Sotomayor, Marilda

Working Paper

The Pareto-stability concept is a natural solution concept for discrete matching markets with indifferences

Working Paper, No. 2008-12

Provided in Cooperation with:
Department of Economics, Brown University

THE PARETO-STABILITY CONCEPT IS A NATURAL SOLUTION CONCEPT FOR DISCRETE MATCHING MARKETS WITH INDIFFERENCES

By

MARILDA SOTOMAYOR[1]

Universidade de São Paulo

Department of Economics, , Cidade Universitária, Av. Prof. Luciano Gualberto 908,

05508-900, São Paulo, SP, Brazil

Brown University

Department of Economics, 64 Waterman Street, Providence, RI, USA

e-mail: marildas@usp.br

August, 2008.

ABSTRACT

In a decentralized setting the game-theoretical predictions are that only strong blockings are allowed to rupture the structure of a matching. This paper argues that, under indifferences, also weak blockings should be considered when these blockings come from the grand coalition. This solution concept requires stability plus Pareto optimality. A characterization of the set of Pareto-stable matchings for the roommate and the marriage models is provided in terms of individually rational matchings whose blocking pairs, if any, are formed with unmatched agents. These matchings always exist and give an economic intuition on how blocking can be done by non-trading agents, so that the transactions need not be undone as agents reach the set of stable matchings. Some properties of the Pareto-stable matchings shared by the Marriage and Roommate models are obtained.

Keywords: Pareto-optimal, stable matching, Pareto-stable matching, simple matching, Pareto-simple matching.

JEL numbers: C78, D78

[1] This paper was partially supported by CNPq-Brazil. The author thanks the hospitality of Brown University where some previous version of this work was written. Many thanks are due to Roberto Serrano for his helpful comments.

INTRODUCTION, MAIN CONCEPTS AND RESULTS

In a decentralized setting in which players can interact with each other and get together in groups, the game theoretic predictions are that a matching that can be upset by a coalition will not occur. When preferences are strict, the outcome of such coalitional interactions should then be a stable matching, if it exists. However, such predictions should be revised in the cases in which preferences are not necessarily strict. In such cases, it is justifiable that recontracts between pairs of agents already allocated according to a stable matching, leading to a weak Pareto improvement of the original matching, should be allowed. In this context, it makes sense to predict that only Pareto-stable matchings, i.e. stable matchings that are Pareto optimal, will occur.

This paper takes up this approach and proves some characteristic properties of the Pareto-stable matchings. It concentrates on the well-known Roommate and Marriage models, both introduced by Gale and Shapley in their famous paper of 1962. We follow the notations and concepts presented in Roth and Sotomayor (1990). The Roommate model is described as the pair (N,P), where $N=\{1,2,\ldots,n\}$ is the set of players and P is the set $\{P(1),\ldots,P(n)\}$, where $P(j)$ is an ordered list of preferences (strict or non-strict) for player j. The Marriage model is regarded as a sub-model of the Roommate model in which $N=M\cup W$, M is a set of men and W is a set of women. For the sake of exposition the main concepts will be introduced along this section, as well as the main results of this paper, which will be presented, motivated, discussed and illustrated with examples. We will not always provide a formal statement. The intuitive proofs will be provided here and the technical proofs will be presented in the next section.

To figure out the kind of coalitional interaction taking place among agents allocated according to a stable matching that is not Pareto-optimal, see Example 1 below.

Example 1. (*Pareto-stability is a natural solution concept for the roommate model)* Consider a decentralized setting where a set of eight boys, 1,2,…,8, wish divide up into pairs of roommates. The boys' preferences over acceptable partners are represented by the following ordered lists, where P(j) denotes boy *j*'s list for all *j*=1,…,8:

P(1)=8, 2, 1 P(5)=8, 6, 5

P(2)=[3, 1], 2 P(6)=[3, 5], 6

P(3)=2,6, 4, 3 P(7)=4, 8, 7

P(4)=[3, 7], 4 P(8)= [1, 5, 7], 8

The brackets in the preference lists of boys 2, 4, 6 and 8 mean that these agents are indifferent among the boys inside the brackets. The matching *z,* where *z(1)=2, z(3)=4, z(5)=6, z(7)=8,* doesn't have any blocking pair, so it is stable. This means that no two boys can be both better off by becoming roommates.

However, we cannot expect to observe this matching as the final outcome. In fact, boy 3 prefers boy 6 to his partner, boy 4; in his turn boy 6 is indifferent between boy 3 and his partner, boy 5; boy 5 prefers boy 8 to his partner, boy 6; boy 8 is indifferent between boy 5 and his partner, boy 7; on the other hand boy 7 prefers boy 4 to his partner, boy 8 and boy 4 is indifferent between boy 7 and his partner boy 3. Thus, boys 3, 5 and 7 can act together and be better off by exchanging their partners 6, 8 and 4 among them. It is natural to expect that this exchange will be accepted by 6, 8 and 4, since these boys are indifferent between their current partners under z and the new proposed mates. It is then reasonable to expect that these boys will form a new set of partnerships, {3,6}, {5,8} and {7,4}, and that matching *w*, such that *w(1)=2, w(3)=6, w(5)=8* and *w(7)=4,* will be the resulting matching of this coalitional interaction. Matching *w* is a weak Pareto improvement of matching *z* via coalition {3,4,5,6,7,8}, which weakly blocks matching *z*. Since a weak Pareto improvement of a matching does not create any blocking pair, and *z* is stable, then matching *w* is also stable.

Considering that an exchange of partners is *acceptable* if it does not hurt anybody, it is then evident that an exchange of partners is acceptable only if (1) the agents involved are either all indifferent between their current partners and the new ones or they form a weak blocking coalition and (2) by matching the agents of the weak blocking coalition among them in an appropriate way, a weak Pareto improvement of the current matching is obtained.

