Chapter 16

TWO-SIDED MATCHING

ALVIN E. ROTH[a] and MARILDA SOTOMAYOR[b]
[a]*University of Pittsburgh and*
[b]*Pontificia Universidade Catolica do Rio de Janeiro*

Contents

1. Introduction 486
2. Some empirical motivation 486
   2.1. The case of American physicians 486
   2.2. Bidder rings in auctions 490
3. Several simple models: Stability, and the polarization of interests in the core 491
   3.1. The marriage model 492
   3.2. The reformulated college admissions model 494
   3.3. Complex preferences over groups 498
   3.4. The assignment model 502
4. The structure of the set of stable matchings 507
   4.1. Size of the core 511
   4.2. The linear structure of the set of stable matchings in the marriage model 512
   4.3. Comparative statics: New entrants 514
5. Strategic results 515
   5.1. Strategic behavior in models of one-to-one matching 518
   5.2. Many-to-one matching: The college admissions model 525
   5.3. Incomplete information 527
6. Empirical overview 530
   6.1. Some further remarks on empirical matters 532
Bibliography 535

Handbook of Game Theory, Volume 1, Edited by R.J. Aumann and S. Hart

1. Introduction

The games we consider in this chapter are "two-sided matching markets". The phrase "two-sided" refers to the fact that agents in such markets belong, from the outset, to one of two disjoint sets – e.g. firms or workers. The term "matching" refers to the bilateral nature of exchange in these markets – e.g. if I work for some firm, then that firm employs me. In recent years the game-theoretic analysis of these markets has proved useful in various empirically oriented studies. To emphasize the close connection between empirical and theoretical work in this area, this chapter begins by describing some of the phenomena the theory should be able to explain. Much of the available theory will be summarized in the body of the chapter, and the chapter will conclude by returning to consider how the theory addresses the empirical questions raised at the beginning.

We will be concerned both with the core of the game, and with the dominant and equilibrium strategies under various rules about how the game might be played. Thus this material will serve to emphasize that the distinction between "cooperative" and "noncooperative" game theory is often somewhat artificial, since the tools of both kinds of theory can be used to study the same phenomena.

This chapter is adapted from our monograph, Roth and Sotomayor (1990a), in which a much more complete treatment can be found.

2. Some empirical motivation

2.1. The case of American physicians

Hospitals began offering newly-graduated medical students internship positions around the turn of the century. Not until 1945 were the relevant medical associations able to institute a single market with uniform dates at which such positions would be offered.[1] Once this was accomplished, however, both students and hospitals were dismayed by the chaotic conditions that developed between the time offers of internships were first made, and the time by which students were required to accept or reject them. The situation is described as follows in Roth (1984a).

[1]See Roth (1984a) for a description of the difficulties encountered in setting uniform appointment dates prior to 1945, which will not be discussed here.

Basically, the problem was that a student who was offered an internship at, say, his third choice hospital, and who was informed he was an alternate (i.e. on a waiting list) at his second choice, would be inclined to wait as long as possible before accepting the position he had been offered, in the hope of evantually being offered a preferable position. Students who were pressured into accepting offers before their alternate status was resolved were unhappy if they were ultimately offered a preferable position, and hospitals whose candidates waited until the last minute to reject them were unhappy if their preferred alternate candidates had in the meantime already accepted positions. Hospitals were unhappier still when a candidate who had indicated acceptance subsequently failed to fulfil his commitment after receiving a preferable offer. In response to pressure originating chiefly from the hospitals, a series of small procedural adjustments were made in the years 1945–51. The nature of these adjustments, described next, makes clear how these problems were perceived by the parties involved.

For 1945, it was resolved that hospitals should allow students ten days after an offer had been made to consider whether to accept or reject it. For 1946, it was resolved that there should be a uniform appointment date (July 1) on which offers should be tendered . . . , and that acceptance or rejection should not be required before July 8. By 1949, [the Association of American Medical Colleges] proposed that appointments should be made by telegram at 12:01 AM (on November 15), with applicants not being required to accept or reject them until 12:00 Noon the same day. Even this twelve-hour waiting period was rejected by the American Hospital Association as too long: the joint resolution finally agreed upon contained the phrase "no specified waiting period after 12:01 AM is obligatory," and specifically noted that telegrams could be filed in advance for delivery precisely at 12:01 AM. In 1950, the resolution again included a twelve-hour period for consideration, with the specific injunction that "Hospitals and/or students shall not follow telegrams of offers of appointment with telephone calls' until after the twelve-hour grace period." [. . . the injunction against telephone calls was two-way, in order to stem a flood of calls both from hospitals seeking to pressure students into an immediate decision, and from students seeking to convert their alternate status into a firm offer].

It was eventually recognized that these problems could not be solved by compressing still further the time allowed for the last stage of the matching process, and it was agreed to try instead a centralized matching algorithm, on a voluntary basis. Students and hospitals would continue to exchange information via applications and interviews as before, but then both students and hospitals would submit rank-orderings of their potential assignments,[2] and the algorithm would be used to suggest a matching of students to hospitals, who would then, it was hoped, sign employment contracts with their suggested assignments.

The first algorithm proposed was abandoned after a year because it was observed to give students the incentive to submit a rank-ordering different from their true preferences. The algorithm proposed in its place was used for the first time in 1951, and remains in use to this day. (This algorithm will be called the NIMP algorithm, for National Intern Matching Program, the name under which the algorithm was initially administered.)

This system of arranging matches was *voluntary* – students and hospitals were free to try to arrange their own matches outside the system, and there was no way to enforce compliance on those who did participate.[3] This makes it all the more remarkable that, in the first years of operation, over 95 percent of eligible students and hospitals participated in the system, and these high rates of participation continued until the early 1970s.

Since then, although the overall rate of participation remains high, increasing numbers of students, particularly those among the growing number of medical students who are married to other medical students, have begun to seek to arrange their own matches, without going through the centralized clearinghouse. Another aspect of this market which has caused some concern in medical circles has been the resulting distribution of physicians among hospitals, with rural hospitals getting fewer interns than they wish, and a much higher percentage of interns who are graduates of foreign medical schools.

The chief phenomena we would like a theory to explain in this case are:

- What accounted for the disorderly operation of the market between 1945 and 1951?
- Why was the centralized procedure instituted in the 1951–52 market able to achieve such high rates of voluntary participation?
- Why did these high rates start to diminish by the 1970s, particularly among the growing number of medical students who were married to other medical students?

We will also want to investigate "strategic" questions of the kind that led to the scrapping of the first algorithm.

- Does the NIMP algorithm, as claimed by the sponsoring medical associations,

[2]Regarding the problem of formulating a rank-ordering, note that the complete job description offered by a hospital program in a given year was customarily specified in advance. Thus the responsibilities, salary, etc. associated with a given internship, while they might be adjusted from year to year in response to a hospital's experience in the previous year's market, were not a subject of negotiation with individual job candidates.

[3]The experience prior to 1950 amply demonstrated that no amount of moral suasion was effective at preventing participants from acting in what they perceived as their own best interests.

give students and hospitals the incentive to submit rank-orderings corresponding to their true preferences?

Finally, we would like to be able to get some idea of which aspects of the market could be influenced by modifying its organization, while preserving those features that have led to high rates of voluntary participation. In this regard, we will want to know:

- Can the defection of married couples be halted?
- Can the distribution of interns to rural hospitals be changed?

A preview of the proposed explanation

When we see that, in the late 1940s, there is a lot of two-way telephone traffic between hospitals and students, who sometimes renege on previous verbal agreements, we can hypothesize that there is some systematic incentive to the parties involved to behave in this way. These incentives must be mutual: if students who called hospitals that had not extended them offers were uniformly told that no places were available, the practice would be unlikely to persist in the virulent form that was observed. Situations in which there are some students and hospitals who are not matched with each other, but who *both* prefer to be matched one to the other, will therefore be called "unstable". By the same token, if the matching suggested by the NIMP algorithm was unstable in this way – i.e. if there were students and hospitals that would prefer to be matched to one another rather than to accept the suggested match – then we would expect that these students and hospitals would continue to try to locate each other, and subsequently decline to accept the assignment suggested by the matching procedure. The very high rates of voluntary participation in the years following the introduction of the NIMP procedure suggest this was not the case, and that the set of suggested assignments produced by the NIMP procedure must be "stable", i.e. must have the property that, if some student would prefer another hospital to his suggested assignment, then that hospital does not return the favor, but prefers the students assigned to it to the student in question. In Section 6 we will see that the NIMP assignments do indeed have this property. So our explanation of why the chaotic market conditions prior to 1951 vanished following the introduction of the NIMP procedure will be that it introduced this kind of stability to the market.

In a similar vein, we will observe that, as married couples became more common in this market, the procedures used to deal with them introduced instabilities once again, so that married couples could find hospitals that they preferred to their assigned matches and that were willing to offer them jobs. This will be the basis of our explanation of the defection of married couples from the system, which became so noticeable in the mid-1970s. We will also argue that the answers to our questions about how much freedom there is to alter the organization of the market while maintaining high rates of voluntary participation also hinge on whether any given organization of the market leads to stable market outcomes.

A complementary set of ideas, having to do with the strategies of individual agents in the market, will be used to explore the question of whether, as claimed, it is always in the interest of all parties to state their true preferences. We will see it is not, and that it cannot be for any procedure that produces stable outcomes. However, it is possible to arrange things so that it is always in the best interest of *some* of the parties to state their true preferences. The development of these ideas will involve us in a number of subtle issues, not the least of which is that we will be forced to reconsider and re-evaluate our conclusions about stability. If the students and hospitals may not be stating their true preferences when they submit rank-order lists for the NIMP algorithm, is there still reason to believe that the outcome is a stable set of assignments? When we look at equilibrium behavior we will see that there is.

2.2. *Bidder rings in auctions*

Strategic considerations of a somewhat different sort arise in the study of auction markets. The opportunities to profitably deviate from straightforward behavior are different for buyers and sellers. The sellers (and their agent the auctioneer) would like prices to be high, and the buyers would like prices to be low. The most commonly reported "strategic" behavior on the part of auctioneers or sellers is to introduce imaginary bids into the proceedings, which when practiced by auctioneers is called by a variety of colorful names, such as "pulling bids off the chandelier". And the most commonly reported strategic behavior on the part of buyers is to form *rings* that agree to coordinate their bidding in an effort to keep down the price. Cassady (1967) reports that in antique and art auctions, the subsequent auction among members of the ring, called a "knockout" auction, serves both to determine which of the ring members will receive what the ring has bought, and what payments shall be made by ring members among themselves. (The *Oxford English Dictionary* cites nineteenth-century sources for this meaning of the word "knockout", suggesting that the organization of bidder rings in this way is not only a widespread phenomenon, but also not a new one.) Cassady remarks that buyer rings are common in many kinds of auctions all over the world, although in auctions of divisible commodities (such as fish in England, timber in the United States, and wool in Australia), rings commonly divide the purchases among themselves, rather than conducting a knockout auction. An unusually detailed description of the strategic behavior of rings and auctioneers in New Jersey machine tool auctions is given by Graham and Marshall (1984). The analysis that follows will shed some light on the strategic opportunities facing the auctioneer and individual bidders, and the opportunities for bidders to organize themselves into rings.

3. Several simple models: Stability, and the polarization of interests in the core

Most theoretical work on this topic traces its history to the papers of Gale and Shapley (1962), and Shapley and Shubik (1972). Gale and Shapley formulated a model of two-sided matching without sidepayments which they called the marriage problem, and Shapley and Shubik formulated a sidepayment game which they called the assignment game. Each paper studied the core of the game, and showed it is nonempty for any preferences of the agents. Curiously, although Gale and Shapley were unaware of the 1951 NIMP algorithm, they showed the core was nonempty by formulating what can be regarded as an equivalent algorithm [Roth (1984a)]. Both for the marriage problem and the assignment game, these early papers demonstrated that, within the core, there is a surprising coincidence of interest among players on the same side of the market, and a polarization of interest between the two sides of the market. These two models are introduced in Subsections 3.1 and 3.4.

Both models involve one-to-one matching, i.e. each agent on one side of the market can be matched to at most one agent on the other side. Gale and Shapley also discussed the case of many-to-one matching, which they called the college admissions problem, but they treated this as essentially equivalent to the marriage problem. Although they considered that agents on one side of the market (e.g. colleges) could be matched to more than one agent on the other side (e.g. students), colleges' preferences were only considered to be defined over individual students, not over groups. For a number of years thereafter, the case of many-to-one matching was regarded as equivalent to one-to-one matching.

That this is not the case was observed in Roth (1985a), where some erroneous conclusions that had been reached about many-to-one matching were considered, and where a model of many-to-one matching was reformulated as a well-defined game. That model, presented in Subsection 3.2, is a straightforward generalization of the marriage model, in that colleges continue to have preferences over individual students, and their preferences over groups of students are constrained by their preferences over individuals in a simple way. The model of Subsection 3.3, in contrast, is a further generalization in which firms' preferences over groups of workers need not reflect an underlying preference over individuals.[4] We will see that while these generalizations and alternative formulations differ in important ways from the simple marriage problem, all of these models also share a number of their most striking properties.

As in any game-theoretic analysis, it will be important to keep clearly in mind the "rules of the game" by which agents may become matched to one another, as these will influence every aspect of the analysis. We will suppose the general rules are that any pair of agents on opposite sides of the market may be matched to one another if they both agree, and any agent is free to remain unmatched. We will consider more detailed descriptions of possible rules (concerning, for example, how proposals are made, or whether a marriage broker plays a role) at various points in the discussion.

3.1. *The marriage model*

The two finite and disjoint sets of agents in the marriage model are the set $M = \{m_1, m_2, \ldots, m_n\}$ of men, and $W = \{w_1, w_2, \ldots, w_m\}$ of women. Each man has preferences over the women, and each woman has preferences over the men. These preferences are transitive and complete, and may be such that a man m, say, would prefer to remain single rather than marry some woman w he does not care for.

The preferences of each man m will be represented by an ordered list, $P(m)$, on the set $W \cup \{m\}$. That is, a man m's preferences might be of the form $P(m) = w_1, w_2, m, w_3, \ldots, w_m$, indicating that his first choice is to be married to woman w_1, his second choice is to be married to woman w_2, and his third choice is to remain single. Similarly, each woman w in W has an ordered list of preferences, $P(w)$, on the set $M \cup \{w\}$. (An agent may also be indifferent between several possible mates.) We will usually describe an agent's preferences by writing only the ordered set of people that the agent prefers to being single. Thus the preferences $P(m)$ described above will be abbreviated by $P(m) = w_1, w_2$.

Denote by $\boldsymbol{P}$ the set of preference lists $\boldsymbol{P} = \{P(m_1), \ldots, P(m_n), P(w_1), \ldots, P(w_m)\}$, one for each man and woman. A specific marriage market is denoted by the triple $(M, W; \boldsymbol{P})$. We write $w >_m w'$ to mean m prefers w to w', and $w \geqslant_m w'$ to mean m likes w at least as well as w'. Similarly

[4]The model presented here is a special case of one formulated in Roth (1984c), which in turn builds upon the work of Kelso and Crawford (1982). (In each of these models we may refer to the agents as firms and workers, but in the marriage model we will also refer to them as men and women, in the reformulated college admissions model as colleges and students, and in the assignment model as buyers and sellers, in order to keep in mind the particular assumptions of those models.)

we write $m >_w m'$ and $m \geqslant_w m'$. Woman w is called *acceptable* to man m if he likes her at least as well as remaining single, i.e. if $w \geqslant_m m$. Analogously, m is *acceptable* to w if $m \geqslant_w w$. If an individual is not indifferent between any two acceptable alternatives, he or she has *strict preferences*.

An outcome of the marriage market is a set of marriages. In general, not everyone may be married – some people may remain single. (We will adopt the convention that a person who is not married to someone is *self-matched*.) Formally we have

Definition 1. A *matching* μ is a one-to-one correspondence from the set $M \cup W$ onto itself of order two [that is, $\mu^2(x) = x$] such that if $\mu(m) \neq m$, then $\mu(m)$ is in W and if $\mu(w) \neq w$, then $\mu(w)$ is in M. We refer to $\mu(x)$ as the *mate* of x.

Note that $\mu^2(x) = x$ means that if man m is matched to woman w [i.e. if $\mu(m) = w$], then woman w is matched to man m [i.e. $\mu(w) = m$]. The definition also requires that individuals who are not single be matched with agents of the opposite set – i.e. men are matched with women. A matching will sometimes be represented as a set of matched pairs, e.g.

$$\mu = \begin{matrix} w_4 & w_1 & w_2 & w_3 & (m_5) \\ m_1 & m_2 & m_3 & m_4 & m_5 \end{matrix} ,$$

has m_1 married to w_4 and m_5 remaining single, i.e. $\mu(m_1) = w_4$ and $\mu(m_5) = m_5$, etc.

