WHAT MATCHINGS CAN BE STABLE? THE REFUTABILITY OF MATCHING THEORY

FEDERICO ECHENIQUE

California Institute of Technology

Abstract. When can a collection of matchings be stable, if preferences are unknown? This question lies behind the refutability of matching theory. A preference profile rationalizes a collection of matchings if the matchings are stable under the profile. Matching theory is refutable if there are observations of matchings that cannot be rationalized. I show that the theory is refutable, and provide a characterization of the matchings that can be rationalized.

1. Introduction

Two-sided matching models are described by two sets of agents (think of workers and firms or men and women) and a preference relation for each agent in each set over potential partners from the other set. The theory studies matchings of agents that have the core property; the core matchings are called "stable." Matching models have been studied very extensively since Gale and Shapley's (1962) seminal paper (Al Roth's online bibliography lists almost 500 papers).

The literature has focused on, given agents' preferences, determining which matchings may occur. It assumes that the stable matchings are the ones that may occur, and proceeds to study the structure of stable matchings. Instead, I study the problem of which matchings can be stable when agents' preferences are not know. Concretely, given a collection of matchings, $\mu_1, \mu_2 \ldots \mu_k$, I ask if there are preferences for the agents involved so that all these matchings are stable. When this is the case, I say that the set of matchings is *rationalizable*.

The problem is important because it is often difficulty to infer agents' preferences. It is important to understand the implications of the

Date: April 24, 2006.

I thank David Ahn, Chris Chambers, Alekos Kechris, Hideo Konishi, Tayfun Sönmez and seminar audiences at UC Berkeley, Boston College, and audiences at the Wallis/Thomson Conference and the Caltech SISL retreat.

theory—its predictions—when preferences are unknown. If one assumes that matchings are observable but preferences are not, one needs to know if a set of matchings can be incompatible with the theory—that is, if the theory has testable implications. And, if the theory is testable, one needs a characterization of the matchings that can be stable in order to empirically determine the validity of the theory in particular instances.

The problem of rationalizing matchings is part of a larger research program of studying refutability in economics. Early results in this program include Samuelson's (1947) and Afriat's (1967) theories of *revealed preference*. Refutability has been studied in General Equilibrium Theory and Non-cooperative Game Theory, but not for matching problems. In matching, one can think of the agents as choosing a partner from the opposite side of the market, but revealed preference theory has no bite because agent 1 not choosing agent 2 does not necessarily mean that 1 is revealed preferred to 2; it can also mean that 2 prefers not to be with 1.

In this paper, I show: (1) that the theory is testable, so there are non-rationalizable sets of matchings; and (2) I provide a series of results, leading up to a characterization, of the rationalizable sets of matchings.

The classical result on stable matchings imply a coincidence of interest within the same side of the market, and opposition of interest across the market. I show that, essentially, stability is characterized by a version of the coincidence/opposition property for any pair of matchings. In the classical results, certain distinguished matchings have a coincidence/opposition of interest for all agents. And for any two matchings, the coincidence/opposition holds for certain agents. I show that there is a coincidence/opposition property that holds for all agents in any pair of matchings, and that this property is essentially the content of the theory.

A simple but important insight is that the testable implications of the theory stem from agents who are matched to the same partner in more than one matching. Thus, in any empirical test of the theory, being able to treat some individuals in different matchings as the same individual is crucial. For example, consider data on a cross-section of matches between buyers and sellers of a certain good. Each match corresponds to the outcome in one market, for example domestic markets for a good that is not traded internationally. One can then assume that firms with similar observable characteristics (size, technology) have the same preferences over potential buyers and are considered to be the same by the buyers. These are exactly the assumptions in (positive) empirical work on matching. One recent example is Choo and Siow

(2006), working on marriage matchings. Choo and Siow assume that there are "types" of men and women, and that individuals of the same type have the same preferences over partners and are considered identical by their potential partners. Choo and Siow build on the theoretical results of Dagsvik (2000), who also assumes that the population can be partitioned in types according to their observable characteristics. An alternative, but related, approach is to model preference as parametrically dependent on the agents' observable characteristics. This is the approach in econometric studies of matching markets (a recent example is Hitsch, Hortaçsu, and Ariely (2006)).