Having this in mind observe that once matching *w* is reached no more acceptable exchange of partners is possible. In fact, boys 7 and 5 are assigned to their first choice, so there is no acceptable exchange involving these boys and their partners. On the other hand, any exchange involving some of the remaining boys will necessarily involve boy 8, partner of boy 5, who will not accept such pairwise interaction. Hence, although *z* and *w* are stable, only *w* can be expected to occur.

The pairs {3,2} and {1,8,} are the only weak blocking pairs of matching *w* but the coalition {3,2,1,8} does not produce any weak Pareto-improvement of *w.* Matching *z* is also weakly blocked but only *w* is Pareto-stable.

Observe that in this example coalition {1,2,3,4,7,8} also weakly blocks matching *z,* and yields a weak Pareto-improvement given by the matching *w',* which assigns 1 to 8, 2 to 3, 4 to 7 and 5 to 6. This new matching is also Pareto-stable. No more acceptable exchange of partners will occur. ■

The Pareto-stability concept can be viewed as an intermediate concept between the stability concept and the strong-stability concept. In fact, the set of strongly stable matchings is contained in the set of Pareto-stable matchings, since if a stable matching is not Pareto-optimal then it has a weak Pareto-improvement via some weak-blocking coalition. When preferences are strict, these two sets coincide with the set of stable matchings, because there is no weak blocking coalitions. With indifferences, the previous example illustrates that the set of strongly stable matchings may be a proper subset of the set of Pareto-stable matchings, which may be a proper subset of the set of stable matchings. In that example the set of strongly stable matchings is empty.

It is immediate that Pareto-stable matchings exist if and only if the set of stable matchings is non-empty. In fact, starting at any stable matching that is not Pareto optimal, a finite sequence of weak Pareto-improvements leads to a Pareto-stable matching. This is due to the fact that any weak Pareto improvement of a stable matching is still stable and the set of stable matchings is non-empty by assumption, it is finite and preferences are transitive. Consequently, a Pareto-stable matching always exists for the Marriage model.

Assuming we have a stable matching, a natural question is how to test it for Pareto optimality. Clearly, if x is a stable matching then matching z is a weak Pareto improvement of x if: (i) the set $S=\{j\in N;\ z(j)\neq x(j)\}$ is a weak blocking coalition of x; (ii) $x(S)=z(S)=S;$ (iii) if $j,\ k\in S$ and $z(j)=k$ then (j,k) is a weak blocking pair of x or both agents are indifferent between each other and their mates under x and (iv) if $j\in S$ and j is unmatched under z then j must be indifferent between being unmatched and being matched to $x(j)$. Equivalently, given a stable matching $x,$ we can say that x is Pareto optimal if none of the following requirements occurs:

(1) There are sequences $(j_1,j_2,\ldots,j_q)$ and $(k_1,k_2,\ldots,k_q)$ with $x(j_1)=k_q,$ $x(j_t)=k_{t-1}$ for all $t=2,\ldots,q,$ and such that either (j_t,k_t) is a weak blocking pair of x or both agents are indifferent between each other and their mates under x, for

all $t=1,\dots,q$. Moreover, (j_t,k_t) is a weak blocking pair of x for some $t=1,\dots,q$.

(2) There are sequences $(j_1,j_2,\dots,j_q)$ and $(k_0,k_1,\dots,k_q)$ where k_q is unmatched under x, $x(j_t)=k_{t-1}$ for all $t=1,\dots,q$, k_0 is indifferent between being unmatched at x and being matched to $j_1=x(k_0)$, and either (j_t,k_t) is a weak blocking pair of x or both agents are indifferent between each other and their mates under x, for all $t=1,\dots,q$. Moreover, (j_t,k_t) is a weak blocking pair of x for some $t=1,\dots,q$.

(3) There are sequences $(j_1,j_2,\dots,j_{q+1})$ and $(k_1,\dots,k_q)$ where j_1 is unmatched under x, $x(j_t)=k_{t-1}$ for all $t=2,\dots,q+1$, j_{q+1} is indifferent between being unmatched at x and being matched to $k_q=x(j_{q+1})$, and either (j_t,k_t) is a weak blocking pair of x or both agents are indifferent between each other and their mates under x, for all $t=1,\dots,q$. Moreover, (j_t,k_t) is a weak blocking pair of x for some $t=1,\dots,q$.

In fact, if (1) occurs a weak Pareto improvement of x is obtained by matching j_t to k_t, for all $t=1,\dots,q$ and keeping the other matches. If (2) occurs then a weak Pareto improvement of x is obtained by matching j_t to k_t, for all $t=1,\dots,q$, leaving k_0 unmatched and keeping the other matches. If (3) occurs then a weak Pareto improvement of x is obtained by matching j_t to k_t, for all $t=1,\dots,q$, leaving j_{q+1} unmatched and keeping the other matches.

The remaining part of this paper is devoted to finding the main properties that characterize the Pareto-stable matchings for the Roommate and Marriage models. Our main finding concerns the role played by the simple matchings and Pareto-simple matchings in the characterization of such outcomes[2]. Simple matchings can be defined as follows:

Definition 1. *Matching x is simple if it is individually rational and all of its blocking pairs, if any, are formed with unmatched agents.*

Simple matchings exist even when stable matchings do not, since the matching where everyone is unmatched is simple. Clearly, every stable matching is simple.

The concept of Pareto- simple matching is the following:

[2] The idea of focusing on simple matchings has already been used in the literature for the proof of existence theorems in several matching models. (See the last section of this paper).

Definition 2. *An individually rational matching z extends the individually rational matching x if z is a weak Pareto improvement of x. If z and x are simple we say that z is a simple extension of x. A matching x is Pareto- simple if it is simple and does not have any simple extension.*

That is, matching x is Pareto- simple if it is simple and it is not weakly-dominated by any other simple matching. Pareto-simple matchings always exist since the set of simple matchings is non-empty, finite and preferences are transitive.