We will assume that each agent's preferences over alternative matchings correspond exactly to his (her) preferences over his own mates at the two matchings. Thus man m, say, prefers matching μ to matching ν if and only if he prefers $\mu(m)$ to $\nu(m)$.

A matching μ is individually irrational if it contains a matched pair (m, w) who are not mutually acceptable, and we say such a matching can be *improved upon* by an individual, since the rules allow any agent to remain single if he or she chooses. Similarly, a matching μ can be improved upon by some pair consisting of a man m and woman w if m and w are not matched to one another at μ, but prefer each other to their assignments at μ, i.e. if $w >_m \mu(m)$ and $m >_m \mu(w)$. The motivation of this terminology should be clear. Suppose such a matching μ should be under consideration – e.g. suppose no agreements have yet been reached, but courtships are under way that if successfully concluded will result in the matching μ. This state of affairs would be unstable in the sense that man m and woman w would have good reason to disrupt it in order to marry each other, and the rules of the game allow them to do so. This leads to the following definition.

Definition 2. A matching μ is *stable* if it cannot be improved upon by any individual or any pair of agents.

Note that unstable matchings are those dominated via coalitions consisting of individuals or pairs, and so unstable matchings are not in the core of the game. But the core is the set of matchings undominated by coalitions of any size, and so the set of stable matchings might strictly contain the core. But for this model of one-to-one matching, that is not the case.

Theorem 1. *The core of the marriage market equals the set of stable matchings.*

Proof. If μ is not in the core, then μ is dominated by some matching μ' via a coalition A. If μ is not individually irrational, this implies $\mu'(w) \in M$ for all w in A, since every woman w in A prefers $\mu'(w)$ to $\mu(w)$, and A is effective for μ'. Let w be in A and $m = \mu'(w)$. Then m prefers w to $\mu(m)$ and the pair (m, w) can improve upon μ, so μ is unstable. □

We will continue to speak of stable (rather than core) matchings since in the more general models of many-to-one matching that follow, the set of stable matchings will be a subset of the core. For the marriage model, Gale and Shapley proved the following.

Theorem 2 [Gale and Shapley (1962)]. *The set of stable matchings is always nonempty. And when all men and women have strict preferences it contains an M-optimal stable matching, which all the men like at least as well as every other stable matching, and, similarly, a W-optimal stable matching.*

We will defer discussion of the proof until the more general model of Subsection 3.3.[5] We turn next to a many-to-one generalization of the marriage model in which it continues to be meaningful to speak of firms as having preferences over individual workers.

3.2. The reformulated college admissions model

There are two finite and disjoint sets, $\mathscr{C} = \{C_1, \ldots, C_n\}$ and $S = \{s_1, \ldots, s_m\}$, of colleges and students, respectively. Each student has preferences over the

[5]Roth and Vande Vate (1990) construct another kind of existence proof, based on the observation that a sequence of matchings generated by allowing randomly chosen blocking pairs to form must converge with probability one to a stable matching. (The difficulty lies in the fact that cycles of unstable matching may arise.)

colleges, and each college has preferences over individual students, exactly as in the marriage model.

The first difference from the marriage model is that, associated with each college C is a positive integer q_C called its *quota*, which indicates the maximum number of positions it may fill. (That all q_C positions are identical is reflected in the fact that students' preferences are over colleges – they do not distinguish between positions.)

An outcome is a matching of students to colleges, such that each student is matched to at most one college, and each college is matched to at most its quota of students. A student who is not matched to any college will be "matched to himself" as in the marriage model, and a college that has some number of unfilled positions will be matched to itself in each of those positions. A matching is bilateral, in the sense that a student is enrolled at a given college if and only if the college enrolls that student. To give a formal definition, first define, for any set X, an *unordered family of elements of* X to be a collection of elements, not necessarily distinct. So a given element of X may appear more than once in an unordered family of elements of X.

Definition 3. A *matching* μ is a function from the set $\mathscr{C} \cup S$ into the set of unordered families of elements of $\mathscr{C} \cup S$ such that:

(1) $|\mu(s)| = 1$ for every student s and $\mu(s) = s$ if $\mu(s) \not\subseteq \mathscr{C}$;

(2) $|\mu(C)| = q_C$ for every college C, and if the number of students in $\mu(C)$, say r, is less than q_C, then $\mu(C)$ contains $q_C - r$ copies of C;

(3) $\mu(s) = C$ if and only if s is in $\mu(C)$.

So $\mu(s_1) = C$ denotes that student s_1 is enrolled at college C at the matching μ, and $\mu(C) = \{s_1, s_3, C, C\}$ denotes that college C, with quota $q_C = 4$, enrolls students s_1 and s_3 and has two positions unfilled.

At this point in our description of the marriage model we had only to say that each agent's preferences over alternative matchings correspond exactly to his preferences over his own assignments at those matchings. We can now say this about students, since at each matching a student is either unmatched or matched to a college, and we have already described student's preferences over colleges. But, while we have described colleges' preferences over students, each college with a quota greater than 1 must be able to compare *groups* of students in order to compare alternative matchings, and we have yet to describe the preferences of colleges over groups. (Until we have described colleges' preferences over matchings, the model will not be a well-defined game.)

The assumption we will make connecting colleges' preferences over groups of students to their preferences over individual students is one insuring that, for example, if $\mu(C)$ assigns college C its third and fourth choice students, and

$\mu'(C)$ assigns it its second and fourth choice students, then college C prefers $\mu'(C)$ to $\mu(C)$. Specifically, let $P^{\#}(C)$ denote the preference relation of college C over all assignments $\mu(C)$ it could receive at some matching μ. A college C's preferences $P^{\#}(C)$ will be called *responsive* to its preferences $P(C)$ over individual students if, for any two assignments that differ in only one student, it prefers the assignment containing the more preferred student. That is, we assume colleges' preferences are responsive, as follows.

Definition 4. The preference relation $P^{\#}(C)$ over sets of students is *responsive* [to the preferences $P(C)$ over individual students] if, whenever $\mu'(C) = \mu(C) \cup \{s_k\} \setminus \{\sigma\}$ for σ in $\mu(C)$ and s_k not in $\mu(C)$, then C prefers $\mu'(C)$ to $\mu(C)$ [under $P^{\#}(C)$] if and only if C prefers s_k to σ [under $P(C)$].

We will write $\mu'(C) >_C \mu(C)$ to indicate that college C prefers $\mu'(C)$ to $\mu(C)$ according to its preferences $P^{\#}(C)$, and $\mu'(C) \geqslant_C \mu(C)$ to indicate that C likes $\mu'(C)$ at least as well as $\mu(C)$, where the fact that $\mu'(C)$ and $\mu(C)$ are not singletons will make clear that we are dealing with the preferences $P^{\#}(C)$, as distinct from statements about C's preferences over individual students. Note that C may be indifferent between distinct assignments $\mu(C)$ and $\mu'(C)$ even if C has strict preferences over individual students.

Note also that different responsive preference orderings $P^{\#}(C)$ exist for any preference $P(C)$, since, for example, responsiveness does not specify whether a college prefers to be assigned its first and fourth choice students instead of its second and third choice students. However, the preference ordering $P(C)$ over individual students can be derived from $P^{\#}(C)$ by considering a college C's preferences over assignments $\mu(C)$ containing no more than a single student (and $q_C - 1$ copies of C). The assumption that colleges have responsive preferences is essentially no more than the assumption that their preferences for sets of students are related to their ranking of individual students in a natural way. (Of course the assumption that colleges have preferences over individual students is nontrivial, and it is this assumption which is relaxed in Subsection 3.3.)

Some of the results that follow will depend on the assumption that agents have strict preferences. Surprisingly, we will only need to assume that colleges have strict preferences over individuals: it will not be necessary to assume they have strict preferences over groups of students. The reasons for this will not become completely clear until Corollary 17, which says that when colleges have strict preferences over individuals, they will not be indifferent between any groups of students assigned to them at stable matchings, even though they may be indifferent between other groups of students.

A matching μ is individually irrational if $\mu(s) = C$ for some student s and college C such that either the student is unacceptable to the college or the college is unacceptable to the student. We will say the unhappy agent can *improve upon* such a matching. Similarly, a college C and student s can improve upon a matching μ if they are not matched to one another at μ, but would both prefer to be matched to one another than to (one of) their present assignments. That is, the pair (C, s) can improve upon μ if $\mu(s) \neq C$ and if $C >_s \mu(s)$ and $s >_C \sigma$ for some σ in $\mu(C)$. [Note that σ may equal either some student s' in $\mu(C)$, or, if one or more of college C's positions is unfilled at $\mu(C)$, σ may equal C.] It should be clear that matchings blocked in this way by an individual or by a pair of agents are unstable in the sense discussed for the marriage model, since there are agents with both the incentive and the power to disrupt such matchings. So, as in the marriage model, we now define stable matchings – although we will immediately have to ask whether the set of stable matchings defined this way can serve the same role as in the marriage model.

Definition 5. A matching μ is *stable* if it cannot be improved upon by any individual agent or any college–student pair.

It is not obvious that this definition will still be adequate, since we now might need to consider coalitions consisting of colleges and several students (all of whom might be able to enroll simultaneously at the college), or even coalitions consisting of multiple colleges and students. However, when preferences are responsive, nothing is lost by concentrating on simple college–students pairs. The set of stable matchings is equal to the core defined by *weak* domination [Roth (1985b)].[6] So it is a subset of the core. To see why an outcome which is not strictly dominated might nevertheless be unstable, suppose college C with quota 2 is the first choice of students s_1, s_2, and s_3, and has preferences $P(C) = s_1, s_2, s_3$. Then a matching with $\mu(C) = \{s_1, s_3\}$ can be improved upon by (C, s_2), but the resulting match, $\mu'(C) = \{s_1, s_2\}$, involves a coalition of three agents, $\{C, s_1, s_2\}$, and s_1 is indifferent between μ and μ', since he is matched to C at both matchings.

We will defer until the next section the proof that the set of stable matchings is always nonempty, and contains optimal stable matchings for each side of the market. But note that if the preferences of the colleges for groups of students are not responsive (to some set of preferences over individual students), the core may be empty.

[6]A matching μ dominates another matching μ' if there is a coalition A of agents which is effective for μ – i.e. whose members can achieve their parts of μ by matching among themselves, without the participation of agents not in A – and such that all members of A prefer their matches under μ to those under μ'. In contrast, a matching μ *weakly* dominates another matching μ' if only some of the members of the effective coalition A prefer μ to μ', so long as no other members of A have the reverse preference. The core is the set of matchings that are not dominated, and the core defined by weak domination is the set of matchings that are not (even) weakly dominated.

3.3. Complex preferences over groups

Let the two sets of agents be n firms $\mathcal{F} = \{F_1, \ldots, F_n\}$, and m workers $W = \{w_1, \ldots, w_m\}$. For simplicity assume all firms have the same quota, equal to m, so each firm could in principle hire all the workers. This will allow us to describe matchings a little more simply, since it will not be necessary to keep track of each firm's quota by saying, for example, that a firm that does not employ any workers is matched to m copies of itself.

Definition 6. A *matching* μ is a function from the set $F \cup W$ into the set of all subsets of $\mathcal{F} \cup W$ such that:

(1) $|\mu(w)| = 1$ for every worker w and $\mu(w) = w$ if $\mu(w) \not\subseteq \mathcal{F}$;

(2) $|\mu(F)| \leq m$ for every firm F [$\mu(F) = \emptyset$ if F is not matched to any workers];

(3) $\mu(w) = F$ if and only if w is in $\mu(F)$.

Workers have preferences over individual firms, just as in the college admissions problem, and firms have preferences over subsets of W. For simplicity assume all preferences are strict. So a worker w's preferences can be represented by a list of acceptable firms, e.g. $P(w) = F_i, F_j, F_k, w$; and a firm's preferences by a list of acceptable subsets of workers, e.g. $P^{\#}(F) = S_1, S_2, \ldots, S_k, \emptyset$; where each S_i is a subset of W. Each agent compares different matchings by comparing his (its) own assignment at those matchings. The preferences of all the agents will be denoted by $\boldsymbol{P} = (P^{\#}(F_1), \ldots, P^{\#}(F_n), P(w_1), \ldots, P(w_m))$. Keep in mind that firms' preferences are over *sets* of employees.

Faced with a set S of workers, each firm F can determine which subset of S it would most prefer to hire. Call this F's choice from S, and denote it by $\mathrm{Ch}_F(S)$. That is, for any subset S of W, F's *choice set* is $\mathrm{Ch}_F(S) = S'$ such that S' is contained in S and $S' >_F S''$ for all S'' contained in S. Since preferences are strict, there is always a single set S' that F most prefers to hire, out of any set S of available workers. (Of course S' could equal S, or it could be empty.)

We will assume that firms regard workers as substitutes rather than complements, as follows.[7]

Definition 7. A firm F's preferences over sets of workers has the property of *substitutability* if, for any set S that contains workers w and w', if w is in $\mathrm{Ch}_F(S)$, then w is in $\mathrm{Ch}_F(S - w')$.

That is, if F has "substitutable" preferences, then if its preferred set of

[7]This kind of condition on preferences was proposed by Kelso and Crawford (1982).

employees from S includes w, so will its preferred set of employees from any subset of S that still includes w. [By repeated application, if $w \in \mathrm{Ch}_F(S)$, then for any S' contained in S with $w \in S'$, $w \in \mathrm{Ch}_F(S')$.] This is the sense in which the firm regards worker w and the other workers in $\mathrm{Ch}_F(S)$ more as substitutes than complements: it continues to want to employ w even if some of the other workers become unavailable.

So substitutability rules out the possibility that firms regard workers as complements, as might be the case of an American football team, for example, that wanted to employ a player who could throw long passes and one who could catch them, but if only one of them were available would prefer to hire a different player entirely. Note that responsive preferences have the substitutability property: in the college admissions model, the choice set from any set of students of a college with quota q is either the q most preferred acceptable students in the set, or all the acceptable students in the set, whichever is the smaller number.

A matching μ can be *improved upon* by an individual worker w if $w >_w \mu(w)$, and by an individual firm F if $\mu(F) \neq \mathrm{Ch}_F(\mu(F))$. Note that μ may be improved upon by an individual firm F without being individually irrational, since it might still be that $\mu(F) >_F \emptyset$. This definition reflects the assumption that workers' preferences are over firms (and not over coworkers), so that F may fire some workers in $\mu(F)$ if it chooses, without affecting other members of $\mu(F)$. Similarly, μ can be improved upon by a worker–firm pair (w, F) if w and F are not matched at μ but would both prefer if F hired w: i.e. if $\mu(w) \neq F$ and if $F >_w \mu(w)$ and $w \in \mathrm{Ch}_F(\mu(F) \cup w)$. If the firms have responsive preferences this is equivalent to the definition we used for the college admissions model. We define stable matchings the same way also.

Definition 8. A matching μ is *stable* if it cannot be improved by any individual agent or any worker–firm pair.

Since "improvement" is now defined in terms of firm's preferences over sets of workers, this definition of stability has a slightly different meaning than the same definition for the college admissions model. Nevertheless, it is still a definition of *pairwise* stability, since the largest coalitions it considers are worker–firm pairs. So we still have to consider whether something is missed by not considering larger coalitions. It turns out that pairwise stability is still sufficient: as when preferences are responsive, we can show that, for any preferences $\boldsymbol{P}$, the set $S(\boldsymbol{P})$ of stable matchings equals the core defined by weak domination, $C_W(\boldsymbol{P})$, and is always nonempty.

Theorem 3. *When firms have substitutable preferences (and all preferences are strict)* $S(\boldsymbol{P}) = C_W(\boldsymbol{P})$.

Theorem 4. *When firms have substitutable preferences, the set of stable matchings is always nonempty.*

The proof of Theorem 4 will be by means of the following algorithm:

In Step 1, each firm proposes to its most preferred set of workers, and each worker rejects all but the most preferred acceptable firm that proposes to it. In each subsequent step, each firm which received one or more rejections at the previous step proposes to its most preferred set of workers that includes all of those workers who it previously proposed to and have not yet rejected it, but does not include any workers who have previously rejected it. Each worker rejects all but the most preferred acceptable firm that has proposed so far. The algorithm stops after any step in which there are no rejections, at which point each firm is matched to the set of workers to which it has issued proposals that have not been rejected.