In sum, in testing matching theory, it is crucial to control agents' preferences using observables. My methods are viable using one procedure for controlling preference, and this procedure is already being used by empirical researchers.

I should make a distinction between the positive empirical applications I have in mind and normative applications of matching theory. The latter have been very successful Roth and Peranson (1999); Roth, Sönmez, and Ünver (2004); Abdulkadiroglu, Pathak, Roth, and Sönmez (2005), and in no way rely on rationalizing matchings. The positive applications, to labor economics and marriage markets, do rely on finding testable implications of matching theory.

I also obtain some secondary results. The first is that, if a collection of matchings is rationalizable, then it is typically rationalizable by a large number of different preference profiles. So matching theory is not exactly identified, in the econometric sense. This confirms an argument Choo and Siow (2006) make informally by counting observations and unknowns. In fact, it is not clear from Choo and Siow's argument that the theory is testable; my results imply that it is testable, despite the existence of fewer observations than unknowns.

Finally, I consider the problem of when purely randomly generated matchings would be rationalizable. I show the, admittedly unsurprising, result that the probability of rationalizing a fixed number of random matchings remains bounded away from zero as the number of agents grows. So for large populations, one needs large samples of matchings for the theory to have power.

2. Statement of the problem.

2.1. **Preliminary definitions.** In this paper, I use the language of graph theory, but no results from graph theory. A *graph* is a pair $G = (V, E)$, where V is a set and E is a binary relation on V, i.e. a subset of $V \times V$. The set V is called the *vertex set* of G, and E is the set of *edges* of G. Say that G is *loop-free* if $(v,v) \notin E$, for all $v \in V$. Say that G is undirected if $(v,v') \in E$ implies that $(v',v) \in E$, that is if E is a symmetric binary relation.

A *path* is a sequence $v_1, v_2, \ldots, v_K$ in V with $K > 1$ and $(v_k, v_{k+1}) \in E$ for all k, $1 \leq k \leq K-1$. Say that v and v' are *connected* if there is a path $v_1, v_2, \ldots, v_K$ with $v = v_1$ and $v' = v_K$ or a path $v_1, v_2, \ldots, v_K$ with $v = v_K$ and $v' = v_1$. Say that v and v' are *disconnected* if they are not connected. A *connected component* of G is a set $C \subseteq V$ such that, for all $v, v' \in C$, v and v' are connected. The set of all connected components of G form a partition of V.

2.2. **The Model.** Let M and W be disjoint, finite, sets. I call men the elements of M and women the elements of W. A *matching* is a function $\mu : M \cup W \rightarrow M \cup W \cup \{\emptyset\}$ such that for all $w \in W$ and $m \in M$,

(1) $\mu(w) \in M \cup \{\emptyset\}$,
(2) $\mu(m) \in W \cup \{\emptyset\}$,
(3) and $m = \mu(w)$ if and only if $w = \mu(m)$.

Denote the set of all matchings by $\mathcal{M}$. The notation $\mu(a) = \emptyset$ has the interpretation that $a \in M \cup W$ is unmatched in μ, while $w = \mu(m)$ denotes that m and w are matched in μ.

A *preference relation* is a linear, transitive and antisymmetric binary relation. A preference relation for a man $m \in M$, denoted $P(m)$ is understood to be over the set $W \cup \{\emptyset\}$. Similarly, $P(w)$, for $w \in W$, denotes a preference relation over $M \cup \{\emptyset\}$. A *preference profile* is a list P of preference relations for men and women, i.e.

$$P = \left((P(m))_{m \in M}, (P(w))_{w \in W} \right).$$

Note that no man or woman is indifferent over two different partners; preferences with this property are normally called *strict*.

Denote by $R(m)$ the weak version of $P(m)$. So $w' \, R(m) \, w$ if $w' = w$ or $w' \, P(m) \, w$. The definition of $R(w)$ is analogous.

Fix a preference profile P. Say that a matching μ is *individually rational* if, for any m and w, $\mu(m) \, R(m) \, \emptyset$ and $\mu(w) \, R(w) \, \emptyset$. Say that a pair (w,m) *blocks* μ if $w \neq \mu(m)$, $w \, P(m) \, \mu(m)$ and $m \, P(w) \, \mu(w)$. A matching is *stable* if it is individually rational and there is no pair that blocks it. Denote by $S(P)$ the set of all stable matchings.