The following example, due to Gale and Shapley (1962), shows that the set of Pareto-simple matchings may be disjoint from the set of Pareto-optimal matchings, as well as from the set of Pareto-stable matchings.

Example 2. (*The set of Pareto-simple matchings, the set of Pareto-optimal matchings and the set of Pareto-stable matchings are disjoint*) Consider the Roommate model where the set of boys is $N=\{1,2,3,4\}$. The boys' preferences over acceptable partners are given by:

P(1)=2,3,4,1 P(3)=1,2,4,3

P(2)=3,1,4,2 P(4)=arbitrary

The set of Pareto-stable matchings is empty. There is no Pareto-simple matching that is Pareto-optimal. In fact, matching x where every agent is unmatched is the only simple matching because any other matching has a blocking pair where at least one boy is matched. Then it is Pareto-simple. However, it is not Pareto-optimal since it is weakly dominated by, for example, matching x^1, which matches 1 to 2 and 3 to 4. Matching x^1 is Pareto-optimal but it is not simple. The set of Pareto-optimal matchings also includes x^2, which matches 1 to 3 and 2 to 4 and x^3 which matches 1 to 4 and 3 to 2. ■

The set of Pareto-stable matchings may be a non-empty proper subset of the set of Pareto-simple matchings and of the set of Pareto-optimal matchings, as illustrated in the example below.

Example 3. *(Pareto-stable matchings is a non-empty proper subset of the set of Pareto-simple matchings and of the set of Pareto-optimal matchings.)* Consider the Roommate model where the set of boys is $N=\{1,2,\ldots,6\}$. The boys' preferences over acceptable partners are given by:

P(1)=2,3,1	P(4)=[5,6],4
P(2)=3,1,2	P(5)=4,3,5
P(3)=1,5,2,3	P(6)=4,6

The set of stable matchings is non-empty since matching y, such that $y(1)=2$, $y(3)=5$ and $y(4)=6$, is stable. This is the only stable matching for this market. Since any Pareto-improvement of y must be stable then y is Pareto-optimal, so it is Pareto-stable and Pareto-simple. The pair $\{5,4\}$ weakly blocks y, so the set of strongly stable matchings is empty. Now, let y' be the matching that assigns 5 to 4 and leaves unmatched the other boys. It is easy to see that y' is simple and unstable. On the other hand, there is no way to extend y' to a simple matching. In fact, boy 5 is matched to his first choice. Consequently, any weak-Pareto-improvement of y' will only involve the unmatched boys. However, any arrangement with these boys will have a blocking pair where at least one boy is matched. Then, any weak-Pareto-improvement of y' is not simple, so y' is a Pareto-simple matching. Since it is not stable then it is not Pareto-stable. Matching y' is not Pareto-optimal, since matching z' that assigns 5 to 4, 1 to 2 and leaves unmatched the other agents, for example, is a weak-Pareto improvement of y'. However, matching z' is not simple since the pair {2,3} blocks it and boy 2 is matched. Then, z' is Pareto-optimal but it is not Pareto-stable. ■

As these examples suggest, the set of Pareto-stable matchings is the intersection of two non-empty Pareto sets:

Theorem 1. *The set of Pareto-stable matchings equals the intersection of the set of Pareto-simple matchings with the set of Pareto-optimal matchings.*

The proof of this result is straightforward. If a matching is Pareto-stable then it is simple and it is not weakly dominated by any individually rational matching, in particular it is not weakly dominated by any simple matching, so it is a Pareto-simple matching. Conversely, if a matching is simple and Pareto-optimal then it must be stable, since otherwise it would have a blocking pair formed with unmatched agents and so, by matching these agents with each other, we would get a weak-Pareto-improvement of the given matching, which would contradict its Pareto-optimality.

Thus, by Theorem 1, in order to show that Pareto-stable matchings exist it is sufficient to find just one Pareto-simple matching that is Pareto-optimal. It turns out that under strict preferences, if Pareto-stable matchings exist then every Pareto-simple matching must be Pareto-optimal, so every Pareto-simple matching must be stable. In fact, Theorem 2 provides a characterization of the set of Pareto-stable matchings as the set of Pareto-simple matchings. For the Roommate model it is required strictness of the preferences and non-emptiness of the set of stable matchings. For the Marriage model it is not imposed any restriction.

Theorem 2. *a) Consider the Roommate model with strict preferences and suppose the set of stable matchings is non-empty. Then the set of Pareto-stable matchings equals the set of Pareto-simple matchings.*
b) Consider the Marriage model. Then the set of Pareto-stable matchings equals the set of Pareto-simple matchings.

The idea of the proof of this result is to show that every Pareto-simple matching is stable. If this is established then every Pareto-simple matching is Pareto optimal, since otherwise there would be a weak Pareto improvement of it, which would still be stable, so it would be simple, which is a contradiction. This is equivalent to show that every unstable and simple matching has a simple extension:

Proposition 1. *a)Consider the Roommate model with strict preferences. If the set of stable matchings is non-empty then every unstable and simple matching has a simple extension.*
b)Every unstable and simple matching for the Marriage model has a simple extension.

The proof of this proposition is given in the next section. Unlike the other results of this paper it is not straightforward. It is easy to obtain an extension B of an unstable and simple matching A for the Roommate model. It is enough to keep the partnerships formed under A, if any, and to add some new partnerships. Of course, these new partnerships are formed with blocking pairs of A. What is not clear is that if the set of stable matchings is non-empty and preferences are strict, then matching B can be constructed so that it is still simple. Without these requirements such construction of B is not always possible. Indeed, to match the correct blocking pairs of A is the inventive part of the proof. (Remember that matching x of Example 2 and matching y' of Example 3 are simple and unstable matchings but they cannot be extended to a simple matching. In the first case there is no stable matchings in the market and in the second case the preference of player 4 is not strict).

The proof of Proposition 1-(a) uses a key lemma. This is a technical result, which is a one-sided version of the Decomposition lemma for the Marriage model from Gale and Sotomayor (1985). For part (b), the proof strongly uses the fact that the Marriage model has two sides.

