{w_1, \ldots, w_n\}$ be the set of *women*. The set of *agents* is $N = M \cup W$. We assume that each men $m \in M$ has a strict preference (a ranking) P_m on the set of women and the prospect of remaining unmatched, that for convenience we identify as being matched to himself. That is, P_m is a complete, antisymmetric and transitive binary relation on the set $W \cup \{m\}$. Given $m \in M$ and $w, w' \in W$, we will write wP_mw' and mP_mw to denote that m prefers to be matched to w rather than to w' and to be unmatched instead of being matched to w, respectively. Similarly, each women $w \in W$ has a strict preference Pw on

[6] Note that, by considering the set of all agents and all singleton sets, if an assignment belongs to the core it has to be efficient and individually rational.
[7] They promoted, together with Dr. Francis Delmonico and Susan Saidman, the New England Program for Kidney Exchange (NEPKE). Many countries have now their corresponding centralized programs, for instance, Spain, The Netherlands, The United Kingdom, Italy and South Korea.
[8] The two basic characteristics of the problem are that agent's preferences are ordinal and matching is one-to-one.

Table 2. In the columns, the agents' preference are listed. Each column indicates the corresponding agent's preference in decreasing order

P_{m_1}	P_{m_2}	P_{m_3}	P_{m_4}	P_{m_5}	P_{w_1}	P_{w_2}	P_{w_3}	P_{w_4}
w_1	w_4	w_4	w_1	w_1	m_2	m_3	m_4	m_1
w_2	w_2	w_3	w_4	w_2	m_3	m_1	m_5	m_4
w_3	w_3	w_1	w_3	w_4	m_1	m_2	m_1	m_5
w_4	w_1	w_2	w_2	m_5	m_4	m_4	m_2	m_2
m_1	m_2	m_3	m_4	w_3	m_5	m_5	m_3	m_3
					w_1	w_2	w_3	w_4

the set $M \cup \{w\}$, where w in the ranking represents the prospect, for w, of remaining unmatched. Given $w \in W$ and $m, m' \in M$, we will write $mP_w m'$ and $wP_w m$ to denote that w prefers to be matched to m rather than to m' and to be unmatched instead of being matched to m, respectively. A (preference) profile is a list of preferences $P = (P_{m_1}, ..., P_{m_n}; P_{w_1}, ..., P_{w_m})$, one for each agent. A *market* (or matching problem) is a triple (M, W, P), where M is the set of men, W is the set of women and P is a profile. Example 2 below contains an instance of a market that will be used later on.

Example 2. Let (M, W, P) be the market where $M = \{m_1, m_2, m_3, m_4, m_5\}$, $W = \{w_1, w_2, w_3, w_4\}$, and P is defined in Table 2 where agents' preferences are columns and each column indicates the corresponding agent's preference in decreasing order, for instance, $w_1 P_{m_5} w_2$ and $m_5 P_{m_5} w_3$.

The assignment problem consists of matching each man with at most a woman and each woman with at most a man with the properties that the matching is bilateral and agents may remain unmatched. Formally,

Definition 1: A *matching* (for market (M, W, P)) is a mapping $\mu : M \cup W \to M \cup W$ such that:
(a) for each $m \in M$, $\mu(m) \in W \cup \{m\}$,
(b) for each $w \in W$, $\mu(w) \in M \cup \{w\}$, and
(c) for each pair $(m, w) \in M \times W$, $\mu(m) = w$ if and only if $\mu(m) = m$.

Figure 3 illustrates a matching for this market.

A matching is stable if it is individually rational and no pair of agents prefer each other rather than the partners proposed to each of them by the matching. Formally,

Definition 2: A matching μ is *stable at* P if
(a) for each $m \in M$, $\mu(m) R_m m$,[9]
(b) for each $w \in W$, $\mu(w) R_w w$, and
(c) there is no pair $(m, w) \in M \times W$ such that $wP_m \mu(m)$ and $mP_w \mu(w)$.