Proof of Theorem 4. The matching μ produced by the above algorithm is stable. The key observation is that, because firms have substitutable preferences, no firm ever regrets that it must continue to offer employment at subsequent steps of the algorithm to workers who have not rejected its earlier offers. That is, at every step in the algorithm each firm is proposing to its most preferred set of workers that does not contain any workers who have previously rejected it. So consider a firm F and a worker w not matched to F at μ such that $w \in \mathrm{Ch}_F(\mu(F) \cup w)$. At some step of the algorithm, F proposed to w and was subsequently rejected, so w prefers $\mu(w)$ to F, and μ is not improvable by the pair (w, F). Since w and F were arbitrary, and since μ is not improvable by any individual, μ is stable. □

We call this algorithm a "deferred acceptance" procedure, to emphasize that workers are able to hold the best offer they have received, without accepting it outright. For the moment we present this algorithm only to show that stable matchings always exist. That is, although the algorithm is presented as if at each step the firms and workers take certain actions, we will not consider until Section 5 whether they would be well advised to take those actions, and consequently whether it is reasonable for us to expect that they would act as described, if the rules for making and accepting proposals were as in the algorithm.

This result also establishes the nonemptiness of the set of stable matchings for the marriage and college admissions models, which are special cases of the present model. The algorithm and proof presented here are simple generalizations of those presented by Gale and Shapley (1962). And, as in the marriage and college admissions models, we can further note the surprising fact that the set of stable matchings contains elements of the following sort.

Definition 9. A stable matching is *firm optimal* if every firm likes it at least as well as any other stable matching. A stable matching is *worker optimal* if every worker likes it at least as well as any other stable matching.

Theorem 5 [Kelso and Crawford (1982)]. *When firms have substitutable preferences, and preferences are strict, the deferred acceptance algorithm with firms proposing produces a firm-optimal stable matching.*

Theorem 5 can be proved by showing that in the deferred acceptance algorithm with firms proposing, no firm is ever rejected by an *achievable* worker, where a worker w is said to be achievable for a firm F if there is some stable matching μ at which $\mu(w) = F$.

Since, unlike the marriage model and like the college admissions model, this model is not symmetric between firms and workers, it is not immediately apparent that a deferred acceptance algorithm with workers proposing will have an analogous result, but it does. In the algorithm with workers proposing, workers propose to firms in order of preference, and a firm rejects at any step all those workers who are not in the firm's choice set from those proposals it has not yet rejected. We can state the following result.

Theorem 6 [Roth (1984c)]. *When firms have substitutable preferences, and preferences are strict, the deferred acceptance algorithm with workers proposing produces a worker-optimal stable matching.*

The key observation for the proof is that, because firms have substitutable preferences, no firm ever regrets that it rejected a worker at an earlier step, when it sees who proposed at the current step. One can then show that no worker is ever rejected by an achievable firm.

These results cannot be generalized in a strightforward way to the symmetric case of many-to-many matching in which workers may take multiple jobs, even when both sides have substitutable, or even responsive, preferences.[8] The reason is not that the analogously defined pairwise stable matchings do not have similar properties in such a model, but that pairwise stable matchings are no longer always in the core.

Before moving on, an example will help clarify things.

Example 7. An example in which firms have substitutable (but nonresponsive) preferences. There are two firms and three workers, with preferences as follows.

[8]See Blair (1988) and Roth (1991).

$$P^{\#}(F_1) = \{w_1, w_2\}, \{w_1, w_3\}, \{w_2, w_3\}, \{w_3\}, \{w_2\}, \{w_1\} ,$$
$$P^{\#}(F_2) = \{w_3\} ,$$
$$P(w_1) = F_1, F_2 ,$$
$$P(w_2) = F_1, F_2 ,$$
$$P(w_3) = F_1, F_2 .$$

Note that

$$\mu = \begin{matrix} F_1 & F_2 \\ \{w_1, w_2\} & \{w_3\} \end{matrix}$$

is the unique stable matching.

If we look just at single workers, we see that F_1 prefers w_3 to w_2 to w_1. But $P^{\#}(F_1)$ is not responsive to these preferences over single workers, since $\{w_1, w_2\} >_{F_1} \{w_1, w_3\}$ even though w_3 alone is preferred to w_2 alone. But the preferences are substitutable. Recall the earlier discussion of why the college admissions model needed to be reformulated to include colleges' preferences over groups, and observe once again that the class of many-to-one matching problems, of which this is an example, would not be well-defined games if we specified only the preferences of firms over individuals. Indeed, if we defined stability only in terms of preferences over individuals, the matching μ would be unstable with respect to the pair (F_1, w_3) since w_3 prefers F_1 to F_2 and F_1 prefers w_3 (by himself) to w_2 (by himself). But μ is not unstable in this example because F_1 does not prefer $\{w_1, w_3\}$ to $\{w_1, w_2\}$. □

3.4. The assignment model

In this model money plays an explicit role. There are two finite disjoint sets of players P and Q, containing m and n players, respectively. Members of P will sometimes be called P-agents and members of Q called Q-agents, and the letters i and j will be reserved for P and Q agents, respectively. Associated with each possible partnership $(i, j) \in P \times Q$ is a non-negative real number α_{ij}. A game in coalitional function form with sidepayments is determined by (P, Q, α), with the numbers α_{ij} being equal to the worth of the coalitions $\{i, j\}$ consisting of one P agent and one Q agent. The worth of larger coalitions is determined entirely by the worth of the pairwise combinations that the coalition members can form. That is, the coalitional function v is given by

$v(S) = \alpha_{ij}$ if $S = \{i, j\}$ for $i \in P$ and $j \in Q$;

$v(S) = 0$ if S contains only P agents or only Q agents; and

$v(S) = \max(v(i_1, j_1) + v(i_2, j_2) + \cdots + v(i_k, j_k))$ for arbitrary coalitions S, with the maximum to be taken over all arrangements of $2k$ distinct players $i_1, i_2, \ldots, i_k$ belonging to S_P and $j_1, j_2, \ldots, j_k$ belonging to S_Q, where S_P and S_Q denote the sets of P and Q agents in S (i.e. the intersection of the coalition S with P and with Q), respectively.

So the rules of the game are that any pair of agents $(i, j) \in P \times Q$ can together obtain α_{ij}, and any larger coalition is valuable only insofar as it can organize itself into such pairs. The members of any coalition may divide among themselves their collective worth in any way they like. An imputation of this game is thus a non-negative vector (u, v) in $\boldsymbol{R}^m \times \boldsymbol{R}^n$ such that $\Sigma_{i\in P} u_i + \Sigma_{j\in Q} v_j = v(P \cup Q)$. The easiest way to interpret this is to take the quantities α_{ij} to be amounts of money, and to assume that agents' preferences are concerned only with their monetary payoffs.

We might think of P as a set of potential buyers of some objects offered for sale by the set Q of potential sellers, and each seller owns and each buyer wants exactly one indivisible object. If each seller j has a reservation price c_j, and each buyer i has a reservation price r_{ij} for object j, we may take α_{ij} to be the potential gains from trade between i and j; that is, $\alpha_{ij} = \max\{0, r_{ij} - c_j\}$. If buyer i buys object j from seller j at a price p, and if no other monetary transfers are made, the utilities are $u_i = r_{ij} - p$ and $v_j = p - c_j$. So, when no other monetary transfers are made, $u_i + v_j = \alpha_{ij}$ when i buys from j. But note that transfers between agents are *not* restricted to those between buyers and sellers; e.g. buyers may make transfers among themselves as in the bidder rings of Subsection 2.2.[9]

We can also think of the P and Q agents as being firms and workers, etc. As in the marriage model, we look here at the simple case of one-to-one matching, with firms constrained to hire at most one worker.[10] In such a case, the α_{ij}'s represent some measure of the joint productivity of the firm and worker, while transfers between a matched firm and worker represent salary. Transfers can also take place between workers (as when workers form a labor union in which the dues of employed members help pay unemployment benefits to unemployed members), or between firms.

The maximization problem to determine $v(S)$ for a given matrix α is called an *assignment problem*, so games of this form are called *assignment games*. We will be particularly interested in the coalition $P \cup Q$, since $v(P \cup Q)$ is the

[9]A model in which it is assumed that transfers cannot be made between agents on the same side of the market is considered by Demange and Gale (1985), who show that many of the results presented here for other models can be obtained in a model of this kind allowing rather general utility functions.

[10]The case of many-to-one matching has some important differences, analogous to those found between the marriage and college admissions models: see Sotomayor (1988).

maximum total payoff available to the players, and hence determines the Pareto set and the set of imputations.

Consider the following linear programming (LP) problem P_1:

$$\text{maximize} \sum_{i,j} \alpha_{ij} \cdot x_{ij}$$

$$\text{subject to (a)} \sum_i x_{ij} \leq 1 ,$$

$$\text{(b)} \sum_j x_{ij} \leq 1 ,$$

$$\text{(c)} \ x_{ij} \geq 0 .$$

We may interpret x_{ij} as, for example, the probability that a partnership (i, j) will form. Then the linear inequalities of type (a), one for each j in Q, say that the probability that j will be matched to some i cannot exceed 1. The inequalities of form (b), one for each i in P, say the same about the probability that i will be matched.

It can be shown [see Dantzig (1963, p. 318)] that there exists a solution of this LP problem which involves only values of zero and one. [The extreme points of systems of linear inequalities of the form (a), (b), and (c) have integer values of x_{ij}, i.e. each x_{ij} equals 0 or 1.] Thus the fractions artificially introduced in the LP formulation disappear in the solution and the (continuous) LP problem is equivalent to the (discrete) assignment problem for the coalition of all players, that is, the determination of $v(P \cup Q)$. Then $v(P \cup Q) = \Sigma\, \alpha_{ij} \cdot x_{ij}$, where x is an optimal solution of the LP problem.

Definition 10. A *feasible assignment* for (P, Q, α) is a matrix $x = (x_{ij})$ (of zeros and ones) that satisfies (a), (b) and (c) above. An *optimal assignment* is a feasible assignment x such that $\Sigma_{i,j}\, \alpha_{ij} \cdot x_{ij} \geq \Sigma_{i,j}\, \alpha_{ij} \cdot x'_{ij}$, for all feasible assignments x'.

So if x is a feasible assignment, $x_{ij} = 1$ if i and j form a partnership and $x_{ij} = 0$ otherwise. If $\Sigma_j\, x_{ij} = 0$, then i is *unassigned*, and if $\Sigma_i\, x_{ij} = 0$, then j is likewise unassigned. A feasible assignment x corresponds exactly to a matching μ as in Definition 1, with $\mu(i) = j$ if and only if $x_{ij} = 1$.

Definition 11. The pair of vectors (u, v), $u \in R^m$ and $v \in R^n$ is called a *feasible payoff* for (P, Q, α) if there is a feasible assignment x such that

$$\sum_{i \in P} u_i + \sum_{i \in Q} v_j = \sum_{i \in P, j \in Q} \alpha_{ij} \cdot x_{ij} .$$

In this case we say (u, v) and x are *compatible* with each other, and we call $((u, v); x)$ a *feasible outcome*. Note again that a feasible payoff vector may involve monetary transfers between agents who are not assigned to one another.

As in the earlier models, the key notion is that of stability.

Definition 12. A feasible outcome $((u, v); x)$ is *stable* [or the payoff (u, v) with an assignment x is stable] if

(i) $u_i \geqslant 0, v_j \geqslant 0$,

(ii) $u_i + v_j \geqslant \alpha_{ij}$ for all $(i, j) \in P \times Q$.

Condition (i) (individual rationality) reflects that a player always has the option of remaining unmatched (recall that $v(i) = v(j) = 0$ for all individual agents i and j). Condition (ii) requires that the outcome cannot be improved by any pair: if (ii) is not satisfied for some agents i and j, then it would pay them to break up their present partnership(s) (either with one another or with other agents) and form a new partnership together, because this could give them each a higher payoff.

From the definition of feasibility and stability it follows that

Lemma 8. *Let $((u, v), x)$ be a stable outcome for (P, Q, α). Then*

(i) $u_i + v_j = \alpha_{ij}$ *for all pairs (i, j) such that $x_{ij} = 1$;*

(ii) $u_i = 0$ *for all unassigned i, and $v_j = 0$ for all unassigned j at x.*

The lemma implies that at a stable outcome, the only monetary transfers that occur are between P and Q agents who are matched to each other. (Note that this is an implication of stability, not an assumption of the model.)

Now consider the LP problem P_1^* that is the dual of P_1, i.e. the LP problem of finding a pair of vectors (u, v), $u \in R^m$, $v \in R^n$, that minimizes the sum

$$\sum_{i \in P} u_i + \sum_{i \in Q} v_j$$

subject, for all $i \in P$ and $j \in Q$, to

(a*) $u_i \geqslant 0, v_j > 0$,

(b*) $u_i + v_j \geqslant \alpha_{ij}$.

Because we know that P_1 has a solution, we know also that P_1^* must have an optimal solution. A fundamental duality theorem [see Dantzig (1963, p. 129)] asserts that the objective functions of these dual LPs must attain the same value. That is, if x is an optimal assignment and (u, v) is a solution of P_1^*, we have that

$$\sum_{i \in P} u_i + \sum_{i \in Q} v_j = \sum_{P \times Q} \alpha_{ij} \cdot x_{ij} = v(P \cup Q) . \tag{1}$$

This means that $((u, v), x)$ is a feasible outcome. Moreover, $((u, v), x)$ is a stable outcome for (P, Q, α), since (a*) ensures individual rationality and $u_i + v_j \geqslant \alpha_{ij}$ for all $(i, j) \in P \times Q$ by (b*). It follows, by the definition of $v(S)$, that for any coalition $S = S_P \cup S_Q$, where S_P is contained in P and S_Q in Q,

$$\sum_{i \in S_P} u_i + \sum_{i \in S_Q} v_j \geqslant v(S) . \tag{2}$$

But (1) and (2) are exactly how the core of the game is determined: (1) ensures the feasibility of (u, v) and (2) ensures its nonimprovability by any coalition. Conversely, any payoff vector in the core, i.e. satisfying (1) and (2), satisfies the conditions for a solution to P_1^*. Hence we have shown

Theorem 9 [Shapley and Shubik (1972)]. *Let (P, Q, α) be an assignment game. Then*

(a) *the set of stable outcomes and the core of (P, Q, α) are the same;*

(b) *the core of (P, Q, α) is the (nonempty) set of solutions of the dual LP of the corresponding assignment problem.*

The following two corollaries make clear why, in contrast to the discrete models considered earlier, we can concentrate here on the payoffs to the agents rather than on the underlying assignment (matching).[11]

Corollary 10. *If x is an optimal assignment, then it is compatible with any stable payoff (u, v).*

Corollary 11. *If $((u,v), x)$ is a stable outcome, then x is an optimal assignment.*

In this model also there are optimal stable outcomes for each side of the market. Note that in view of Corollary 10, the difference between different stable outcomes in this model has to do only with the payments to each player, not to whom players are matched.

Theorem 12 [Shapley and Shubik (1972)]. *There is a P-optimal stable payoff $(\bar{u}, \underline{v})$, with the property that for any stable payoff (u, v), $\bar{u} \geqslant u$ and $\underline{v} \leqslant v$; there is a Q-optimal stable payoff $(\underline{u}, \bar{v})$ with symmetrical properties.*

[11]Becker (1981), who uses the assignment model to study marriage and household economics, makes use of the fact that stable outcomes all correspond to optimal assignments, and that the optimal assignment is typically unique, to study which men are matched to which women (e.g. if high wage earners marry good cooks), for different assumptions about how the assignment matrix is derived.

4. The structure of the set of stable matchings

In each of the models we have described, the set of stable matchings is nonempty.[12] In fact, for each side of the market, there exists an optimal stable matching that all agents on that side of the market like at least as well as any other stable matching.[13] That such side-optimal stable matchings always exist is more than a little surprising in models in which firms compete with one another for good workers, and workers compete with one another for desirable jobs. It turns out to be only the tip of the iceberg, in terms of the welfare comparisons which can be made between different stable matchings. In this section we describe some of these.

One question is whether the optimal stable matching for agents on one side of the market is Pareto optimal for them as well. This turns out to be one of the respects in which many-to-one matching is not equivalent to the special case of one-to-one matching. In the marriage model the optimal stable matching for each side of the market is weakly Pareto optimal for that side. (Since the market is symmetric between men and women, we consider here only the man-optimal stable matching μ_M.)

Theorem 13 [Roth (1982a)]. Weak Pareto optimality for the men. *In the marriage model there is no individually rational matching μ (stable or not) such that $\mu >_m \mu_M$ for all m in M.*

However, the following example shows that this result cannot be strengthened to strong Pareto optimality.

Example 14 [Roth (1982a)]. Let $M = \{m_1, m_2, m_3\}$ and $W = \{w_1, w_2, w_3\}$ with preferences over the acceptable people given by:

$$P(m_1) = w_2, w_1, w_3 ; \qquad P(w_1) = m_1, m_2, m_3 ;$$
$$P(m_2) = w_1, w_2, w_3 ; \qquad P(w_2) = m_3, m_1, m_2 ;$$
$$P(m_3) = w_1, w_2, w_3 ; \qquad P(w_3) = m_1, m_2, m_3 .$$

Then

$$\mu_M = \begin{matrix} w_1 & w_2 & w_3 \\ m_1 & m_3 & m_2 \end{matrix}$$

[12]This nonemptiness is related to the two-sidedness of the models: one-sided and three-sided models may have empty cores.