By Proposition 1, in order to conclude that the set of Pareto-stable matchings for the Roommate model with strict preferences is empty, it is enough to find just one Pareto-simple and unstable matching. See the example below.

Example 4. (*An application of Proposition 1-(a)*) Consider the Roommate model where the set of boys is $N=\{1,2,\ldots,7\}$. The boys' preferences over acceptable partners are given by:

P(1)=5, 6, 1	P(4)=6, 5, 4	P(7)=2, 1, 3, 7
P(2)=3, 7, 2	P(5)=4, 1, 6, 5	
P(3)=7, 2, 3	P(6)=1, 4, 6	

The matching that assigns 4 to 5, 1 to 6 and leaves the other agents unmatched is simple and unstable. Any extension of this matching will match a pair of agents in $\{2,3,7\}$. However, one of the agents in the pair will form a blocking pair with the agent left unmatched. Hence, the original matching does not have a simple extension. Since the preferences are strict, we need not check that every Pareto-simple matching is unstable. (Observe that the matching that assigns 4 to 6, 1 to 5 and leaves the other agents unmatched is also Pareto-simple and unstable). Proposition 1 implies that the set of stable matchings is empty, so the set of Pareto-stable matchings is also empty. ■

The following corollary is then immediate:

Corollary 1. (a) *Suppose the preferences in the Roommate model are strict. The set of stable matchings is non-empty if and only if every unstable and simple matching has a simple extension.*

(b) *The set of stable matchings for the Marriage model is always non-empty.*

The fact that the condition in (a) is necessary is immediate from Proposition 1. It is sufficient since, if every unstable and simple matching has a simple extension then the Pareto-simple matchings must be stable. The conclusion follows since Pareto-simple matchings always exist. The proof of part (b) is immediate from Proposition 1 b), since a Pareto-simple matching always exists and cannot have a simple extension.

It is easy to construct examples for the Marriage model where, as in the Roommate model, the set of strongly stable matchings is empty. However, it is well known that the existence of two sides in the Marriage market causes fundamental differences between the two models. There are properties of the Marriage model which depend on the two-sidedness of the market, as the non-emptiness of the set of stable matchings under any kind of preferences and the lattice property of the set of stable matchings when preferences are strict. This last property guarantees the existence of the optimal stable matchings for each side of the market. Moreover, it implies that if the two optimal stable matchings coincide then the set of stable matchings is a singleton. When preferences need not be strict, the lattice property may fail to hold even when the man-optimal and the woman-optimal stable matchings exist. Moreover, the man-optimal stable matching may coincide with the woman-optimal stable matching when the set of stable matchings is not a singleton. See the example below.

Example 5. (*The woman-optimal and the man-optimal stable matchings coincide but the set of stable matchings is not a singleton*) Consider the Marriage model where the set of agents are $M=\{m_1, m_2\}$ and $W=\{w_1,w_2\}$. Agent m_1 is indifferent between w_1 and w_2; m_2 prefers w_1 to w_2; w_1 is indifferent between m_1 and m_2 and w_2 prefers m_1 to m_2. Both matchings under which no agent is unmatched are stable and are the only stable matchings. The matching y^1 where $y^1(m_1)=w_1$ and $y^1(m_2)=w_2$ is not Pareto-optimal and is not strongly stable. It is weakly Pareto improved by matching y^2 where $y^2(m_1)=w_2$ and $y^2(m_2)=w_1$. Matching y^2 is strongly stable. Matching y^2 is clearly optimal for the men and for the women but matching y^1 is also stable. ■

The key lemma mentioned above is also used in this paper to extend, to the Roommate model with strict preferences, two well-known properties for the Marriage model with strict preferences. The first result reflects an opposition of interests between the two players involved in a partnership regarding two Pareto-stable matchings. It asserts that if x and y are Pareto-stable matchings and j prefers x to y then j is matched under both matchings and both of his mates prefer y to x. The second result implies that the set of trading agents at a simple matching can be regarded as a sort of *stable coalition* in the sense that such agents always make their transactions under a stable matching within the same pool. In particular, the set of matched agents under a Pareto-stable matching is the same under any Pareto-stable matching. The proof of both results will be given in the next section.

The present work also addresses the case of non-necessarily strict preferences. Similar results to those stated under the assumption of strict preferences, by focusing on strongly stable matchings and strongly simple matchings, are obtained and presented in section 3. The proofs of these results follow the lines of the proofs of the corresponding results under strict preferences and are left to the reader. Some final conclusions and related work are presented in the last section.

2. TECHNICAL PROOFS

The following result is a technical lemma that will be used to prove some of our results. The idea is the following. Given a pair of matchings, one stable and the other simple, non-necessarily stable, the set of agents who prefer one matching to the other can be decomposed into two disjoint sets, such that the agents from one set are matched to the agents of the other set under both matchings. A special case of this result, in which both matchings are stable, was obtained in Gale and Sotomayor (1985) for the Marriage model and called by these authors Decomposition lemma.

Lemma 1. *Suppose the preferences are strict. Consider the Roommate model* (N,P)*. Let* x *be a simple matching and let* y *be a stable matching. Let* $T=\{j\in N;\ x(j)\neq j\}$, $M_x=\{j\in N;\ x(j) >_j y(j)\}$ *and* $M_y=\{j\in T;\ y(j) >_j x(j)\}$. *Then* $x(M_x)= y(M_x)=M_y$ *and* $x(M_y)=y(M_y)=M_x$.