The matching μ in Fig. 2 is not stable at profile P of Example 2 since $w_2 P_{m_1} w_3 = \mu(m_1)$ and $m_1 P_{w_2} m_5 = \mu(w_2)$. Fix M and W. Given P, let $S(P)$ be the set of stable matchings at P. Gale and Shapley state and prove that the set of

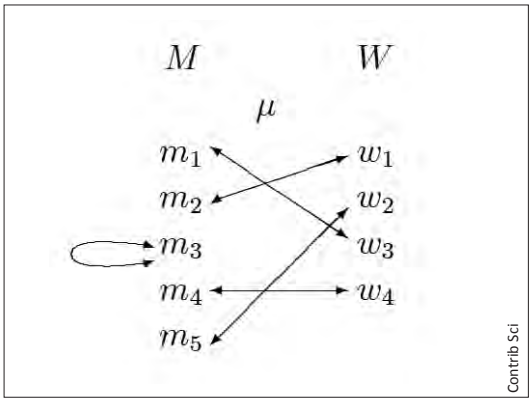

Fig. 3. A matching for the market of Example 2.

[9] Given agents $x, y, z \in N$ we write $xR_y z$ to denote that either $x = z$ or else $xP_y z$; namely, either x and z are the same agent or else y strictly prefers to be matched to x rather than to z.
[10] Moreover, it coincides with the core of the one-to-one matching problem; namely, intermediate coalitions of agents have no additional blocking power.

Table 3. Four steps of the deferred acceptance algorithm in which men make offers to women, applied to the market (M, W, P) of Example 2

Step 1	Step 2	Step 3	Step 4	Final
$m_1 \to w_1$ Yes	$m_1 \to w_1$ Yes	$m_1 \to w_1$ Yes	$m_1 \to w_1$ Yes	$\mu_M(m_1) = w_1$
$m_2 \to w_4$ Yes	$m_2 \to w_4$ No	$m_2 \to w_2$ Yes	$m_2 \to w_2$ Yes	$\mu_M(m_2) = w_2$
$m_3 \to w_4$ No	$m_3 \to w_3$ Yes	$m_3 \to w_3$ Yes	$m_3 \to w_3$ Yes	$\mu_M(m_3) = w_3$
$m_4 \to w_1$ No	$m_4 \to w_4$ Yes	$m_4 \to w_4$ Yes	$m_4 \to w_4$ Yes	$\mu_M(m_4) = w_4$
$m_5 \to w_1$ No	$m_5 \to w_2$ Yes	$m_5 \to w_2$ No	$m_5 \to w_4$ No	$\mu_M(m_5) = w_5$

stable matchings is always non-empty [7].[10] Formally,

Theorem 1: *Let P be a profile. Then, $S(P)$ is non-empty.*

Gale and Shapley prove that, for any P, the set $S(P)$ is non-empty by showing that it contains two stable matchings, the *men-optimal stable matching* (denoted by μ_M) and the *women-optimal stable matching* (denoted by μ_W) [7]. The two matchings have the properties that for any stable matching $\mu \in S(P)$ the following two conditions hold: (a) for all $m \in M$, $\mu_M(m) R_m \mu(m) R_m \mu_W(m)$ and (b) for all $w \in W$, $\mu_W(w) R_w \mu(w) R_w \mu_M(w)$; namely, all men agree that the partner that they receive at $\mu_M(\mu_W)$ is the best (worst) among all partners that they receive at any stable matching and, simultaneously, all women agree that the partner that they receive at $\mu_W(\mu_M)$ is the best (worst) among all partners that they receive at any stable matching.[11] Gale and Shapley propose two versions of the deferred acceptance algorithm (*DAA*) to compute the two optimal stable matchings μ_M and μ_W [7]. I describe the version of the algorithm in which men make offers to women, denoted by DAA_M (the other is symmetric, replacing the role of men and women and it is denoted by DAA_W). At any step of the DAA_M, each man offers himself to his most-preferred woman amongst the set of women who have not already rejected him, while each woman accepts the most-preferred men amongst all men whose proposals along the algorithm she has not rejected yet. The algorithm terminates when no woman rejects a man. It turns out that the outcome of the DAA_M is μ_M and the outcome of the DAA_W is μ_W.