[13]For the discrete markets this is only the case when preferences are strict: when they are not, it is easy to see that although the set of stable matchings remains nonempty, it may not contain any such side-optimal matchings.

is the man-optimal stable matching. Nevertheless,

$$\mu = \begin{matrix} w_1 & w_2 & w_3 \\ m_3 & m_1 & m_2 \end{matrix}$$

leaves m_2 no worse than under μ_M, but benefits m_1 and m_3. So there may in general be matchings that all men like at least as well as the M-optimal stable matching, and that some men prefer. We shall return to this fact in our discussion of the strategic options available to coalitions of men.

Theorem 13 cannot be generalized to both sides of the college admissions model. We can state the following result instead.

Theorem 15 [Roth (1985a)]. *When the preferences over individuals are strict, the student-optimal stable matching is weakly Pareto optimal for the students, but the college-optimal stable matching need not be even weakly Pareto optimal for the colleges.*

However, as we have already seen through the existence of optimal stable matchings for each side of the market, there are some important properties concerning welfare comparisons within the set of stable matchings that hold both for one-to-one and many-to-one matching. There are also welfare comparisons that can be made in the case of many-to-one matching that have no counterpart in the special case of one-to-one matching.

We first consider some comparisons of this latter sort, for the college admissions model, concerning how well a given college might do at different stable matchings. Theorem 16 says that for every pair of stable matchings, each college will prefer every student who is assigned to it at one of the two matchings to every student who is assigned to it in the second matching but not the first. An immediate corollary is that in a college admissions problem in which all preferences over individuals are strict (and responsive), no college will be indifferent between any two (different) groups of students that it enrolls at stable matchings. The manner in which these results are mathematically unusual can be understood by noting that this corollary, for example, can be rephrased to say that if a given matching is stable (and hence in the core), and if some college is indifferent between the entering class it is assigned at that matching and a different entering class that it is assigned at a different matching, then the second matching is *not* in the core. We thus have a way of concluding that an outcome is not in the core, based on the direct examination of the preferences of only *one* agent (the college). Since the definition of the core involves preferences of coalitions of agents, this is rather unusual.

Theorem 16 [Roth and Sotomayor (1989)]. *Let preferences over individuals be strict, and let μ and μ' be stable matchings for a college admissions problem $(S, \mathscr{C}, \boldsymbol{P})$). If $\mu(C) >_C \mu'(C)$ for some college C, then $s >_C s'$ for all $s \in \mu(C)$ and $s' \in \mu'(C) - \mu(C)$. That is, C prefers every student in its entering class at μ* to every student who is in its entering class at μ' *but not at* μ.

Given that colleges have responsive preferences, the following corollary is immediate.

Corollary 17 [Roth and Sotomayor (1989)]. *If colleges and students have strict preferences over individuals, then colleges have strict preferences over those groups of students that they may be assigned at stable matchings. That is, if μ and μ' are stable matchings, then a college C is indifferent between $\mu(C)$ and $\mu'(C)$ only if $\mu(C) = \mu'(C)$.*

And since the set of stable matchings depends only on the preferences over individuals, and not on the preferences over groups (so long as these are responsive to the preferences over individuals) the following result is also immediate.

Corollary 18 [Roth and Sotomayor (1989)]. *Consider a college C with preferences $P(C)$ over individual students, and let $P^{\#}(C)$ and $P^*(C)$ be preferences over groups of students that are responsive to $P(C)$ (but are otherwise arbitrary). Then for every pair of stable matchings μ and μ', $\mu(C)$ is preferred to $\mu'(C)$ under the preferences $P^{\#}(C)$ if and only if $\mu(C)$ is preferred to $\mu'(C)$ under $P^*(C)$.*

An example will illustrate Theorem 16 and Corollaries 17 and 18.

Let the preferences over individuals be given by

$$P(s)_1 = C_5, C_1 ; \qquad P(C)_1 = s_1, s_2, s_3, s_4, s_5, s_6, s_7 ;$$
$$P(s)_2 = C_2, C_5, C_1 ; \qquad P(C)_2 = s_5, s_2 ;$$
$$P(s)_3 = C_3, C_1 ; \qquad P(C)_3 = s_6, s_7, s_3 ;$$
$$P(s)_4 = C_4, C_1 ; \qquad P(C)_4 = s_7, s_4 ;$$
$$P(s)_5 = C_1, C_2 ; \qquad P(C)_5 = s_2, s_1 ;$$
$$P(s)_6 = C_1, C_3 ;$$
$$P(s)_7 = C_1, C_3, C_4 ,$$

and let the quotas be $q_{C_1}=3$, $q_{C_j}=1$ for $j=2,\ldots,5$. Then the set of stable outcomes is $\{\mu_1, \mu_2, \mu_3, \mu_4\}$ where

	C_1	C_2	C_3	C_4	C_5
$\mu_1 =$	$s_1s_3s_4$	s_5	s_6	s_7	s_2
$\mu_2 =$	$s_3s_4s_5$	s_2	s_6	s_7	s_1
$\mu_3 =$	$s_3s_5s_6$	s_2	s_7	s_4	s_1
$\mu_4 =$	$s_5s_6s_7$	s_2	s_3	s_4	s_1

Note that these are the only stable matchings, and

$$\mu_1(C_1) >_{C_1} \mu_2(C_1) >_{C_1} \mu_3(C_1) >_{C_1} \mu_4(C_1),$$

for any responsive preferences.

We turn next to consider welfare comparisons involving more than one agent, on the set of stable matchings. Again, we concentrate primarily on the college admissions model. (The proofs all involve some version of Theorem 16.) However, these results [which are proved in Roth and Sotomayor (1990a)] all have parallels in the case of one-to-one matching, where they were first discovered.

We begin with a result which says that if an agent prefers one stable matching to another, then any agents on the other side of the market who are matched to that agent at either matching have the opposite preferences.[14]

Theorem 19. *If μ and μ' are two stable matchings for $(S, \mathscr{C}, P)$ and $C=\mu(s)$ or $C=\mu'(s)$, with $C\in\mathscr{C}$ and $s\in S$, then if $\mu(C)\geq_C \mu'(C)$ then $\mu'(s)\geq_s \mu(s)$* [and if $\mu'(s)>_s \mu(s)$ then $\mu(C)\geq_C \mu'(C)$].

The equivalent result for the assignment model, which is easy to prove, says that if i prefers a stable payoff (u, v) to another stable payoff (u', v'), his mate(s) will prefer (u', v').

Theorem 20. *Let $((u, v), x)$ and $((u', v'), x')$ be stable outcomes for (P, Q, α). Then if $x'_{ij}=1$, $u'_i>u_i$ implies $v'_j<v_j$.*

Proof. Suppose $v'_j\geq v_j$. Then $\alpha_{ij}=u'_i+v'_j>u_i+v_j\geq\alpha_{ij}$, which is a contradiction.

[14] The case of the marriage model was shown by Knuth (1976), and an extended version of this result was given by Gale and Sotomayor (1985a), who show its usefulness as a lemma in a number of other proofs.

The next result concerns the common preferences of agents on the same side of the market. Stated here for the college admissions model, it also holds for the assignment model. We write $\mu >_{\mathscr{C}} \mu'$ to mean that every college likes μ at least as well as μ', and some college strictly prefers μ, i.e. $\mu(C) \geqslant_C \mu'(C)$ for all $C \in \mathscr{C}$ and $\mu(C) >_C \mu'(C)$ for some $C \in \mathscr{C}$. So the relation $>_{\mathscr{C}}$ represents the common preferences of the colleges, and we define the relation $>_S$ analogously, to represent the common preferences of the students. The relations $>_{\mathscr{C}}$ and $>_S$ are only partial orders on the set of stable matchings, which is to say that there may be stable matchings μ and μ' such that neither $\mu >_S \mu'$ nor $\mu' >_S \mu$. An additional definition will help us summarize the state of affairs.

Definition 13. A *lattice* is a partially ordered set L any two of whose elements x and y have a "sup", denoted by $x \vee y$ and an "inf", denoted by $x \wedge y$. A lattice L is *complete* when each of its subsets X has a "sup" and an "inf" in L.

Hence, any nonempty complete lattice P has a least element and a greatest element. The next result therefore accounts for the existence of optimal stable matchings for each side of the market.

Theorem 21. *When all preferences over individuals are strict, the set of stable matchings in the college admissions model is a lattice under the partial orders* $>_{\mathscr{C}}$ *and* $>_S$. *Furthermore, these two partial orders are duals: if* μ *and* μ' *are stable matchings for* $(S, \mathscr{C}, \boldsymbol{P})$, *then* $\mu >_{\mathscr{C}} \mu'$ *if and only if* $\mu' >_S \mu$.

This theorem provides a more complete description of those structural properties of the set of stable matchings that account for the existence of optimal stable matchings for each side of the market. And the theorem shows that the optimal stable matching for one side of the market is the worst stable matching for the other side. Knuth (1976) attributes the lattice result for the marriage model to J.H. Conway. Shapley and Shubik (1972) established the same result for the assignment model.

4.1. Size of the core

Knuth (1976) examined the computational efficiency of the deferred acceptance procedure for the marriage model, and observed that the task of computing a single stable matching is not computationally onerous (it can be completed in polynomial time). However, even in the marriage model, the task of computing *all* the stable matchings can quickly become intractable as the size of the problem grows, for the simple reason that the number of stable matchings can grow exponentially. The next result, which follows a construction found in Knuth (1976), describes the case of a marriage problem in which there are n men and n women, which we will speak of as a problem of size n.

Theorem 22 [Irving and Leather (1986)]. *For each $i \geq 0$ there exists a stable marriage problem (M, W, P) of size $n = 2^i$ with at least 2^{n-1} stable matchings.*

However, because of the special structure of the core in these games, we can answer some questions about the core without computing all its elements. For example, suppose we simply wish to know which pairs of agents may be matched to one another at some stable matching, i.e. which pairs of agents are achievable for one another. The following result says that these can be found by following any path through the lattice from the man-optimal stable matching μ_M to the woman optimal stable matching μ_W.

Theorem 23 [Gusfield (1985)]. *Let $\mu_M = \mu_0 >_M \mu_1 >_M \mu_2 >_M \cdots >_M \mu_t = \mu_W$ be a sequence of stable matchings encountered on any path through the lattice of stable matchings of a marriage problem. Then every achievable pair appears in at least one of the matchings in the sequence.*

For the assignment model, since the core is a convex polyhedron we cannot ask how many elements it contains, but we can ask how many extreme points it might have. We can state the following result.

Theorem 24 [Balinski and Gale (1990)]. *In the assignment game, the core has at most $\binom{2m}{m}$ extreme points, where $m = \min\{|P|, |Q|\}$.*

4.2. *The linear structure of the set of stable matchings in the marriage model*

That the marriage and assignment models share so many properties has been a long-standing puzzle, since many of these results (e.g. the existence of optimal stable outcomes for each side of the market, and the lattice structure throughout the set of stable outcomes) require the assumption of strict preferences in the marriage model, while in the assignment model all admissible preferences must allow agents to be indifferent between different matches if prices are adjusted accordingly.[15] However, a structural similarity between the two models is seen in the rather remarkable result of Vande Vate (1989) that

[15] Roth and Sotomayor (1990b) show, however, that the two sets of results can be derived under common assumptions if one requires merely that the core defined by weak domination coincides with the core.

finding the stable matchings in the marriage model can also be represented as a linear programming problem.[16] The argument proceeds by first showing that the problem can be phrased as an integer program, and then observing that when the integer constraints are relaxed, the problem nevertheless has integer solutions.

For simplicity consider the special case in which $|M| = |W|$ and every pair (m, w) is mutually acceptable, and all preferences are strict. Thus, every man is matched to some woman and vice versa, under any stable matching. Let the *configuration* of a matching μ be a matrix x of zeros and ones such that $x_{mw} = 1$ if $\mu(m) = w$ and $x_{mw} = 0$ otherwise.

We will also consider matrices x of dimension $|M| \times |W|$ the elements of which may not be integers, i.e. matrices which may not be the configuration of any matching. Let $\Sigma_i x_{iw}$ denote the sum over all i in M, $\Sigma_j x_{mj}$ denote the sum over all j in W, $\Sigma_{j_m > w} x_{mj}$ denote the sum over all those j in W that man m prefers to woman w, and $\Sigma_{i_w > m} x_{iw}$ denote the sum over all those i in M that woman w prefers to man m.

We can characterize the set of stable matchings by their configurations:

Theorem 25 (Vande Vate). *A matching is stable if and only if its configuration x is an* integer *matrix of dimension $|M| \times |W|$ satisfying the following set of constraints*:

(1) $\Sigma_j x_{mj} = 1$ *for all m in M*,

(2) $\Sigma_i x_{iw} = 1$ *for all w in W*,

(3) $\Sigma_{j_m > w} x_{mj} + \Sigma_{i_w > m} x_{iw} + x_{mw} \geq 1$ *for all m in M and w in W*,

and

(4) $x_{mw} \geq 0$ *for all m in M and w in W.*

Constraints (1), (2) and (4) require that if x is integer it is the configuration of a matching, i.e. its elements are 0's and 1's and every agent on one side is matched to some agent on the opposite side. It is easy to check that constraint (3) is equivalent to the nonexistence of blocking pairs. [To see this, note that if x is a matching, i.e. a matrix of 0's and 1's satisfying (1), (2), and (4), then (3) is not satisfied for some m and w only if $\Sigma_{j_m > w} x_{mj} = \Sigma_{i_w > m} x_{iw} = x_{mw} = 0$, in which case m and w form a blocking pair.]

Thus, an integer $|M| \times |W|$ matrix x is the configuration of a stable matching if and only if x satisfies (1)–(4). Of course there will in general be an infinite set of noninteger solutions of (1)–(4) also, and these are not matchings. However, we may think of them as corresponding to "fractional matchings", in which x_{mw} denotes something like the fraction of the time man m and woman w are matched, or the probability that they will be matched.

[16] Subsequent, simpler proofs are found in Rothblum (1992) and Roth, Rothblum and Vande Vate (1992).

The surprising result is that the integer solutions of (1)–(4), i.e. the stable matchings, are precisely the extreme points of the convex polyhedron defined by the linear constraints (1)–(4). That is, we have the following result.

Theorem 26 (Vande Vate). *Let C be the convex polyhedron of solutions to the linear constaints* (1)–(4). *Then the integer points of C are precisely its extreme points. That is, the extreme points of the linear constraints* (1)–(4) *correspond precisely to the stable matchings.*

4.3. Comparative statics: New entrants

The results of this subsection concern the effect of adding a new agent to the market. Following Kelso and Crawford (1982), who established the following result for a class of models including the assignment model, a number of authors have examined the effect on the optimal stable matchings for each side of the market of adding an agent on one side of the market. Briefly, the results are that, measured in this way, agents on opposite sides of the market are complements, and agents on the same side of the market are substitutes.[17] This result seems to be robust, with a recent paper by Crawford (1988) establishing the result for a general class of models with substitutable preferences introduced in Roth (1984c). As it applies to the simple model with substitutable preferences described in Subsection 3.3, his result is the following.

Theorem 27 [Crawford (1991)]. *Suppose $\mathcal{F}$ is contained in $\mathcal{F}^*$ and μ_W and μ_F are the W-optimal and F-optimal stable matchings, respectively, for a market with substitutable preferences $(W, \mathcal{F}, \boldsymbol{P})$ and let μ_W^* and μ_F^* be the W- and F-optimal stable matchings for $(W, \mathcal{F}^*, \boldsymbol{P}^*)$, where $\boldsymbol{P}^*$ agrees with $\boldsymbol{P}$ on $\mathcal{F}$. Then*

$$\mu_W^* \geqslant_W \mu_W \text{ and } \mu_F^* \geqslant_W \mu_F \text{ under } \boldsymbol{P}^*; \text{ and } \mu_W \geqslant_F \mu_W^*, \text{ and } \mu_F \geqslant_F \mu_F^* .$$

Symmetrical results are obtained if S is contained in S^*.

The next result, which we state for the assignment model, shows that when a new agent enters the market there will be some P and Q agents for whom we can unambiguously compare *all* stable outcomes of the two markets.[18] Suppose some P agent i^* enters the market $M = (P, Q, \alpha)$. The new market is then $M^{i^*} = (P \cup \{i^*\}, Q, \alpha')$, where $\alpha'_{ij} = \alpha_{ij}$ for all $i \in P$ and $j \in Q$.

[17] Cf. Shapley (1962) for a related linear programming result.
[18] A similar result for the marriage market is given in Roth and Sotomayor (1990a).