Proof. First observe that all j in M_x are matched under x and all j in M_y are matched under y. If j is in M_x then $x(j)\neq y(j)$ and $k=x(j)$ is in M_y, for otherwise the strictness of the preferences will imply that $j=x(k) >_k y(k)$, so y will be blocked by j and k, which contradicts the assumption that y is stable. On the other hand, if k is in M_y then $y(k)\neq x(k)$ and $j=y(k)$ is in M_x, for if not the strictness of the preferences will imply that $k=y(j) >_j x(j)$, so x will be blocked by j and k. However, k is in T, so k is matched under x, which contradicts the fact that x is simple. Therefore,

$x(M_x)\subseteq M_y$ and $y(M_y)\subseteq M_x$. Since x and y are one-to-one and M_x and M_y are finite sets, the conclusion follows. Hence the proof is complete.■

Proposition 1. (a) *Consider the Roommate model (N,P). Let x be an unstable and simple matching. If the set of stable matchings is non-empty then x can be extended to a simple matching.*

(b) *Consider the Marriage model (M,W,P). Let x be an unstable and simple matching. Then x can be extended to a simple matching.*

Proof of part (a). Let y be a stable matching. Using the notation of Lemma 1, set $S\equiv M_x\cup M_y$. To prove that x has a simple extension define matching z to agree with x on S and with y on N-S. Lemma 1 implies that all of S are matched among them under x and y, so z is well defined. Clearly, z is individually rational and restricted to S (because x is simple) or N-S (because y is stable) is stable. Then, if there is a blocking pair $\{j,k\}$ we must have that $j\in N$-S and $k\in S$. Then $j>_k z(k)=x(k)$ and $k>_j z(j)=y(j)\geq_j x(j)$, so $\{j,k\}$ blocks x. However, k is matched at x, which contradicts the fact that x is simple. Hence, z is stable. It is also clear that z extends x. In fact, we have that $z(j)=x(j)$ for every $j\in S$ and $z(j)=y(j)\geq_j x(j)$ for every $j\notin S$. Furthermore, $x\neq z$ due to the fact that x is unstable and z is stable. Hence, $z(j)\geq_j x(j)$ for every j, with strict preference for at least one $j\in N$. Then, z extends x and we have proved part (a).

Proof of part (b). The fact that x is simple and unstable implies that every blocking pair is formed with unmatched agents. Choose (m_1,w_1) such that w_1 is one of m_1's favorite blocking partners (m_1 may have more than one favorite blocking partner since preferences need not be strict). Now let x_1 be the matching that matches m_1 with w_1 and agrees with x on every other agent. Clearly, every woman weakly prefers x_1 to x and w_1 strictly prefers x_1 to x. Also every man weakly prefers x_1 to x. Therefore x_1 is a weak Pareto-improvement of x. If x_1 is unstable then choose (m_2,w_2) such that w_2 is one of the m_2's favorite blocking partners. (Note that w_2 might be w_1). Of course, (m_2,w_2) also blocks x and since x is simple we must have that m_2 and w_2 are unmatched under x. By construction of x_1 we have that m_2 is also unmatched at x_1. Now let x_2 be the matching that matches m_2 to w_2, leaves m_1 unmatched in case $w_2=w_1$, and otherwise agrees with x_1 on every other agent. Clearly, all women weakly prefer x_2 to x_1, and by transitivity they prefer x_2 to x; w_2 strictly prefers x_2 to x_1, so she strictly prefers x_2 to x; every man weakly prefers x_2 to x. Then, x_2 is a weak Pareto-improvement of x. Again, if x_2 is unstable then choose (m_3,w_3) such that w_3 is one of the m_3's favorite blocking partners. (Note that w_3 might be w_1 or w_2). Of course, (m_3,w_3) also blocks x and since x is simple we must have that m_3 and w_3 are unmatched under x. By construction of x_2 we have that m_3 is also unmatched at x_2. Then, let x_3 be the matching that matches m_3 with $w_3,$ leaves m_j unmatched in case $w_3=w_j$ for some $j\in\{1,2\},$ and otherwise agrees with x_2 on every other agent, and so on. Following this procedure, we can construct a sequence of matchings, $x_1, x_2, x_3,\ldots$, where every term of the sequence is a weak Pareto-improvement of $x,$ it is weakly preferred by every woman and it is strictly preferred by at least one woman to the previous terms, so the matchings of this sequence are distinct. Since the number of matchings is finite we must have that this sequence ends with a matching x^* which is a weak Pareto-improvement of x and does not have any blocking pair. Hence x^* is stable (so it is simple) and extends x, so the proof is complete. ■

Remark 1. In the proof of Proposition 1-b), matchings $x_1, x_2, x_3,\ldots$, are not necessarily simple. They belong to a special class of matchings, which we call here *semi-simple matchings*. They are *the individually rational matchings such that all of its blocking pairs, if any, have a single man*[3]. Therefore, every simple matching and every stable matching are semi-simple. We learned from the proof of Proposition 1-b) that we can construct a sequence of distinct semi-simple matchings, starting at any semi-simple matching, by satisfying some blocking pairs conveniently chosen, leading to a stable matching. Unlike the algorithm of Gale and Shapley, this final matching is not necessarily an extreme point of the lattice of the stable matchings.■

The following two properties of the Pareto-stable matchings, already discussed in the previous section, can also be derived from Lemma 1 under the assumption of strict preferences.

Property 1. *Suppose the preferences are strict. Let x and y be Pareto-stable matchings. If j prefers x to y then $k=x(j)\neq j$, for some k, and $h=y(j)\neq j$, for some h with $h\neq k$. Furthermore, both k and h prefer y to x.*

[3] This concept was first used in the literature in Sotomayor (1996).

Proof. Using the notation of Lemma 1, if $x(j)>_j y(j)$ then $j \in M_x$. It follows from this lemma that j is matched under x and under y and both mates belong to M_y. ■

Property 2. *Suppose the preferences are strict. Let x be a simple matching and let y be a stable matching. If $j \in N$ is unmatched under y then j is unmatched under x.*

Proof. Using the notation of Lemma 1, if j was unmatched under y, but he was matched under x, then he would belong to M_x, so he would be matched under y by Lemma 1, contradiction. ■

The following corollary is immediate:

Corollary 2. *Suppose the preferences are strict. Then the set of unmatched agents under a Pareto-stable matching is the same for every Pareto-stable matching.*

3. WHEN PREFERENCES NEED NOT BE STRICT

If we relax the assumption of strictness of the preferences, we can obtain similar results to those presented in the previous section by focusing on strongly stable matchings and strongly simple matchings. The proofs of these results are trivial adaptations of the proofs presented in section 2 and will be left to the reader.