Table 3 summarizes the 4 steps of the DAA_M applied to the market (M, W, P) of Example 2, where $m \to w$ represents an offer of m to w, Yes means that w accepts it, and No that w rejects it.

Table 4 describes the unique step of the DAA_W applied to the market (M, W, P) of Example 2. Observe that $\mu_M \neq \mu_W$ and that $\mu_M(w_5) = \mu_W(w_5) = w_5$.[12]

Roth has made relevant contributions to the study of the strategic incentives induced by the *DAAs* when they are understood as direct revelation mechanisms.[13] Moreover, he has proposed modifications of some mechanisms used to solve real-life assignment problems. Some of the modifications are partially driven by the aim of fixing mechanisms that induce wrong strategic incentives to agents. At the end of this section I will be a bit more specific about two of these modifications. But first, to consider the strategic incentives

Table 4. Unique step of the DAA_W applied to the market (M, W, P) of Example 2

Step 1	Final
$w_1 \to m_2$ Yes	$\mu_W(w_1) = m_2$
$w_2 \to m_3$ Yes	$\mu_W(w_2) = m_3$
$w_3 \to m_4$ Yes	$\mu_W(w_3) = m_4$
$w_4 \to m_1$ Yes	$\mu_W(w_4) = m_5$
	$\mu_W(w_5) = m_5$

[11] Ref. [8] shows that $S(P)$ is a (dual) complete lattice with the unanimous partial ordering of men (women) $\geq_M$ ($\geq_W$), where for any $\mu, \mu' \in S(P)$, $\mu \geq_M \mu'$ if and only if $\mu(m) R_m \mu'(m)$ ($\mu \geq_W \mu'$ if and only if $\mu(w) R_w \mu'(w)$). Moreover, $\mu \geq_M \mu'$ if and only if $\mu' \geq_W \mu$. Then, μ_M is the supremum and μ_W is the infimum of the set $S(P)$ according to $\geq_M$, and μ_W is the supremum and μ_M is the infimum of the set $S(P)$ according to $\geq_W$.

[12] The following property of the set of stable matchings $S(P)$ always holds. For any agent $x \in M \cup W$ if $\mu \in S(P)$ and $\mu(x) = x$ then, for all $\mu' \in S(P)$, $\mu'(x) = x$. Namely, to be unmatched is a global property of the set of stable matchings.

[13] A direct revelation mechanism asks each agent to report his preferences and proposes a matching depending on the declared profile of preferences.

faced by participants in these markets, observe that whether a matching is stable depends on the agents' preferences. But each agent's preferences are private information and hence, they have to be elicited by a mechanism. Fix the sets *M* and *W*. Let $\mathcal{M}$ be the set of all matchings among *M* and *W* and let $\mathcal{P}$ be the set of all preference profiles. A *social choice function* is a mapping $f:\mathcal{P}\to\mathcal{M}$ selecting, for each preference profile $P\in\mathcal{P}$, a matching $f(P)\in\mathcal{M}$. Given a social choice function $f:\mathcal{P}\to\mathcal{M}$, a profile $P\in\mathcal{P}$ and an agent $x\in M\cup W$ we denote by $f^x(P)$ the partner assigned to x by the social choice function f at profile P (i.e., $f^x(P)\equiv f(P)(x)$, because $f(P)$ is the matching selected by f at P). Given agent $x\in M\cup W$, a profile $P\in\mathcal{P}$ and a preference P'_x denote by (P'_x, P_{-x}) the new profile obtained from P after replacing P_x by P'_x in P. Agent $x\in M\cup W$ *manipulates* the social choice function $f:\mathcal{P}\to\mathcal{M}$ if there exist $P\in\mathcal{P}$ and P'_x such that $f^x(P'_x,P_{-x})P_x f^x(P_x,P_{-x})$; namely, agent x (with preference P_x) obtains an strictly preferred partner by reporting to *f* a false preference P'_x. A social choice function $f:\mathcal{P}\to\mathcal{M}$ is *strategy-proof* if no agent can manipulate it.[14] A social choice function $f:\mathcal{P}\to\mathcal{M}$ is *stable* if it always selects stable matchings; namely, for all $P\in\mathcal{P}$, $f(P)\in S(P)$. Roth shows that strategy-proofness and stability are incompatible [13].