Theorem 28. Strong dominance [Mo (1988)]. *If i^* is matched under some optimal assignment for M^{i*}, then there is a nonempty set A of agents in $P \cup Q$ such that every Q agent in A is better off and every P agent in A is worse off at any stable outcome of the new market than at any stable outcome of the old market. That is, for all (u', v') and (u, v) stable for M^{i*} and M, respectively, we have*

(a) *if a P agent i is in A, then* $u_i \geq u'_i$;
(b) *if a Q agent j is in A, then* $v_j \leq v'_j$.

The final result of this subsection can be thought of as describing how much the entry of an agent i^* in the assignment model can move the core of the game. There will be some agents whose worst core payoff in one of the two games (with and without i^*) is exactly equal to their best core payoff in the other.

Corollary 29 [Mo (1988)]. *Let $(\bar{u}', \underline{v}')$ be the P-optimal stable payoff for M^{i*}. Let $(\underline{u}, \bar{v})$ be the Q-optimal stable payoff for M. If i^* is matched under some optimal assignment for M^{i*}, there exists a nonempty set A of agents in $P \cup Q$ such that*

(a) *if a P agent i is in A, then* $\bar{u}'_i = \underline{u}_i$;
(b) *if a Q agent j is in A, then* $\underline{v}'_j = \bar{v}_j$.

5. Strategic results

We now turn to a different class of questions, motivated by the claim made in the literature distributed to participants in the hospital-intern market that the NIMP algorithm makes it unprofitable for either students or hospitals to state anything other than their true preferences. While we will defer consideration of the NIMP algorithm itself until Section 6, we consider here the extent to which it is possible to minimize the strategic complexity of matching, and what can be said about the strategic properties of procedures which lead to stable matchings.

To set the stage, consider the procedure by which graduating students at the United States Naval Academy obtain their first posts as Naval officers. The following description is taken from the *New York Times* (30 January 1986, p. 8).

> Midshipmen who will graduate from the Naval Academy in June decided this week whether they wanted to be aviators or nuclear submariners, destroyermen or engineers, marines or oceanographers . . . . From late Thursday afternoon through the wee hours of Friday morning, the first classmen, or seniors, lined up according to their standing in the class, walked

up to a long table lined with officers from each specialty, and made their choices on a first-come, first-served basis . . . .

It is easy to see that each agent in this procedure has a dominant strategy, since a student can do no better than to select his first choice of those specialties remaining when his turn comes, and since the various Naval specialties have no choices of any sort to make. And if the preferences of each specialty over the students correspond exactly to students' class standings, then the matching which results from this procedure is stable. Of course if any of the specialties have different preferences, the matching may not be stable, but the rules by which the Navy is run do not permit specialties to refuse positions to some students and offer positions to students they prefer but who have lower class standings.[19]

However, in markets that allow the agents on the two sides of the market to freely negotiate with one another, the empirical evidence suggests that the stability or instability of the final matching is important. So we will want to consider whether any procedures exist which yield stable matchings for all preferences, and which give each agent a dominant strategy. It will be sufficient for this purpose to confine our attention to the special class of "revelation mechanisms" which are functions from the stated preferences of the agents to the set of matchings. We will call a revelation mechanism which always chooses a matching that is stable with respect to the stated preferences a *stable matching mechanism*.[20] If any procedures with the desired properties exist, then there will exist a revelation mechanism which is a stable mechanism and which makes it a dominant strategy for each agent to state his true preferences.[21]

The next theorem states that no such mechanism exists for the marriage model. Since the marriage model is a special case of the college admissions and substitutable preferences models, the theorem implies that no such mechanism exists for those models either.[22]

Theorem 30. Impossibility Theorem [Roth (1982a)]. *No stable matching mechanism for the marriage model exists for which stating the true preferences is a dominant strategy for every agent.*

[19]That is, in the Navy's game, the outcome of this procedure is in the core, even if it can be improved upon by some student–specialty pair, since the rules do not permit the specialties to be active players.

[20]Note that the Naval Academy procedure just described can be thought of as a revelation mechanism, albeit one in which the preferences of the specialties are ignored. However, it is not a stable matching mechanism, since although it produces a stable matching for some preferences, there are (many) preferences for which the matching it produces is unstable.

[21]Various formalizations of this observation go under the name of the *revelation principle*, and are widely used in game-theoretic proofs.

[22]Notice that impossibility theorems are strongest when stated on the narrowest domain, since if no mechanism exists which works for all examples of the narrow domain, then certainly no mechanism exists which works for all examples of wider domains.

Proof*. Since a matching mechanism is a function that produces a matching for *any* stated preferences, to prove the theorem it is sufficient to demonstrate some particular marriage market such that, for any stable matching mechanism, truthtelling will not be a dominant strategy for all agents. So consider a market with two men and two women, with preferences $\boldsymbol{P}$ given by $P(m_1) = w_1, w_2$; $P(m_2) = w_2, w_1$; $P(w_1) = m_2, m_1$; $P(w_2) = m_1, m_2$. Then there are two stable matchings, μ and ν, given by $\mu(m_i) = w_i$ for $i \in \{1, 2\}$, and $\nu(m_i) = w_j$ for $i, j \in \{1, 2\}$, $j \neq i$. So any stable mechanism must choose one of μ or ν when preferences $\boldsymbol{P}$ are stated: suppose the mechanism chooses μ. Observe that if w_2, say, changes her stated preference from $P(w_2)$ to $Q(w_2) = m_1$ while everyone else states their true preferences, then ν is the only stable matching with respect to the stated preferences $\boldsymbol{P}' = (P(m_1), P(m_2), P(w_1), Q(w_2))$, and so any stable mechanism must select ν when the stated preferences are $\boldsymbol{P}'$. So it is not a dominant strategy for all agents to state their true preferences, since w_2 does better to state $Q(w_2)$. Similarly, if the mechanism chooses ν when the preferences $\boldsymbol{P}$ are stated, then m_2 can profitably mis-state his preferences. □

The same result can be stated for the assignment model.

Since we have defined a matching mechanism as a procedure which can be applied to any marriage market (i.e. as a function defined for all marriage markets), the Impossibility Theorem says we cannot find a stable mechanism that will not *sometimes* give some agent an incentive to mis-state his or her preferences. But we might hope to find a stable matching mechanism that only seldom gave agents such incentives, in which case the problem of incentives might not be very important. The following result, which can be thought of as a corollary of the proof of the Impossibility Theorem, and which strengthens it, states that no such mechanism can be found. Instead, that at least one agent will have incentive to behave strategically seems to be the usual case.

Corollary 31 [Roth and Sotomayor (1990a)]. *When any stable mechanism is applied to a marriage market in which preferences are strict and there is more than one stable matching, then at least one agent can profitably misrepresent his or her preferences, assuming the others tell the truth. (This agent can misrepresent in such a way as to be matched to his or her most preferred achievable mate under the true preferences at every stable matching under the mis-stated preferences.)*

*Alcade and Barbera (1991) have strengthened the impossibility theorem by observing that there exists no efficient and individually rational matching mechanism for the marriage model, for which stating the true preferences is a dominant strategy for every agent.

The proof of Corollary 31 depends on demonstrating the following. Suppose a stable mechanism selects a point different from the W-optimal stable matching μ_W, say. Then a woman w who prefers μ_W can profitably misrepresent her preferences by removing from her stated preference list of acceptable men all men who rank below $\mu_W(w)$ (as in the proof of the Impossibility Theorem). Similarly, if the mechanism selects a point different from the M-optimal stable matching, some man can profitably misrepresent his preferences.

The Impossibility Theorem and the parallel result for the assignment model tell us that in each of the models considered here there will be no way to organize the market so as to achieve a stable matching without sometimes presenting at least some of the agents with nontrivial strategic decisions. And Corollary 31 shows that only in rare cases will it be an equilibrium for all agents to state their true preferences. So we turn next to investigating which agents may have incentives to misrepresent their preferences, and what equilibrium behavior looks like, as a function of how the market is organized.

It was observed in Roth (1985a) that the answers to these questions differ in important ways depending on whether we are considering one-to-one or many-to-one matching, and so we shall deal with these two cases separately. We begin with our models of one-to-one matching, namely the marriage and assignment models.

5.1. Strategic behavior in models of one-to-one matching

The first result for the marriage model states that the incentive to state other than true preferences can be confined to the agents on one or the other side of the market.

Theorem 32 [Dubins and Freedman (1981), Roth (1982a)]. *In the marriage model, the mechanism that yields the M-optimal stable matching (in terms of the stated preferences) makes it a dominant strategy for each man to state his true preferences. (Similarly, the mechanism that yields the W-optimal stable matching makes it a dominant strategy for every woman to state her true preferences.)*

To place in context the parallel result for the assignment model, it will be helpful to first consider the case in which there is only a single agent on one side of the market. This can be thought of as a market consisting of a single seller, who owns one unit of an indivisible object, and n buyers, each of whom is interested in purchasing it. Each buyer b places a monetary value $\$r_b$ on the object, which is the maximum amount he is willing to pay, and the seller similarly places a value $\$r_s$ on the object, which is the price below which he will not sell. We will call these monetary values the *reservation prices* of the agents.

An example of this market is given by the vector of reservation prices $r = (r_1, \ldots, r_n, r_{n+1})$. In order to characterize the core and the set of stable payoff vectors, it will be convenient to define a *reordering* of the players, $1^*, 2^*, \ldots, n+1^*$ so that $r_{1^*} \geq r_{2^*} \geq \cdots \geq r_{n+1^*}$. That is, under this alternative ordering, player 1^* is that player in N who has the highest reservation price (or one of the highest, if there is a tie), and $n+1^*$ is the player with the lowest reservation price.

It is straightforward to verify that the core of this game (which by Theorem 9 equals the set of stable outcomes) corresponds to those transactions in which the object is sold to the agent with the highest reservation price, at a price between the highest and second highest reservation prices.[23] So at the seller-optimal stable outcome the price equals r_{1^*} and at the buyer-optimal stable outcome it equals r_{2^*}. It is easy to see why the Impossibility Theorem applies to this model (and therefore to the general assignment model as well), since if the seller does not have the highest reservation price, he can raise his payoff by stating a reservation price equal to the highest stated reservation price whenever the seller optimal outcome is not chosen, and similarly, if the buyer optimal outcome is not chosen, the buyer with the highest reservation price can profit by lowering his stated price to just above the second highest stated price.

As in the marriage model, however, it is possible to make it a dominant strategy for either side of the market to state true reservation prices, by using the mechanism that always selects the optimal core outcome for that side. We concentrate here on the mechanism that, for any stated reservation prices, chooses the buyer optimal core outcome. This is a well-known mechanism, variants of which are used in the auction of some U.S. government securities, for example. It is called the *sealed-bid, second-price auction* mechanism, and can be thought of as follows: each buyer writes down a number (his bid, or stated reservation price) in an envelope, without knowing what number will be written down by any other buyer. The seller also writes down a number. All the envelopes are opened, and placed in order $r_{1^*} \geq \cdots \geq r_{n+1}$, with the seller being player 1^* only if his number is strictly greater than all the others, in which case there is no sale. Otherwise, buyer 1^* receives the object, and pays the seller the price $p = r_{2^*}$. This mechanism is sometimes also called a *Vickrey auction*, after the economist who first observed the following result in a celebrated paper.

Theorem 33 [Vickrey (1961)]. *In a second-price, sealed bid auction (which always yields the buyer optimal core outcome in terms of the stated reservation prices), it is a dominant strategy for every buyer to state his true reservation price.*

[23] Unless the seller has the highest reservation price, in which case there is no sale. By Lemma 8 no monetary transfers other than the transfer of the selling price can take place in the core.

Proof. Consider a buyer b who states his true reservation price r_b, resulting in a vector $\boldsymbol{r}$ of stated reservation prices. Given the stated reservation prices of the others, b could not have helped himself, and could have hurt himself, if he had instead stated some reservation price different from the true one. If $b = 1^*$ with respect to these stated prices, i.e. if his true reservation price is the highest stated price, then he gets the object at price $p = r_{2^*}$, which gives him a positive profit whenever r_{2^*} is strictly less than $r_b = r_{1^*}$. If he had stated a different reservation price, the outcome would not change at all so long as his stated price remains above r_{2^*}. But if he states a reservation price $r'_b < r_{2^*}$ (where by 2^* we still mean the player with the second highest of the original reservation prices), buyer b will forgo his profit, and receive a payoff of 0. (What happens when $r_b = r_{2^*}$ depends on what tie-breaking rule is used, but does not change the argument.) Now suppose that $b \neq 1^*$. Then b receives a payoff of 0, and would continue to do so for any stated preference $r'_b \leq r_{1^*}$. The only way b can change his payoff is by stating a reservation price $r'_b > r_{1^*}$, but in this case he buys the object at a price greater than his true reservation price, which gives him a negative profit. So it is a dominant strategy for each buyer to state his true reservation price.[24] □

This brings us back to the case of the general assignment model. The following lemma shows a critical way in which the Vickrey second price auction is generalized by the mechanism which gives P agents their optimal stable outcome $(\bar{u}, \underline{v})$. Just as the second price auction gives the winning buyer his marginal contribution $r_{1^*} - r_{2^*}$ (and gives each other buyer his marginal contribution, which is 0), the P-optimal stable mechanism gives each P agent his marginal contribution.

Lemma 34 [Demange (1982), Leonard (1983)]. *For all i in P, $\bar{u}_i = v(P, Q) - v(P - \{i\}, Q)$.*

This permits the following parallel to Theorem 32.

Theorem 35 [Demange (1982), Leonard (1983)]. *The mechanism that yields the P-optimal stable outcome $(\bar{u}, \underline{v})$ makes truthtelling a dominant strategy for each P agent.*

[24]Note that an important feature of this mechanism is that the price stated by a bidder determines if he is the winner, but does not determine the price he pays (as it would in a conventional first-price sealed bid auction in which the high bidder 1* pays r_{1^*}). Of course, this is not the whole argument: a useful exercise for the reader to check that he has understood is to consider why a *third-price* sealed bid auction, i.e. one at which buyer 1* receives the object but pays price r_{3^*}, does not make it a dominant strategy for each buyer to state his true reservation price.

Returning to the marriage model, we state the following two theorems which strengthen and amplify Theorem 32.

Theorem 36 [Dubins and Freedman (1981)]. *Let* $\boldsymbol{P}$ *be the true preferences of the agents, and let* $\bar{\boldsymbol{P}}$ *differ from* $\boldsymbol{P}$ *in that some coalition* $\bar{M}$ *of the men mis-state their preferences. Then there is no matching* μ*, stable for* $\bar{\boldsymbol{P}}$*, which is preferred to* μ_M *by all members of* $\bar{M}$.

The original proofs of Theorems 32 and 36 in Roth (1982a) and Dubins and Freedman (1981) were rather lengthy. A short proof of the following result, gives a much shorter proof of those two theorems.

Theorem 37. Limits on successful manipulation [Demange, Gale and Sotomayor (1987)]. *Let* $\boldsymbol{P}$ *be the true preferences (not necessarily strict) of the agents, and let* $\bar{\boldsymbol{P}}$ *differ from* $\boldsymbol{P}$ *in that some coalition C of men and women mis-state their preferences. Then there is no matching* μ*, stable for* $\bar{\boldsymbol{P}}$*, which is preferred to* every *stable matching under the true preferences* $\boldsymbol{P}$ *by all members of C.*

To see that Theorem 37 will provide a proof of Theorems 32 and 36, consider the special case where all the coalition members are men. Then Theorem 37 implies that no matter which stable matching with respect to $\bar{\boldsymbol{P}}$ is chosen, at least one of the liars is no better off than he would be at the M-optimal matching under $\boldsymbol{P}$.[25]

Note also that Theorem 36 implies Theorem 32. Initially Theorem 36 was sometimes further interpreted as stating that no *coalition* of men could profitably misrepresent their preferences in one-to-one matching situations of the kind modelled by the marriage model, when an M-optimal stable mechanism was employed. That this is not a robust interpretation can be seen by re-examining Example 14, and observing that if man m_2 in that example were to misrepresent his preferences by listing w_3 as his first choice, then the M-optimal stable matching with respect to the stated preferences $\boldsymbol{P}'$ in which all agents but m_2 state their true preferences is equal to μ. That is, if m_2 misrepresents his preferences in this way under an M-optimal stable matching mechanism, then the resulting matching is $\mu'_M = \mu$ instead of μ_M. So m_2 is able to help the other men at no cost to himself. Note, however, that if there were any way at all in which the other men could pay m_2 for his services, then it

[25] When preferences are not strict, there may of course not be an M-optimal stable matching, and so we have to rephrase Theorem 32 to avoid speaking of *the* M-optimal stable mechanism. Instead, we can consider the deferred acceptance procedure with men proposing, and with a tie-breaking procedure.

would be possible for a coalition of men to form and collectively profit from this misrepresentation. Since m_2 receives the same mate at both matchings, presumably even a very small payment would make it worth his while to become part of a coalition to change the final outome from μ_M to μ, and since the gains to the other men in this coalition might be substantial, there would be ample motivation for such a coalition to form. Thus the negative implications of Theorem 36 (and also of Theorem 37) for strategic behavior by coalitions depend on the fact that, in the model of the marriage market that we are working with, we have assumed that no possibility whatsoever exists for such "sidepayments" between agents.[26] If this assumption is relaxed even a little, we see that coalitions of men can profitably manipulate even the M-optimal stable mechanism. We turn next to consider this in detail for the case of one seller and many buyers considered in connection with Theorem 33.