Definition 3. *The matching x is* ***strongly simple*** *if it is individually rational and no matched agent is part of a weak blocking pair. Matching x is called* ***Pareto-strongly simple*** *if it is* ***strongly simple*** *and it is not weakly dominated by any* ***strongly simple*** *matching.*

Clearly, the matching where every agent is unmatched is strongly simple. Also, if a matching is strongly stable then it is strongly simple. From the definition above, if a matching is Pareto-simple and strongly simple then it is Pareto-strongly simple. If a matching is strongly stable then it is Pareto-optimal and strongly simple, so it is Pareto-strongly simple. A sort of converse is given by Proposition 2 that asserts that under the assumption that the set of strongly stable matchings is non-empty, we have that every Pareto-strongly simple matching is strongly stable. Properties 3 and 4 are the corresponding extensions of Properties 1 and 2 to the case in which preferences are non-necessarily strict. The proofs of these three results use the straightforward extension of Lemma 1 where stable and simple are replaced by strongly stable and strongly simple, respectively, with the adequate adaptations.

Proposition 2. *Suppose the set of strongly stable matchings is non-empty. Then, the set of Pareto-strongly simple matchings equals the set of strongly stable matchings.*

Property 3. *Let x and y be strongly stable matchings for the Roommate model (N,P). If j prefers x to y then $k=x(j)\neq j$, for some k, and $h=y(j)\neq j$, for some $h\neq k$. Furthermore, both k and h prefer y to x.*

The following corollary asserts that there is an opposition of interests between the two sides of the Marriage market along the whole set of strongly-stable matchings.[4]

Corollary 3. *Consider the Marriage market (M,W,P). Let x and y be strongly stable matchings. Then, all men like x at least as well as y if and only if all women like y at least as well as x.*

Proof. Suppose all men like x at least as well as y. If there is some woman w such that $x(w)>_w y(w)$ then $x(w)=m$ for some $m\in M$ and m prefers y to x by Property 1, contradiction. Hence, no woman prefers x to y and so all women like y at least as well as x. The other direction is proved similarly.■

Given two strongly stable matchings, x and y, the trading agents at x, who are not indifferent between the two outcomes, trade among themselves at y. Consequently, an unmatched agent under y either is also unmatched under x or is indifferent between his/her mate under x and being unmatched. Formally,

Property 4. *Let x and y be strongly stable matchings. If j is unmatched under y then j is indifferent between $x(j)$ and being unmatched.*[5]

---

[4] Knuth (1976) proved this result for the case where preferences are strict.

[5] This result was proved for the Marriage market and for the College admission market with strict preferences by Gale and Sotomayor (1985a,b). A different proof is provided by McVitie and Wilson (1970) for the particular Marriage market where all men and women are mutually acceptable. For the College admission model, Roth (1986) proved that if a college does not fill its quota under some stable matching then it is matched to the same set of students under every stable matching.

4. CONCLUSION AND RELATED WORK.

Basically, we can identify four sets, each a superset of the next: The weak Pareto frontier, the set of stable matchings, the set of Pareto-stable matchings and the set of strongly stable matchings. The set of stable matchings is always a subset of the set of weak Pareto-optimal matchings, understood as those allocations for which the grand coalition cannot strongly block, i.e., the grand coalition does not cause a strict Pareto improvement. We proved that when preferences are strict the three last sets coincide. With indifferences, Pareto-stability plays the role of an intermediate solution concept: the set of Pareto-stable matchings may be a proper subset of the set of stable matchings and may properly contain the set of strongly stable matchings. It is non-empty when the set of stable matchings is non-empty.

Of course we may have Pareto-optimal matchings that are unstable. However, if a Pareto-optimal matching is simple then it is stable. In fact, we characterized the set of Pareto-stable matchings as the intersection of two non-empty sets: the set of Pareto-simple matchings and the set of Pareto-optimal matchings. For the Marriage model, without imposing any restriction, and for the Roommate model, under the assumption of strict preferences and non-emptiness of the set of stable matchings, the Pareto-stable matchings are exactly the Pareto-simple matchings. With indifferences, Example 3 showed that there may be Pareto-simple matchings for the Roommate model that are not Pareto-stable.

From the conceptual point of view, it is natural that in a decentralized setting, recontracts between pairs of agents already allocated according to a stable matching leading to a weak Pareto improvement of the original matching should be expected.

From the technical point of view, the characterization of Pareto-stable matchings in terms of Pareto-simple matchings provides us more understanding of the role of unmatched versus matched agents. Simple matchings capture a sort of dynamic flavor to coalition formation, without an explicit model of dynamics. For the Roommate model under strict preferences, starting from an unstable and simple matching (for example, the matching where every player is unmatched, if it is unstable), it is possible to gradually increase cooperation by making weak-Pareto improvements and still staying within the set of simple matchings, until no pairwise transaction is able to benefit all agents involved, or until the matching cannot be simple anymore. In the former case a stable matching has been reached. In the latter case, the set of stable matchings is empty, so increase in payoffs is only available through non-optimal (in the selfish sense) cooperation of some agents. These weak-Pareto improvements produce a finite sequence of simple matchings that keep the current trades and add new ones, leading to a Pareto-simple matching. Thus, once a transaction is done, it will not be undone at the subsequent matching. Only agents who are not currently trading are able, by trading among them, to be better off.