Proposition 1: *There is no social choice function $f:\mathcal{P}\to\mathcal{M}$ that is simultaneously strategy-proof and stable* [13].

However, the two DAAs understood as social choice functions induce good incentive properties to the side of the market that makes the offers. To state that, let $f^M:\mathcal{P}\to\mathcal{M}$ be the social choice function that selects for each preference profile the men-optimal stable matching and let $f^W:\mathcal{P}\to\mathcal{M}$ be the social choice function that selects for each preference profile the women-optimal stable matching; namely, for each $P\in\mathcal{P}$, $f^M(\ )=\mu_M$ and $f^W(P)=\mu_W$. A social choice function $f:\mathcal{P}\to\mathcal{M}$ is *strategy-proof for the men* if it can not be manipulated by any men and $f:\mathcal{P}\to\mathcal{M}$ is *strategy-proof for the women* if it can not be manipulated by any women. The following result may explain why these two social choice functions are used so widely to solve many real-life centralized two-sided matching problems.

Theorem 2: *The social choice function $f^M:\mathcal{P}\to\mathcal{M}$ is strategy-proof for the men and the social choice function $f^W:\mathcal{P}\to\mathcal{M}$ is strategy-proof for the women* [4,13].

Finally, I mention two successful market design applications of the theory of stable allocations. First, Roth [15] reports that, since the academic year 1951–1952 (ten years earlier than Gale and Shapley's paper [7]), the problem of matching each year medical students with hospital internship programs in North-America was solved by the Association of American Medical Colleges (AAMC) by asking to medical students and hospital to report their ranked preferences lists and by applying to the declared preference profile the *DAA* in which hospitals make offers.

Before 1951, and as earlier as the beginning of the 20th century, the matching process had many problems. In particular, the market unraveled in the sense that hospitals were looking (and making offers in a decentralized setting) to medical students earlier and earlier while they were still at college, and needed almost two additional years of college before finishing. The AAMC tried to stop these practices without much success until 1953–1954 when the centralized *ADD* mechanism was adopted under voluntary basis. The procedure worked well with high participation rates until the mid-1990's (around 20,000 medical students were assigned yearly) when more couples where looking coordinately for hospitals located in the same city, some links had to be done between different subspecialties to fulfill the internship requirements, and students were arguing that the system was favoring hospitals and that they could "game the system" by reporting false preference lists. The AAMC asked Roth to modify the mechanism to fix those problems and he redesigned the algorithm to be able to accommodate them satisfactorily. In 1998 the match was completed using (with small modifications) the *DAA* in which students make offers [18]. This intervention may be seen as the first (conscious) practice of market design.

Abdulkadiroğlu and Sönmez [3] used a two-sided matching model to study the yearly problem of assigning students to public schools in a city. The main issue of the assignment problem is to let parents to choose the school of their children. Boston and New York cities were using a centralized mechanism (known as the Boston mechanism) that is similar to the *DAA* but with the very important difference that provi-

[14] That is, truth-telling is a (weakly) dominant strategy in the game induced by the social choice function f.