It is clear in that model that a *coalition* of bidders may be able, by suppressing some bids, to lower the price at which the object is sold in a second-price, sealed bid auction, or an ascending bid auction[27] (or for that matter in virtually any kind of auction). We will concentrate here on the second-price, sealed bid auction. Consider a vector $\boldsymbol{r}$ of reservation prices for which the seller's reservation price is strictly less than the second highest, so that the sale price, $p = r_{2^*}$, is greater than the seller's (auctioneer's) reservation price. Suppose the seller has the $(k+1)$st highest reservation price, i.e. the seller is player $(k+1)^*$. Then the coalition consisting of bidders 1^* through k^* can, by suppressing $k-1$ bids (or submitting only one bid greater than the seller's reservation price), obtain the object at price $p' = r_{k+1^*} < r_{2^*}$.

Of course, if this was the end of the matter, the buyer who took possession of the object would benefit, but his co-conspirators would not. However, there is money in this model, so the k members of the coalition can share the wealth, for example by having a subsequent auction among themselves, with the proceeds distributed among the coalition members. Thus it is possible for a

[26] We have also assumed that each agent is concerned only with his own mate at any matching, and not with the mates of any other agents, and that the game is played only once, so that there is no possibility of a coalition forming to trade favors over time. In Subsection 5.3 we will also see how this result breaks down if we relax the assumption of complete information.

[27] One reason the second-price, sealed bid auction is of great interest is because of the relationship it has to the more commonly observed *ascending bid* (also called "English") auctions, in which the auctioneer keeps raising the price so long as two or more bidders indicate that they are still interested, and stops as soon as the next-to-last bidder drops out of the bidding. At that point the sale is made to the remaining bidder at the price at which the next to last bidder dropped out. (If the price at which the next to last bidder drops out is lower than the auctioneer's reservation price, the auctioneer acts as if there were a bidder who continued bidding until the auctioneer's reservation price is reached.) Suppose for simplicity that the bidders cannot see which other bidders are still bidding: then the problem facing a bidder b in this auction is simply to decide at what price to drop out of the auction. So in this case these two auctions are *strategically equivalent*, and the incentives facing the players are the same.

coalition of bidders acting together (a "bidder ring") to profit from understating their bids and sharing the benefits among themselves, even though it is not possible for a single bidder acting alone to do better than to state his true reservation price.

Note how this compares with the results for the marriage model. In both models it is a dominant strategy for an individual agent to state his true preferences when his choice consists of what preferences to state to the stable mechanism that chooses the optimal stable outcome for his side. In both models, no coalition of these agents may, by mis-stating their preferences, arrange so that they all do better under such a mechanism than when they all state their true preferences, *unless they are able to make sidepayments within the coalition*. That is, the conclusion of Theorem 36 is true in this model as well: if some coalition of bidders mis-states its reservation prices so that the vector of reservation prices is $\bar{r}$ instead of r, then there is no outcome *in the core with respect to* $\bar{r}$ that all members of the ring prefer to the result of truthful revelation. This is because no money other than the purchase price is transferred at core outcomes. But, as we have just seen, a coalition can profit by understating its preferences and then making sidepayments among its members.

Having gotten some idea of what can be said about dominant strategies and the limits on how much an individual agent can manipulate a stable mechanism, and what possibilities are open to coalitions, we now turn to the questions associated with equilibrium behavior.

5.1.1. Equilibrium behavior

The first result suggests that we may see matchings that are stable with respect to the true preferences even when agents do not state their true preferences.

Theorem 38 [Roth (1984b)]. *Suppose each man chooses his dominant strategy and states his true preferences, and the women choose any set of strategies (preference lists) $P'(w)$ that form an equilibrium for the matching game induced by the M-optimal stable mechanism. Then the corresponding M-optimal stable matching for $(M, W, \boldsymbol{P}')$ is one of the stable matchings of $(M, W, \boldsymbol{P})$.*

Theorem 38 states that *any* equilibrium in which the men state their true preferences produces a matching that is stable with respect to the *true* preferences. Note that the conclusion would not hold if we did not restrict our attention to equilibria in which the men play undominated strategies. For example, when every agent states that no other agent is acceptable, the result is an equilibrium at which all agents remain single.

When preferences are strict, the next result presents a sort of converse to

Theorem 38, since it says that any matching μ which is stable under the true preferences can be obtained by an equilibrium set of strategies.

Theorem 39 [Gale and Sotomayor (1985b)]. *When all preferences are strict, let μ be any stable matching for (M, W, P). Suppose each woman w in $\mu(M)$ chooses the strategy of listing only $\mu(w)$ on her stated preference list of acceptable men (and each man states his true preferences). This is an equilibrium in the game induced by the M-optimal matching mechanism (and μ is the matching that results).*

The next theorem describes an equilibrium even for the case when preferences need not be strict. Furthermore, this equilibrium is a "strong equilibrium for the women", in that no coalition of women can achieve a better outcome for all of its members by having its members change their strategies.

Theorem 40 [Gale and Sotomayor (1985b)]. *Let P' be a set of preferences in which each man states his true preferences, and each woman states a preference list which ranks the men in the same order as her true preferences, but ranks as unacceptable all men who ranked below $\mu_W(w)$. These preferences P' are a strong equilibrium for the women in the game induced by an M-optimal stable matching mechanism (and μ_W is the matching that results).*

Note that these last two theorems describes strategies which put a great burden on the amount of information the women must have in order to implement them. In Subsection 5.3 we will relax the assumption that agents know one another's preferences. In the meantime, it should be clear that advising a woman to play the strategy of Theorem 40, for example, will be singularly unhelpful in most of the practical situations to which we might want to apply a theory of matching, since the strategy requires each woman w to know $\mu_W(w)$. This leads us to consider what advice we can give in environments in which information about other players' preferences may not be readily available to the players.

5.1.2. Good and bad strategies

The problems of coordination and information that may arise in implementing equilibria do not arise in the same way for players who have a dominant strategy. In particular, Theorem 32 implies that when an M-optimal stable matching procedure is used, a man may confidently state his true preferences, without regard to what the preferences of the other men and women may be. So this is a good strategy for the men, and other strategies are, in comparison, bad. Although we have seen that stating the true preferences is not a good strategy in the same way for the women, we turn now to considering what classes of strategies might be bad, in the sense of being dominated by other available strategies.

The first result states that, although it may not be wise for a woman to state her true preferences when the M-optimal stable matching mechanism is used, it can never help her to state preferences in which her first choice mate according to her stated preferences is different from her true first choice.

Theorem 41 [Roth (1982a)]. *Any strategy $P'(w)$ in which w does not list her true first choice at the head of her list is strictly dominated, in the game induced by the M-optimal stable mechanism.*

Theorem 42 states that Theorem 41 describes essentially all the dominated strategies.

Theorem 42 [Gale and Sotomayor (1985b)]. *Let $P'(w)$ be any strategy for w in which w's true first choice is listed first, and the acceptable men in $P'(w)$ are also acceptable men in w's true preference list $P(w)$. Then $P'(w)$ is not a dominated strategy when the M-optimal stable mechanism is used.*

5.2. Many-to-one matching: The college admissions model

We return now to the case of many-to-one matching, and the kind of strategic question that caused the initial 1950 algorithm in the hospital-intern labor market to be abandoned in favor of the NIMP algorithm: Is it always in agents' best interest to state their true preferences? From the Impossibility Theorem for the special case of the marriage market (Theorem 30) we know that no stable matching mechanism can have this property for all agents. But in the marriage market we observed that a mechanism that produced the optimal stable matching for one side of the market made it a dominant strategy for agents on that side to state their true preferences (Theorem 32). We might therefore hope that the parallel result holds for the college admissions model. However, this is not the case: as the next theorem shows, Theorem 32 is one of those results that does not generalize from the case of one-to-one matching to the case of many-to-one matching.

Theorem 43 [Roth (1985a)]. *No stable matching mechanism exists which makes it a dominant strategy for all colleges to state their true preferences.*

An immediate corollary of the proof of Theorem 43 is that Theorem 37 is another of the results which does not generalize from the special case of the marriage model. That is, we have

Corollary 44 [Roth and Sotomayor (1990a)]. *In the college admissions model, the conclusions of Theorem 37 for the marriage model do not hold. A coalition of agents (in fact even a single agent) may be able to misrepresent its preferences so that it does better than at any stable matching.*

Although Theorem 43 shows that no stable matching mechanism gives colleges a dominant strategy, the situation of students is as in the marriage problem. That is, we have the following result.

Theorem 45 [Roth (1985a)]. *A stable matching mechanism for the college admissions model which yields the student-optimal stable matching makes it a dominant strategy for all students to state their true preferences.*

As in the case of the marriage model, these results do little to help us identify "good" strategies for either the students or the colleges when the college-optimal stable mechanism is used. No agents have dominant strategies under that mechanism, so they all face potentially complex decision problems. And we cannot even say as much about equilibria as we could for the marriage market, since there are lots of Nash equilibria, and no easy way to distinguish among them, since the lack of dominant strategies prevents us from eliminating unreasonable equilibria as in Theorem 38. However, since Theorem 45 establishes that the student-optimal stable mechanism makes it a dominant strategy for students to state their true preferences, we might hope to have at least a one-sided generalization of Theorem 38, which would say that every equilibrium of the student-optimal stable mechanism at which students state their true preferences is stable with respect to the true preferences. But this is another result which fails to generalize, even in this partial way, from the special case of the marriage model. Again, the result is a corollary of the proof of Theorem 43.

Corollary 46 [Roth and Sotomayor (1990a)]. *In the college admissions model, the conclusions of Theorem 38 for the marriage model do not hold, even for the student-optimal stable mechanism. When all students state their true preferences, there may be equilibria of the student-optimal stable mechanism which are not stable with respect to the true preferences.*

In general, although there are equilibrium misrepresentations that yield stable matchings with respect to the true preferences, there are also equilibrium misrepresentations that yield any individually rational matching, stable or not.

Theorem 47 [Roth (1985a)]. *There exist Nash equilibrium misrepresentations under any stable matching mechanism that produce any individually rational matching with respect to the true preferences.*

But the equilibria referred to in this theorem may require a great deal of both information and coordination, since, for example, an individually rational matching μ may be achieved at equilibrium if each agent x states that $\mu(x)$ is his or her only acceptable mate.

5.3. Incomplete information

As we have seen, the (implicit) assumption of complete information makes its presence felt in a burdensome way in some of the equilibrium strategies which arise (cf. Theorems 39, 40, and 47). In this subsection we consider which of the results we have discussed so far are robust to a relaxation of the complete information assumption, and which are not.

A one-to-one marriage game with incomplete information about others' preferences will be given by a collection

$$\Gamma = (N = M \cup W, \{D_i\}_{i \in N}, g, U = X_{i \in N} U_i, F) .$$

The set N of players consists of the men and women to be matched. The sets D_i describe the decisions facing each player in the course of any play of the game (i.e. an element d_i of D_i specifies the action of player i at each point in the game at which he has decisions to make). The function g describes how the actions taken by all the agents correspond to matchings and lotteries over matchings, i.e. $g: X_{i \in N} D_i \to L[\mathcal{M}]$, where $\mathcal{M}$ is the set of all matchings between the sets M and W, and $L[\mathcal{M}]$ is the set of all probability distributions (lotteries) over $\mathcal{M}$. The set U_i is the set of all expected utility functions defined over the possibile mates for player i and the possibility of remaining single, and F is a probability distribution over n-tuples of utility functions $u = \{u_i\}_{i \in N}$, for u_i in U_i. The interpretation is that a player's "type" is given by his utility function, and at the time players must choose their strategies each player knows his own type, and the probability distribution F over vectors u is common knowledge. The special case of a game of complete information occurs when the distribution F gives a probability of one to some vector u of utilities. We will typically be concerned with games in which only a countable subset of U has positive probability. In any event, since each player i knows his own utility function u_i, he can compute a conditional probability $p_i(u_{-i} \mid u_i)$ for each vector of other players' utilities u_{-i} in $U_{-i} \equiv X_{j \neq i} U_j$, by applying Bayes' rule to F.

This is not the most general kind of incomplete information model we might consider. The only unknown information is the other players' utilities. In particular, players know their own utilities for being matched with one another even though they do not know what "type" the other is. Each player's utility payoff depends on his own type, and on the actions of all the players (through the matching that results), but not on the types of the other players, i.e. players' types do not effect their desirability, only their desires. This seems like a natural assumption for elite professional markets for entry level positions. For example, in the hospital-intern market, after the usual interviewing has been completed, top students are able to rank prestigious programs, and vice versa. But agents do not know how their top choices rank *them*. (Note the difference between this kind of model and one in which the interviewing process itself is modelled, in which agents would in effect be uncertain about their own preferences.)

A *strategy* for player i is a function σ_i from his type (which in this case is his utility function) to his decisions, i.e. $\sigma_i: U_i \to D_i$. If $\sigma = \{\sigma_i\}_{i\in N}$ denotes the strategy chosen by each player, then for each vector u of players' utility functions, $\sigma(u) = \{d_i \in D_i\}_{i\in N}$ describes the decisions made by the players, which result in the matching (or lottery over matchings) $g(\sigma(u))$. Consequently, a set of strategy choices σ results in a lottery over matchings, the probabilities of which are determined by the probability distribution F over vectors u, and by the function g. The expected utility to player i who is of type u_i is given by

$$u_i(\sigma) = \sum_{u_{-i}\in U_{-i}} p_i(u_{-i} \mid u_i)u_i[g(\sigma(u_{-i}, u_i))] .$$

A *Bayesian equilibrium*[28] is a σ^* such that, for all players i in N and all utility functions u_i in U_i, $u_i(\sigma^*) \geq u_i(\sigma^*_{-i}, \sigma_i)$ for all other strategies σ_i for player i. That is, when player i's utility is u_i the strategy σ^*_i determines player i's decision $d^*_i = \sigma^*_i(u_i)$, and the equilibrium condition requires that for all players i and all types u_i which occur with positive probability, player i cannot profitably substitute another decision $d_i = \sigma_i(u_i)$.

Recall that a general matching game with incomplete information about others' preferences is given by $\Gamma = (N = M \cup W,\ \{D_i\}_{i\in N},\ g,\ U = X_{i\in N}U_i, F)$. We may call $[\{D_i\}_{i\in N}, g]$ the *mechanism*, and $[U, F]$ the *state of information* of the game. Then a game Γ is specified by a set of players, a mechanism, and a state of information. Note that we are here considering much more general kinds of mechanisms than the simple "revelation mechanisms" of the kind observed in the NIMP algorithm, for example, in which

[28]See Chapter 5 in this Handbook on incomplete information.

agents are just asked to state their preferences. Since we will be stating an impossibility theorem, we want to consider very general mechanisms.

The first result is an impossibility theorem that provides a strong negation to the conclusions of Theorem 38 about equilibria in the complete information case when the M-optimal stable mechanism is employed. It says that, in the incomplete information case, no equilibrium of any mechanism can have the stability properties that every equilibrium[29] of the M-optimal stable mechanism has in the complete information case. (The strategy of the proof is to observe that, by the revelation principle, if any such mechanism existed then there would be a stable revelation mechanism with truthtelling as an equilibrium, and then to show that no such revelation mechanism exists.)

Theorem 48 [Roth (1989)]. *If there are at least two agents on each side of the market, then for* any *general matching mechanism* $[\{D_i\}_{i\in N}, g]$ *there exist states of information* $[U, F]$ *for which* every *equilibrium* σ *of the resulting game* Γ *has the property that* $g(\sigma(u)) \not\subseteq L[S(u)]$ *for some* $u \in U$. *(And the set of such* u *with* $g(\sigma(u)) \not\subseteq L[S(u)]$ *has positive probability under* F.*) That is, there exists no mechanism with the property that at least one of its equilibria is always stable with respect to the true preferences at every realization of a game.*

The next theorem states that the conclusion of Theorem 36 also does not generalize to the case of incomplete information. It is possible for coalitions of men, by mis-stating their preferences, to obtain a preferable matching (even) from the M-optimal stable mechanism. This is so even though, as we will briefly discuss, it remains a dominant strategy for each man to state his true preferences.