This dynamics can be identified, for example, in the steps of the Top Trading Cycles algorithm due to Gale, for the Housing market of Shapley and Scarf (1974). Each cycle formed by the Top Trading Cycles algorithm produces a simple allocation, defined in the straightforward manner for the Housing model (see Sotomayor, 2005). In each step one more cycle is added. The sequence of such simple allocations converges to a core outcome as soon as no more cycles can be added.

For the Marriage model with non-necessarily strict preferences, starting with any simple matching (for example, the simple and semi-simple matching in which every agent is unmatched), we can construct a sequence of weak Pareto-improvements of the original matching and still stay within the set of semi-simple matchings. Each term of the sequence is obtained by matching some blocking pair where the man is currently unmatched and the woman is one of his favorite blocking partners. In this sequence of pairwise interactions the women currently matched stay matched (not necessarily to the same mate). Furthermore, the trades can be done without hurting any woman. This sequence of distinct and semi-simple matchings clearly ends with a matching that does not have any blocking pair, so the final matching is stable. This is the general basis which underlies the construction of the algorithm of Gale and Shapley with the men proposing. The matching produced in each step of this algorithm is semi-simple and it is a weak Pareto-improvement of the simple matching in which every agent is unmatched. However our approach is different since: (i) we do not require to break ties when preferences are not strict; (ii) the initial point of our sequence need not be the matching where everyone is unmatched, so (iii) the final matching is not necessarily the man-optimal stable matching.

This justifies Pareto-stability.

Stability and Pareto optimality only for the students were required in Erdil and Ergin (2007) to replace the standard concept of student optimal stable matching for the school choice model. That is, according to this concept, a stable matching is called a "student optimal stable matching" if no stable matching is weakly preferred by all students to that matching. These authors present a simple procedure to compute a “student optimal stable matching”, under the assumption that priorities are weak and the preferences of the students are strict. In this model, since preferences of schools are not considered, the exchange of partners does not always produce a weak Pareto improvement of the given matching and it does not always yield a stable matching. Then, in order to compute a stable and Pareto optimal matching for the students, the authors considered exchange of partners only inside a *stable improvement cycle*, defined as a cycle of students who each prefer the school to which the next student in the cycle is matched, and each of whom is one of the school's most preferred candidates among the students who prefer that school to their current match. Therefore, a stable improvement cycle produces a stable matching in which some schools may be worse off and the students of the cycle are strictly better off. Starting with a student-proposing deferred acceptance algorithm with arbitrary tie-breaking of non-strict preferences by schools, the authors construct a computationally efficient algorithm which, in each step, improves the current matching for the students, by finding and satisfying stable improvement cycles, until no more remain. The outcome of such an algorithm is then a stable matching that is Pareto optimal with respect to students.

The idea of proving the non-emptiness of the core by showing that every Pareto-simple outcome must be stable has been explored in the literature via adaptations of the concept formulated here of simple matching. In Sotomayor (2005), for example, it was introduced the concept of simple allocation for the one-sided market (not matching market) of Shapley and Scarf (1974). There, it was proved that every Pareto-simple allocation must be in the core.

For the two-sided matching models it has been more convenient to work with the concept of *semi-simple outcome,* whose discrete version was introduced in Sotomayor (1996) for the Marriage market. The extension of this concept to a general discrete many-to-many matching market (in particular, to the College Admission model), where preferences are substitutable and non-necessarily strict, was introduced in Sotomayor (1999) to prove the non-emptiness of the set of pairwise-stable matchings. A similar idea was used in Sotomayor (2004) to show the existence of Nash equilibria in an implementation mechanism for the discrete many-to-many matching model. The continuous version of that concept was used in Sotomayor (2000) for the Assignment game of Shapley and Shubik (1972) and for a unified two-sided matching model.

The possibility of obtaining a stable matching for the discrete matching models, by starting from an arbitrary unstable matching and successively satisfying blocking pairs, has been the subject of several papers, motivated by an open problem posed by Knuth (1976): *Let* μ *be an unstable matching for the Marriage model (M,W,P) with strict preferences. Is there a sequence of matchings* $\mu=\mu_1, \ldots, \mu_k$, *such that* μ_k *is stable and for each* $i=1,\ldots,k-1$, *there is a blocking pair* $(m_i,w_i) \in MxW$ *for* μ_i, *such that the subsequent matching* μ_{i+1} *matches* m_i *to* w_i, $\mu_i(m_i)$ *to* $\mu_i(w_i)$ *in case* m_i *and* w_i *are matched under* μ_i, *in case only one of these agents is matched then the corresponding mate is left unmatched, and the other matches at* μ_i *are kept unchanged?*

Knuth illustrates with an example that if the sequence of unstable matchings is not conveniently formed then it may cycle. Unlike the unstable simple matchings, the members of the selected blocking pairs of these unstable matchings need not be unmatched. In case they are matched, it is a crucial point in Knuth's problem that *their current mates must be matched to each other in the subsequent matching.* Thus, it is not correct the assertion that has been made in the literature that Roth and Vande Vate (1990) answered Knuth's question in the affirmative. In the approach treated by Roth *et al.*, if the blocking pair (m_i,w_i) is selected then $\mu_i(m_i)$ and $\mu_i(w_i)$ are left unmatched. This case is easily solved by using a version of the deferred-acceptance algorithm of Gale and Shapley.

According to our results, Knuth's problem is solved for the Marriage model when μ is unstable and semi-simple. It is also solved for the Roommate model with strict preferences when μ is unstable and simple. Under indifferences this problem may have no solution for the Roommate problem, even when the set of stable matchings is non-empty. In fact, in our Example 3, matching y is the only stable matching and matching y' is unstable. If we start with y', the procedure proposed by Knuth never reaches y; it always cycles.