[15] For instance, in the application of the DAA $_M$ to the profile P of Example 2, m_3 is assigned to w_3, his second choice, while in the outcome of the Boston mechanism he is unassigned (his worse choice). However, under the Boston mechanism m_3 could be assigned to w_2 by declaring, instead of Pm_3, any preference $P'm_3$ with w_2 as his top choice.

sional matches along the application of the algorithm are made definitive, and hence, it was highly manipulable.[15]

Abdulkadiroğlu and others have reported that, following their advice, the two cities changed the assignment procedure and replaced the Boston mechanism by the *DAA* in which students make offers [1,2]. Recently, many other cities are adopting the *DAA* to organize the assignment of their students to public schools.

Aknowledgements. I thank Saptarshi Mukherjee for helpful comments, and the financial support received from the Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centers of Excellence in R&D (SEV-2015-0563) and through grants ECO2008-0475-FEDER (Grupo Consolidado-C) and ECO2014-53051, and from the Generalitat de Catalunya, through grant SGR2014-515.

Competing interests: None declared.

References

1. Abdulkadiroğlu A, Pathak P, Roth A (2005) The New York city high school match. Am Econ Rev P and P 95:364-367
2. Abdulkadiroğlu A, Pathak P, Roth A, Sönmez T (2005) The Boston Public School Match. Am Econ Rev P and P 95:368-371
3. Abdulkadiroğlu A, Sönmez T (2003) School choice: a mechanism design approach. Am Econ Rev 93:729-747
4. Dubins L, Freedman D (1981) Machiavelli and the Gale-Shapley algorithm. Am Math Mon 88:485-494
5. Erev I, Roth A (1998) Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. Am Econ Rev 88:848-881
6. Gale D (1962) The theory of linear economic models. McGraw-Hill, New York
7. Gale D, Shapley L (1962) College admissions and the stability of marriage. Am Math Mon 69:9-15
8. Knuth D (1991) Stable marriages and its relation to other combinatorial problems. CRM Proceedings and Lecture Notes 10, American Mathematical Society, Providence
9. Ma J (1994) Strategy-proofness and the strict core in a market with indivisibilities. Int J Game Theory 23:75-83
10. Massó J (2012) La moderna teoria de l'elecció social: de la impossibilitat a la possibilitat. Butlletí de la Societat Catalana de Matemàtiques 27:177-231
11. Roth A (1977) The Shapley Value as a von Neumann-Morgenstern utility. Econometrica 45:657-664
12. Roth A (1979) Independence of irrelevant alternatives, and solutions to Nash's bargaining problem. J Econ Theory 16:247-251
13. Roth A (1982) Incentive compatibility in a market with indivisible goods. Econ Lett 9:127-132
14. Roth A (1982) The economics of matching: stability and incentives. Math Oper Res 7:617-628
15. Roth A (19984) The evolution of the labor market for medical interns and residents: a case study in Game Theory. J Polit Econ 92:991-1016
16. Roth A (ed) (1988) The Shapley Value: Essays in honor of Lloyd S. Shapley. Cambridge University Press, Cambridge
17. Roth A, Ockenfels A (2002) Last-minute bidding and the rules for ending second-price auctions: evidence from eBay and Amazon auctions on the internet. Am Econ Rev 92:1093-1103
18. Roth A, Peranson E (1999) The redesign of the matching market for American physicians: some engineering aspects of economic design. Am Econ Rev 89:748-780
19. Roth A, Postlewaite A (1977) Weak versus strong domination in a market with indivisible goods. J Math Econ 4:131-137
20. Roth A, Sönmez T, Ünver U (2004) Kidney exchange. Q J Econ 119: 457-488
21. Roth A, Sotomayor M (1990) Two-sided matching: a study in game-theoretic modelling and analysis. Cambridge University Press and Econometric Society Monographs 18
22. Shapley L, Scarf H (1974) On cores and indivisibilities. J Math Eco 1:23-28