Theorem 49 [Roth (1989)]. *In games of incomplete information about preferences, the M-optimal stable mechanism may be group manipulable by the men.*

As discussed earlier, the fact that, even in the case of complete information it is possible for a coalition of men to mis-state their preferences in a way that does not hurt any of them and helps some of them, means that the conclusion from Theorem 36 that coalitions of men cannot collectively manipulate the M-optimal mechanism to their advantage cannot be expected to be very robust. Once there is any possibility that the men can make any sort of sidepayments among themselves, this conclusion is no longer justified. The proof of Theorem 49 depends on observing that uncertainty about the preferences of other agents allows some transfers in an expected utility sense, with

[29] In undominated strategies.

men able to trade a gain in one realization for a gain in another. Building on Example 14, it is not hard to show that this can occur even when there is arbitrarily little uncertainty about the preferences.

In contrast to the results for equilibria, the results concerning dominant strategies in the complete information case do generalize to the case of incomplete information. This can be seen by a pointwise argument on realizations of the types of the players. In this way Roth (1989) observed that the conclusions of Theorem 32 and Theorem 41 generalize to the present case: when an M-optimal stable mechanism is used, it is a dominant strategy for each man to state his true preferences, and any strategy for a woman is dominated if her stated first choice is not her true first choice for each of her possible types.

6. Empirical overview

We return now to see what the theory described here can tell us about the principal example which we used to motivate our consideration of stability in two-sided matching markets, namely the hospital-intern labor market. We begin with the formal statement of the result promised in the preview given in Subsection 2.1.1.

Theorem 50 [Roth (1984a)]. *The NIMP algorithm is a stable matching mechanism, i.e. it produces a stable matching with respect to any stated preferences. (In fact, it produces the hospital-optimal stable matching.)*

This result lends support to the conjecture offered in the first part of Subsection 2.1.1 that the difference between the chaotic markets of the late 1940s and the orderly operation of the market with such high rates of voluntary participation starting in the early 1950s can be attributed to the stability of the matchings produced by the centralized procedure.[30]

However in Subsection 2.1 we also referred to the fact that, at least as early as 1973, significant numbers of married couples declined to take part in the NIMP procedure, or to accept the jobs assigned to them by that procedure. If it is the stability of the matching which contributes to voluntary participation in a centralized matching procedure, this should make us suspect that something about the presence of couples introduced instabilities into the market. In fact, the NIMP program included a specific procedure for handling couples that will

[30]The theorem also explains the way in which the NIMP algorithm is equivalent to the deferred acceptance procedure with hospitals proposing, since it also produces the hospital-optimal stable matching. However the internal working of the two algorithms differ in ways that are important for their implementation – see Roth (1984a) and Roth and Sotomayor (1990a).

make it fairly clear how these instabilities arose (and why they were so prevalent), at least until 1983, when the procedure for married couples was modified.

Briefly, the situation prior to 1983 was this. Couples graduating from medical school at the same time, and wishing to obtain two positions in the same community, had two options. One option was to stay outside of the NIMP program and negotiate directly with hospital programs. Alternatively, they could (after being certified by the Dean of their medical school as a legitimate couple) enter the NIMP program together to be matched by a special "couples algorithm".

This couples algorithm can be described roughly as follows. The couple was required to specify one of its members as the "leading member", and to submit a rank ordering of positions for each member of the couple, i.e. a couple submitted two preference lists, one for each member. The leading member of the couple was then matched to a position in the usual way, the preference list of the other member of the couple was edited to remove distant positions, and the second member was then matched if possible to a position in the same vicinity as the leading member.

It is easy to see why instabilities often result. Consider a couple $\{s_1, s_2\}$ whose first choice is to have two particular jobs in Boston, and whose second choice is to have two particular jobs in New York. Under the couples algorithm, the designated "leading member" might be matched to his or her first choice job in Boston, while the other member might be matched to some relatively undesirable job in Boston. If s_1 and s_2 were ranked by their preferred New York jobs higher than students matched to those jobs, an instability would now exist, since the couple would prefer to take the two New York jobs, and the New York hospitals would prefer to have s_1 and s_2.

Notice that, to describe this kind of instability, we are implicitly proposing a modification of the basic model of agents in the market. A couple consists of a pair of students who have a single preference ordering over pairs of positions. Part of the problem with the couples algorithm just described is that it did not permit couples to state their preferences over pairs of positions. Starting with the 1983 match, modifications were made so that couples could for the first time express such preferences within the framework of the centralized matching scheme. However, the following theorem shows that the problem goes deeper than that.

Theorem 51 [Roth (1984a)].[31] *In the hospital-intern problem with couples, the set of stable matchings may be empty.*

[31] This result was independently proved by Sotomayor in an unpublished note.

In view of the evidence in favor of the proposition that high voluntary rates of participation are associated with the stability of the matching mechanism, this suggests that the problem with married couples may be a persistent one. In a similar way, the next theorem suggests that the distribution of interns to rural hospitals discussed in Subsection 2.1 also is not likely to respond to any changes in procedures which achieve high degrees of voluntary compliance.

Theorem 52 [Roth (1986)]. *When all preferences over individuals are strict, the set of interns employed and positions filled is the same at every stable matching. Furthermore, any hospital that does not fill its full quota at some stable matching is matched with exactly the same set of interns at every stable matching.*

6.1. Some further remarks on empirical matters

There are several reasons why we have devoted some attention, in a survey largely concerned with mathematical theory, to the way that American physicians get their first jobs. One reason is to suggest why we think that the body of theory developed here has empirical content. Another reason is simply to give readers an idea of what empirical work connected with theory of this kind might look like. And a third reason is because it seems likely that the lessons learned from the rather special market for American medical interns may generalize to a much wider variety of entry level labor markets and other matching processes.

Regarding the empirical content of the theory, we have laid great weight in our explanation of the history of the medical market on the fact that the centralized market mechanism introduced in 1951 is a stable matching mechanism, and on the fact that the growing numbers of married couples in the market introduce instabilities. It might be objected that these are coincidental features of the market, and that the true explanations of, for example, the rates of participation lie elsewhere. For example, it might be postulated that *any* centralized market organization would have solved the problems experienced prior to 1951, and that the difficulties with having married couples in the market have less to do with instabilities of the kind dealt with here than with the difficulties that young couples have in making decisions.

Ideally, we would like to be able to conduct carefully controlled experiments designed to distinguish between any such alternative hypotheses.[32] But for theories involving the histories of complex natural organizations, we often have to settle for finding "natural experiments" which let us distinguish as well as we

[32]And laboratory experimentation is indeed becoming more common: see the chapter by Shubik on that subject in a forthcoming volume of this Handbook, or see *Handbook of experimental economics* [Kagel and Roth (1992)].

can between competing hypotheses. A very nice natural experiment involving these matters can be found when we look across the Atlantic ocean and examine how new physicians in the United Kingdom obtain their first jobs. The following very brief description is taken from Roth (1991).

Around the middle of the 1960s, the entry level market for physicians in England, Scotland, and Wales began to suffer from some of the same acute problems that had arisen in the American market in the 1940s and 1950s. Chief among these was that the date of appointment for "preregistration" positions (comparable to American internships, and required of new medical school graduates) had crept back in many cases to years before the date a student would graduate from medical school. The market for these positions is regional rather than national, and this problem occurred more or less in the same way in many of the regional markets. (These regional markets have roughly 200 positions each, so they are two full orders of magnitude smaller than the American market.)

The British medical authorities were aware of the experience of the American market, and in many of the regional markets it was decided to introduce a centralized market mechanism using a computerized algorithm to process preference lists obtained from students and hospitals, modelled loosely after the American system, but adapted to local conditions. Most of these algorithms were not stable matching mechanisms, and it appears that a substantial majority of those that were not failed to solve the problems they were designed to address, and were eventually abandoned. [Before being abandoned at least some experienced serious incentive problems, the evidence being a lack of volunatary participation, or a variety of unstraightforward strategic behavior. Some of the ways in which mechanisms failed, and the kind of strategic behavior they elicited, are extremely instructive; see Roth (1991) or Roth and Sotomayor (1990a) for details.] As far as can so far be determined, only two stable matching mechanisms were introduced. Both were largely successful and remain in use to this day. The similarity of the British experience in markets with unstable mechanisms to the American situation prior to 1951, and the similarity of the British experience in the markets with stable mechanisms to the American experience after 1951, support the argument that stability plays at least something like the role we have attributed to it.

The nature of this kind of empirical investigation is of course very different from the purely mathematical investigation of abstract cases. Particular models adapted to the institutional details of the markets in question must be considered (just as considering instabilities involving married couples required us to extend the basic hospital intern model). To give a bit of the flavor of this, one example comes to mind.

One of the stable matching procedures was introduced in a region of Scotland where, in keeping with previous custom, certain kinds of hospital programs were permitted to specify that they did not wish to employ more than one female physician at any time. A program taking advantage of this option might submit a preference list on which several women graduates were highly ranked, but nevertheless stipulate that no more than one of these should be assigned to it. In analyzing such a model, it is of course necessary to consider whether the introduction of such "discriminatory quotas" influences the existence of stable matchings. We leave as an exercise for the reader to show that the model of many-to-one matching with substitutable preferences can be used to address this question, and to prove the following proposition.

Proposition 53 [Roth (1991)]. *In the hospital-intern model with discriminatory quotas, the set of stable matchings is always nonempty.*

Regarding directions for future empirical work, we remark that the two studies discussed here [Roth (1984a, 1991)] are both part of a line of work that seeks to identify markets in which it is possible to establish a particularly close connection between the observed market outcome and the set of stable outcomes. This connection can be made so closely because the markets in question used computerized matching procedures which can be examined to determine the precise relationship between the submitted preferences and the market outcome. But the kind of theory developed here is by no means limited to such markets, and as more becomes known about the behavior of other entry level labor markets, for example, we should be better able to associate certain phenomena with markets that achieve stable outcomes, and other phenomena with markets that achieve unstable outcomes. In this way it should be possible to extend the empirical investigation of the predictions of this kind of theory to two-sided matching markets which are operated in a completely decentralized manner.

An interesting intermediate case, which has been described in Mongell (1987) and Mongell and Roth (1991), concerns the procedures by which the social organizations known as sororities, which operate on many American college campuses, are matched each year with new members. A centralized procedure is employed which in general would not lead to a stable matching, but because the agents in that market respond to the incentives which the procedure gives them not to state their full true preferences, much of the actual matching in that market is done in a decentralized after-market. In the data examined by Mongell and Roth, the strategic behavior of the agents led to stable matches. (This study reaffirms the importance of examining systems of rules from the point of view of how they will behave when participants respond strategically to the incentives which the rules create.)

Finally, what more general conclusions can be draw from the empirical observations we have so far been able to make of two-sided matching markets?

While some of these have been widely interpreted as evidence that "game theory works", our own view is that a somewhat more cautious interpretation is called for. First, while there is a wide variety of game-theoretic work concerning a diversity of environments, there has so far been very much less empirical work that provides tests of game-theoretic predictions. This is no doubt due to the difficulty of gathering the kind of detailed information about institutions and agents that game-theoretic theories employ, and for this reason much of the most interesting empirical work has involved controlled experiments under laboratory conditions.[33] What has made the empirical work on two-sided matching markets different is that it has proved possible to identify naturally occurring markets for which the necessary information can be found. Which brings us to the question: How does the theory fare when tested on the markets observed to date?

Even here, the answer is a little complex. We certainly cannot claim that the evidence supports the simple hypothesis that the outcome of two-sided matching markets will always be stable, since we have observed markets that employ unstable procedures and produce unstable matchings at least some of the time. And even those markets that eventually developed procedures to produce stable matchings operated for many years without such procedures before the problems they encountered in doing so led them to develop the rules they successfully use today.

However the evidence is much clearer when we turn from simple predictions to conditional predictions. The available evidence strongly supports the hypothesis that if matching markets are organized in ways that produce unstable matchings, then they are prone to a variety of related problems and market failures that can largely be avoided if the markets are organized in ways that produce stable matchings. So the kinds of empirical work described here go a long way towards supporting the contention that (at least parts of) game theory may reasonably be thought of as a source of useful theories about complex natural phenomena, and not merely of idealized or metaphorical descriptions of the behavior of perfectly rational agents.

Bibliography

Alcalde, José and Salvador Barbera (1991) 'Top dominance and the possibility of strategy proof stable solutions to the marriage problem', Universitat Antònoma de Barcelona, mimeo.

Alkan, Ahmet (1988a) 'Existence and computation of matching equilibria', Bogazici University, mimeo.

Alkan, Ahmet (1988b) 'Auctioning several objects simultaneously', Bogazici University, mimeo.

[33]For discussions of experiments, see a forthcoming volume of this Handbook, the surveys in Roth (1987a, 1987b), or *Handbook of experimental economics* (Kagel and Roth (1992)].

Alkan, Ahmet (1988c) 'Nonexistence of stable threesome matchings', *Mathematical Social Sciences*, 16: 207–209.
Alkan, Ahmet and David Gale (1990) 'A constructive proof of non-emptiness of the core of the matching game', *Games and Economic Behavior*, 2: 203–212.
Allision, Lloyd (1983) 'Stable marriages by coroutines', *Information Processing Letters*, 16: 61–65.
Balinski, M.L. and David Gale (1990) 'On the core of the assignment game', in: L.J. Leifman, ed., *Functional analysis, optimization and mathematical economics*. Oxford: Oxford University Press, pp. 274–289.
Bartholdi, John J. III and Michael A. Trick (1986) 'Stable matching with preferences derived from a psychological model', *Operations Research Letters*, 5: 165–169.
Becker, Gary S. (1981) *A treatise on the family*. Cambridge, Mass.: Harvard University Press.
Bennett, Elaine (1988) 'Consistent bargaining conjectures in marriage and matching', *Journal of Economic Theory*, 45: 392–407.
Bergstrom, Theodore and Richard Manning (1982) 'Can courtship be cheatproof?' (personal communication).
Bird, Charles G. (1984) 'Group incentive compatibility in a market with indivisible goods', *Economic Letters*, 14: 309–313.
Blair, Charles (1984) 'Every finite distributive lattice is a set of stable matchings', *Journal of Combinatorial Theory* (Series A), 37: 353–356.
Blair, Charles (1988) 'The lattice structure of the set of stable matchings with multiple partners', *Mathematics of Operations Research*, 13: 619–628.
Brams, Steven J. and Philip D. Straffin, Jr. (1979) 'Prisoners' dilemma and professional sports drafts', *American Mathematical Monthly*, 86: 80–88.
Brissenden, T.H.F. (1974) 'Some derivations from the marriage bureau problem', *The Mathematical Gazette*, 58: 250–257.
Cassady, Ralph Jr. (1967) *Auctions and auctioneering*. Berkeley: University of California Press.
Checker, Armand (1973) 'The national intern and resident matching program, 1966–72', *Journal of Medical Education*, 48: 106–109.
Crawford, Vincent P. (1991), 'Comparative statics in matching markets', *Journal of Economic Theory*, 54: 389–400.
Crawford, Vincent P. and Elsie Marie Knoer (1981) 'Job matching with heterogeneous firms and workers', *Econometrica*, 49: 437–450.
Crawford, Vincent P. and Sharon C. Rochford (1986) 'Bargaining and competition in matching markets', *International Economic Review*, 27: 329–348.
Curiel, Imma J. (1988) *Cooperative game theory and applications*, Doctoral dissertation. Katholieke Universiteit van Nijmegen.
Curiel, Imma J. and Stef H. Tijs (1985) 'Assignment games and permutation games', *Methods of Operations Research*, 54: 323–334.
Dantzig, George B. (1963) *Linear programming and extensions*. Princeton: Princeton University Press.
Demange, Gabrielle (1982) 'Strategyproofness in the assignment market game', mimeo, Laboratoire d'Econometrie de l'Ecole Polytechnique, Paris.
Demange, Gabrielle (1987) 'Nonmanipulable cores', *Econometrica*, 55: 1057–1074.
Demange, Gabrielle and David Gale (1985) 'The strategy structure of two-sided matching markets', *Econometrica*, 53: 873–888.
Demange, Gabrielle, David Gale and Marilda Sotomayor (1986) 'Multi-item auctions', *Journal of Political Economy*, 94: 863–872.
Demange, Gabrielle, David Gale and Marilda Sotomayor (1987) 'A further note on the stable matching problem', *Discrete Applied Mathematics*, 16: 217–222.
Diamond, Peter and Eric Maskin (1979) 'An equilibrium analysis of search and breach of contract, I: Steady states', *Bell Journal of Economics*, 10: 282–316.
Diamond, Peter and Eric Maskin (1982) 'An equilibrium analysis of search and breach of contract, II: A non-steady state example', *Journal of Economic Theory*, 25: 165–195.
Dubins, L.E. and D.A. Freedman (1981) 'Machiavelli and the Gale–Shapley algorithm', *American Mathematical Monthly*, 88: 485–494.
Francis, N.D. and D.I. Fleming (1985) 'Optimum allocation of places to students in a national university system', *BIT*, 25: 307–317.