Although the result of Roth and Vande Vate (1990) does not solve Knuth's problem, an immediate corollary of it is that a random process that begins from an arbitrary matching and continues by satisfying a randomly selected blocking pair must eventually converge with probability one to a stable matching, provided each blocking pair has a probability of being selected that is bounded away from zero. This corollary has given origin to several other papers. Chung (2000), for example, proves that this random-paths-to-stability result applies to the Roommate model under the assumption of non-necessarily strict preferences, as long as some sufficient condition for the existence of stable matchings, called *no-odd-rings*, is satisfied. If the no-odd-ring condition is not satisfied but preferences are strict and stable matchings exist then the Roth and Vande Vate path still converges to a stable matching (Diamantoudi *et al.* 2004).

Inarra, Larrea and Molis (2008) generalizes for the Roommate model the result of Diamantoudi et al. (2004), and consequently that of Chung (2000) under strict preferences and the one of Roth and Vande Vate (1990). They prove that from any matching for the Roommate model with strict preferences, there exists a path, given by a proposal-rejection procedure, which reaches a specific matching[6] that has the property to be stable when the set of stable matchings is non-empty.

Kojima and Unver (2006) study the convergence to stability in many-to-many matching models. Klaus and Klijn (2007) analyze this convergence for matching markets with couples.

The presence of indifference in the preferences of the agents in the discrete matching models, affecting the existence of strongly stable matchings, has been considered by several authors. Irving (1994), for example, formulated an $O(n^4)$ algorithm for determining whether a given instance of the Marriage model, with n men and n women and complete lists of non-necessarily strict preferences, admits a strongly stable matching and for constructing one if it does. An extension of this algorithm by allowing incomplete lists of preferences is presented in Manlove (1999). Irving, Manlove and Scott (2000) present an algorithm to determine whether a given instance of the Hospitals/Residents Problem with indifferences admits a *super-stable matching* and, if it does, to construct such a matching. They define a super-stable matching x as a stable matching for which there is no pair, resident/hospital, not matched to each other, such that (a) the resident is indifferent between his partner and the hospital and (b) the hospital is indifferent between its worst partner and the resident.

REFERENCES.

Abeledo, H. and Isaak, G. (1991), "A characterization of graphs that ensure the existence of stable matchings", *Mathematical Social Sciences* vol. 22, Issue 1, 93-96.

Chung, Kim-Sau (2000) "On the existence of stable roommate matchings", *Games and Economic Behavior* 33, 206-230.

[6] This matching is called P-stable matching because it is associated to a stable partition P, concept introduced by Tan (1991) for the Roommate model with strict preferences. It is a partition of the agents into ordered sets satisfying a notion of stability between sets and also within each set. This author proves that if there is a stable partition containing an odd ring, then there is no stable matching.

Diamantoudi E., Miyagawa E., Xue L. (2004) "Random paths to stability in the roommate problem", *Games and Economic Behavior* 48, 18-28.

Erdil, A. and Ergin, H. (2007) "What's the matter with tie-breaking? Improving efficiency in school choice?", *working paper.*

Gale, D. and Shapley, L. (1962) "College admissions and the stability of marriage", *American Mathematical Monthly* 69, 9-15.

Gale, D. and Sotomayor, M. (1985-a) "Some remarks on the stable matching problem" *Discrete Applied Mathematics,* 11, 223-32.

Gale, D. and Sotomayor M. (1985-b) "Ms. Machiavelli and the stable matching problem", *American Mathematical Monthly,* 92, 261-268.

Inarra, E., Larrea, C. and Molis, E. (2008) "Random paths to P-stability in the roommate problem", *International Journal of Game Theory* vol.36, 461-471.

Irving, R. W. (1994) "Stable marriage and indifference", *Discrete Applied Mathematics,* 48, 261-272.

Irving, R. W., Manlove, D. F. and Scott, S. (2000) "The hospitals/residents problem with ties", *Lecture Notes in Computer Science* 1851, 259-271.

Klaus, B. and Klijn, F. (2007), "Paths to stability for matching markets with couples", *Games and Economic Behavior* 58, 154-171.

Knuth, D. E. (1976) "Marriages Stables" (*Les Presses de L'Universite de Montreal, Montreal, Que.*

Kojima, F. and Unver, MU. (2006) "Random paths to pairwise-stability in many-to-many matching problems: a study on market equilibration", *International Journal of Game theory,* v. 36, 473-488.

Manlove, D. (1999) "Stable marriage with ties and unacceptable partners", Technical Report TR-1999-29, University of Glasgow.

McVitie, D. G. and Wilson L. B. (1970) " Stable marriage assignments for unequal sets*", BIT,* 10, 295-309.

Roth, A. (1986) "On the allocation of residents to rural hospitals: a general property of two-sided matching markets", *Econometrica,* 54,425-427.

Roth, A. and Sotomayor, M (1990) "Two-sided matching. A study in game-theoretic modeling and analysis" *Eonometric Society Monographs,* 2nd ed., Vol. 18, Cambridge University Press.

Roth, A. and Vate V. (1990) "Random paths to stability in two-sided matching" *Econometrica* 58, 1475-1480.

Shapley, L. and Scarf, H. (1974) “On cores and indivisibility” *Journal of Mathematical Economics,* 1, 23-8.

Shapley, L. and Shubik, M. (1972) "The assignment game I: The core", *Journal of Game Theory* 1, 111-130.

Sotomayor, M. (2005) “An elementary non-constructive proof of the non-emptiness of the core of the housing-market of Shapley and Scarf” *Mathematical Social Sciences* 50, 298-303.

_____________(2004) "Implementation in the many to many matching market", *Games and Economic Behavior,* v. 46, p. 199-212.

___________(2000) "Existence of stable outcomes and the lattice property for a unified matching market", *Mathematical social Science* 39 , 119-132.

___________(1999) "Three remarks on the many-to-many stable matching problem", *Mathematical Social Sciences* 38, 55-70

___________(1996) "A non-constructive elementary proof of the existence of stable marriages", *Games and Economic Behavior* 13, 135-137.

Tan, J.J.M. (1991) “A necessary and sufficient condition for the existente of a complete stable matching”, *Journal of Algorithms* 12, 154-178.