Gale, David (1968) 'Optimal assignments in an ordered set: An application of matroid theory', *Journal of Combinatorial Theory*, 4: 176–180.

Gale, David (1984) 'Equilibrium in a discrete exchange economy with money', *International Journal of Game Theory*, 13: 61–64.

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

Gale, David and Marilda Sotomayor (1985a) 'Some remarks on the stable matching problem', *Discrete Applied Mathematics*, 11: 223–232.

Gale, David and Marilda Sotomayor (1985b) 'Ms Machiavelli and the stable matching problem', *American Mathematical Monthly*, 92: 261–268.

Gardenfors, Peter (1973) 'Assignment problem based on ordinal preferences', *Management Science*, 20: 331–340.

Gardenfors, Peter (1975) 'Match making: Assignments based on bilateral preferences', *Behavioral Science*, 20: 166–173.

Graham, Daniel A. and Robert C. Marshall (1984) 'Bidder coalitions at auctions', Duke University Department of Economics, mimeo.

Graham, Daniel A. and Robert C. Marshall (1987) 'Collusive bidder behavior at single object second price and English auctions', *Journal of Political Economy*, 95: 1217–1239.

Graham, Daniel A., Robert C. Marshall and Jean-Francois Richard (1987) 'Auctioneer's behavior at a single object English auction with heterogeneous non-cooperative bidders', Working paper #87-01, Duke University Institute of Statistics and Decision Sciences.

Graham, Daniel A., Robert C. Marshall and Jean-Francois Richard (1990) 'Differential payments within a bidder coalition and the Shapley value', *American Economic Review*, 80: 493–510.

Granot, Daniel (1984) 'A note on the room-mates problem and a related revenue allocation problem', *Management Science*, 30: 633–643.

Gusfield, Dan (1987) 'Three fast algorithms for four problems in stable marriage', *SIAM Journal on Computing*, 16: 111–128.

Gusfield, Dan (1988) 'The structure of the stable roommate problem: Efficient representation and enumeration of all stable assignments', *SIAM Journal on Computing*, 17: 742–769.

Gusfield, Dan and Robert W. Irving (1989) *The stable marriage problem: Structure and algorithms*. Cambridge, Mass.: MIT Press.

Gusfield, Dan, Robert W. Irving, Paul Leather and M. Saks (1987) 'Every finite distributive lattice is a set of stable matchings for a *small* stable marriage instance', *Journal of Combinatorial Theory* A, 44: 304–309.

Harrison, Glenn W. and Kevin A. McCabe (1988) Stability and preference distortion in resource matching: An experimental study of the marriage market', mimeo, Department of Economics, University of New Mexico.

Hull, M. Elizabeth C. (1984) 'A parallel view of stable marriages', *Information Processing Letters*, 18: 63–66.

Hwang, J.S. (1978) 'Complete unisexual stable marriages', *Soochow Journal of Mathematics*, 4: 149–151.

Hwang, J.S. (1986) 'The algebra of stable marriages', *International Journal of Computer Mathematics*, 20: 227–243.

Hwang, J.S. (undated) Modelling on college admissions in terms of stable marriages', mimeo.

Hwang, J.S. and H.J. Shyr (1977) 'Complete stable marriages', *Soochow Journal of Mathematical and Natural Sciences*, 3: 41–51.

Hylland, Aanund and Richard Zeckhauser (1979) 'The efficient allocation of individuals to positions', *Journal of Political Economy*, 87: 293–314.

Irving, Robert W. (1985) 'An efficient algorithm for the stable room-mates problem', *Journal of Algorithms*, 6: 577–595.

Irving, Robert W. (1986) 'On the stable room-mates problem', mimeo, Department of Computing Science, University of Glasgow.

Irving, Robert W. and Paul Leather (1986) 'The complexity of counting stable marriages', *SIAM Journal of Computing*, 15: 655–667.

Irving, Robert W., Paul Leather and Dan Gusfield (1987) 'An efficient algorithm for the "optimal" stable marriage', *Journal of the ACM* 34: 532–543.

Itoga, Stephen Y. (1978) 'The upper bound for the stable marriage problem', *Journal of the Operational Research Society*, 29: 811–814.

Itoga, Stephen Y. (1981) 'A generalization of the stable marriage problem', *Journal of the Operational Research Society*, 32: 1069–1074.

Itoga, Stephen Y. (1983) 'A probabilistic version of the stable marriage problem', *BIT*, 23: 161–169.

Jones, Philip C. (1983) 'A polynomial time market mechanism', *Journal of Information and Optimization Sciences*, 4: 193–203.

Kagel, John and Alvin E. Roth, eds. (1992) *Handbook of experimental economics*. Princeton, NJ: Princeton University Press.

Kamecke, Ulrich (1987) 'A generalization of the Gale–Shapley algorithm for monogamous stable matchings to the case of continuous transfers', Discussion paper, Rheinische Friedrich-Wilhelms Universitat, Bonn.

Kamecke, Ulrich (1989) 'Non-cooperative matching games', *International Journal of Game Theory*, 18: 423–431.

Kamecke, Ulrich (1992) 'On the uniqueness of the solution to a large linear assignment problem', *Journal of Mathematical Economics*, forthcoming.

Kaneko, Mamoru (1976) 'On the core and competitive equilibria of a market with indivisible goods', *Naval Research Logistics Quarterly*, 23: 321–337.

Kaneko, Mamoru (1982) 'The central assignment game and the assignment markets', *Journal of Mathematical Economics*, 10: 205–232.

Kaneko, Mamoru (1983) 'Housing markets with indivisibilities', *Journal of Urban Economics*, 13: 22–50.

Kaneko, Mamoru and Myrna Holtz Wooders (1982) 'Cores of partitioning games', *Mathematical Social Sciences*, 3: 313–327.

Kaneko, Mamoru and Myrna Holtz Wooders (1985) 'The core of a game with a continuum of players and finite coalitions: Nonemptiness with bounded sizes of coalitions', mimeo, Institute for Mathematics and its Applications, University of Minnesota.

Kaneko, Mamoru and Myrna Holtz Wooders (1986) 'The core of a game with a continuum of players and finite coalitions: The model and some results', *Mathematical Social Sciences*, 12: 105–137.

Kaneko, Mamoru and Yoshitsugu Yamamoto (1986) 'The existence and computation of competitive equilibria in markets with an indivisible commodity', *Journal of Economic Theory*, 38: 118–136.

Kapur, Deepak and Mukkai S. Krishnamoorthy (1985) 'Worst-case choice for the stable marriage problem', *Information Processing Letters*, 21: 27–30.

Kelso, Alexander S., Jr. and Vincent P. Crawford (1982) 'Job matching, coalition formation, and gross substitutes', *Econometrica*, 50: 1483–1504.

Knuth, Donald E. (1976) *Marriages stables*. Montreal: Les Presses de l'Universite de Montreal.

Leonard, Herman B. (1983) 'Elicitation of honest preferences for the assignment of individuals to positions', *Journal of Political Economy*, 91: 461–479.

Masarani, F. and S.S. Gokturk (1988) 'On the probabilities of the mutual agreement match', *Journal of Economic Theory*, 44: 192–201.

McVitie, D.G. and L.B. Wilson (1970a) 'Stable marriage assignments for unequal sets', *BIT*, 10: 295–309.

McVitie, D.G. and L.B. Wilson (1970b) 'The application of the stable marriage assignment to university admissions', *Operational Research Quarterly*, 21: 425–433.

McVitie, D.G. and L.B. Wilson (1971) 'The stable marriage problem', *Communications of the ACM*, 14: 486–492.

Mo, Jie-ping (1988) 'Entry and structures of interest groups in assignment games', *Journal of Economic Theory*, 46: 66–96.

Moldovanu, Benny (1990) 'Bargained equilibria for assignment games without side payments', *International Journal of Game Theory*, 18: 471–477.

Mongell, Susan J. (1987) *Sorority rush as a two-sided matching mechanism: A game-theoretic analysis*, Ph.D. dissertation. Department of Economics, University of Pittsburgh.

Mongell, Susan J. and Alvin E. Roth (1986) 'A note on job matching with budget constraints', *Economics Letters*, 21: 135–138.

Mongell, Susan J. and Alvin E. Roth (1991) 'Sorority rush as a two-sided matching mechanism', *American Economic Review*, 81: 441–464.
Mortensen, Dale T. (1982) 'The matching process as a Noncooperative bargaining game', in: J. McCall, ed., *The Economics of Information and Uncertainty*. Chicago: University of Chicago Press, pp. 233–258.
Owen, Guillermo (1975) 'On the core of linear production games', *Mathematical Programming*, 9: 358–370.
Prasad, Kislaya (1987) 'The complexity of games II: Assignment games and indices of power', mimeo, Department of Economics, Syracuse University.
Proll, L.G. (1972) 'A simple method of assigning projects to students', *Operational Research Quarterly*, 23: 195–201.
Quinn, Michael J. (1985) 'A note on two parallel algorithms to solve the stable marriage problem', *Bit*, 25: 473–476.
Quint, Thomas (1987a) 'Elongation of the core in an assignment game', Technical report, IMSSS, Stanford.
Quint, Thomas (1987b) 'A proof of the nonemptiness of the core of two sided matching markets', CAM report #87–29, Department of Mathematics, UCLA.
Quint, Thomas (1988) 'An algorithm to find a core point for a two-sided matching model', CAM report #88–03, Department of Mathematics, UCLA.
Quint, Thomas (1991) 'The core of an m-sided assignment game', *Games and Economic Behavior*, 3: 487–503.
Quinzii, Martine (1984) 'Core and competitive equilibria with indivisibilities,' *International Journal of Game Theory*, 13: 41–60.
Rochford, Sharon C. (1984) 'Symmetrically pairwise-bargained allocations in an assignment market', *Journal of Economic Theory* 34: 262–281.
Ronn, Eytan (1986) *On the complexity of stable matchings with and without ties*, Ph.D. dissertation, Yale University.
Ronn, Eytan (1987) 'NP-complete stable matching problems', *Journal of Algorithms*, forthcoming.
Roth, Alvin E. (1982a) 'The economics of matching: Stability and incentives', *Mathematics of Operations Research*, 7: 617–628.
Roth, Alvin E. (1982b) 'Incentive compatibility in a market with indivisible goods', *Economics Letters*, 9: 127–132.
Roth, Alvin 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.
Roth, Alvin E. (1984b) 'Misrepresentation and stability in the marriage problem', *Journal of Economic Theory*, 34: 383–387.
Roth, Alvin E. (1984c) 'Stability and polarization of interests in job matching', *Econometrica*, 52: 47–57.
Roth, Alvin E. (1985a) 'The college admissions problem is not equivalent to the marriage problem', *Journal of Economic Theory*, 36: 277–288.
Roth, Alvin E. (1985b) 'Common and conflicting interests in two-sided matching markets', *European Economic Review* (Special issue on Market Competition, Conflict, and Collusion), 27: 75–96.
Roth, Alvin E. (1985c) 'Conflict and coincidence of interest in job matching: Some new results and open questions', *Mathematics of Operations Research*, 10: 379–389.
Roth, Alvin E. (1986) 'On the allocation of residents to rural hospitals: A general property of two-sided matching markets', *Econometrica* 54: 425–427.
Roth, Alvin E. (1987a) 'Laboratory experimentation in Economics', in: Truman Bewley, ed., *Advances in economic theory*, *Fifth World Congress*. Cambridge University Press, pp 269–299. (Preprinted in *Economics and Philosophy*, Vol. 2, 1986, 245–273.)
Roth, Alvin E., ed. (1987b) *Laboratory experimentation in economics: Six points of view.* Cambridge: Cambridge University Press.
Roth, Alvin E. (1988) 'Laboratory experimentation in economics: A methodological overview', *Economic Journal*, 98: 974–1031.
Roth, Alvin E. (1989) 'Two sided matching with incomplete information about others' preferences', *Games and Economic Behavior*, 1: 191–209.
Roth, Alvin E. (1991) '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: 415–440.

Roth, Alvin E. and Andrew Postlewaite (1977) 'Weak versus strong domination in a market with indivisible goods', *Journal of Mathematical Economics*, 4: 131–137.

Roth, Alvin E. and Marilda Sotomayor (1988a) 'Interior points in the core of two-sided matching problems', *Journal of Economic Theory*, 45: 85–101.

Roth, Alvin E. and Marilda Sotomayor (1989) 'The college admissions problem revisited', *Econometrica*, 57: 559–570.

Roth, Alvin E. and Marilda Sotomayor (1990a) *Two-sided matching: A study in game-theoretic modelling and analysis*, Econometric Society Monograph Series. Cambridge: Cambridge University Press.

Roth, Alvin E. and Marilda Sotomayor (1990b) 'Stable outcomes in discrete and continuous models of two-sided matching: A unified treatment', University of Pittsburgh, mimeo.

Roth, Alvin E. and John H. Vande Vate (1990) 'Random paths to stability in two-sided matching', *Econometrica*, 58: 1475–1480.

Roth, Alvin E., Uriel G. Rothblum and John H. Vande Vate (1992) 'Stable matchings, optimal assignments, and linear programming', *Mathematics of Operations Research*, forthcoming.

Rothblum, Uriel G. (1992) 'Characterization of stable matchings as extreme points of a polytope', *Mathematical Programming*, forthcoming.

Samet, Dov and Eitan Zemel (1984) 'On the core and dual set of linear programming games', *Mathematics of Operations Research*, 9: 309–316.

Sasaki, Hiroo (1988) 'Axiomatization of the core for two-sided matching problems', Economics Discussion paper #86, Faculty of Economics, Nagoya City University, Nagoya, Japan.

Sasaki, Hiroo and Manabu Toda (1986) 'Marriage problem reconsidered: Externalities and stability', mimeo, Department of Economics, University of Rochester.

Satterthwaite, Mark A. (1975) 'Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions', *Journal of Economic Theory*, 10: 187–217.

Scotchmer, Suzanne and Myrna Holtz Wooders (1988) 'Monotonicity in Games that Exhaust Gains to Scale', mimeo, University of California, Berkeley.

Shapley, Lloyd S. (1962) 'Complements and substitutes in the optimal assignment problem', *Naval Research Logistics Quarterly*, 9: 45–48.

Shapley, Lloyd S. and Herbert Scarf (1974) 'On cores and indivisibility', *Journal of Mathematical Economics*, 1: 23–28.

Shapley, Lloyd S. and Martin Shubik (1972) 'The assignment game I: The core', *International Journal of Game Theory*, 1: 111–130.

Sondak, Harris and Max H. Bazerman (1987) 'Matching and negotiation processes in quasi-markets', *Organizational Behavior and Human Decision Processes*, forthcoming.

Sotomayor, Marilda (1986a) 'On incentives in a two-sided matching market', Working paper, Department of Mathematics, Pontificia Universidade Catolica do Rio de Janeiro.

Sotomayor, Marilda (1986b) 'The simple assignment game versus a multiple assignment game', Working paper, Department of Mathematics, Pontificia Universidade Catolica do Rio de Janeiro.

Sotomayor, Marilda (1987) 'Further results on the core of the generalized assignment game', Working paper, Department of Mathematics, Pontificia Universidade Catolica do Rio de Janeiro.

Sotomayor, Marilda (1988) 'The multiple partners game', William Brock and Mukul Majumdar, eds., in: *Equilibrium and dynamics: Essays in honor of David Gale*, in preparation.

Stalnaker, John M. (1953) 'The matching program for intern placement: The second year of operation', *Journal of Medical Education*, 28: 13–19.

Thompson, Gerald L. (1980) 'Computing the core of a market game', in: A.V. Fiacco and K.O. Kortanek, eds., *Extremal methods and systems analysis*, Lecture Notes in Economics and Mathematical Systems #174. Berlin: Springer, pp. 312–334.

Thompson, William (1986) 'Reversal of asymmetries of allocation mechanisms under manipulation', *Economics Letters*, 21: 227–230.

Toda, Manabu (1988) 'The consistency of solutions for marriage problems', Department of Economics, University of Rochester, mimeo.

Tseng, S.S. and R.C.T. Lee (1984) 'A parallel algorithm to solve the stable marriage problem', *Bit*, 24: 308–316.

Vande Vate, John H. (1989) 'Linear programming brings marital bliss', *Operations Research Letters*, 8: 147–153.

Vickrey, W. (1961), 'Counterspeculation, auctions, and competitive sealed tenders', *Journal of Finance*, 16: 8–37.

Wilson, L.B. (1972) 'An analysis of the stable marriage assignment algorithm', *BIT*, 12: 569–575.

Wilson, L.B. (1977) 'Assignment using choice lists', *Operational Research Quarterly*, 28: 569–578.

Wood, Robert O. (1984) 'A note on incentives in the college admissions market', mimeo, Stanford University.