This model was first studied in Gale and Shapley (1962); see Roth and Sotomayor (1990) for an exposition of the theory. It should be clear that one can adapt the definition of the core as a solution for this model, and that the set of stable matchings coincides with the core.

2.3. **Statement of the problem.** Let $\mathcal{H} = \{\mu_1, \dots \mu_K\} \subseteq \mathcal{M}$ be a set of matchings. The problem I study is: *When is there a preference profile P such that $\mathcal{H} \subseteq S(P)$.* I shall say that $\mathcal{H}$ can be *rationalized* when this is the case, and that P *rationalizes* $\mathcal{H}$.

Note that I assume the same sets of agents are involved in each of the matchings in $\mathcal{H}$. See Section 8 on the consequences of relaxing this assumption.

Assume that M and W have the same number of elements, and that $\mu(m) \neq \emptyset$ and $\mu(w) \neq \emptyset$, for all m and w, and for all $\mu \in \mathcal{H}$. These assumptions are without loss of generality for the purpose of studying rationalizability. The reason is that, if $\mathcal{H}$ is rationalizable, then the single agents must be the same for all the matchings in $\mathcal{H}$ (see Roth and Sotomayor (1990)) and we can therefore ignore them and assume that the number of men and women is the same.

I start with two very simple motivating results. The first (Proposition 1) is that not all matchings can be rationalized, so there is potential for refuting matching theory. The second (Proposition 2) says that the source of refutability is quite specific: that some agents match with the same partner in different matchings.

Proposition 1. *If $|M| \geq 3$, then $\mathcal{M}$ is not rationalizable.*

Proof. Suppose, by way of contradiction, that there is P with $\mathcal{M} \subseteq S(P)$. Let $\mu_M = \bigvee S(P)$ and $\mu_W = \bigwedge S(P)$. Since $|M| = |W| \geq 3$, there is a pair (m, w) such that $m \neq \mu_M(w)$ and $w \neq \mu_W(m)$.

Let $\mu' \in \mathcal{M}$ be such that $\mu'(m) = \mu_W(m)$ and $\mu'(w) = \mu_M(w)$. There is a matching μ'' such that $\mu''(m) = w$. Since $\mathcal{M} \subseteq S(P)$, and $\mu''(m) \neq \mu_W(m)$, $w = \mu''(m)\ P(m)\ \mu_W(m)$. Similarly, $m\ P(w)\ \mu_M(w)$. Then (m, w) blocks μ'. So $\mu \notin S(P)$ and $\mathcal{M} \nsubseteq S(P)$. □

Proposition 2. *If, for all m, $\mu_i(m) \neq \mu_j(m)$ for all $\mu_i, \mu_j \in \mathcal{H}$, then $\mathcal{H}$ is rationalizable.*

Proof. For each m, define $P(m)$ by $w'\ P(m)\ w$ if and only if there is $\mu_i, \mu_j \in \mathcal{H}$ with $\mu_i(m) = w', \mu_j(m) = w$ and $i < j$. And $\emptyset\ P(m)\ w$ if $w \neq \mu(m)$, for all $\mu \in \mathcal{H}$.

For each w, define $P(w)$ by $m'P(w)m$ if and only if there is $\mu_i, \mu_j \in \mathcal{H}$ with $\mu_i(w) = m', \mu_j(w) = m$ and $i > j$. And $\emptyset\ P(w)\ m$ if $m \neq \mu(w)$, for all $\mu \in \mathcal{H}$.

Let P be the resulting preference profile. It is clear that all matchings in $\mathcal{H}$ are individually rational under P. In addition, for any (m, w) and $\mu \in \mathcal{H}$ with $m \neq \mu(w)$, $w\ P(m)\ \mu(m)$ implies that $\mu(w)\ P(w)\ m$. So there can be no blocking pair of μ. So $\mathcal{H} \subseteq S(P)$. □

The following example shows that the constructed preferences in the proof of Proposition 2 do not imply $\mathcal{H} = S(P)$.

Example 3. *Let $M = \{m_1, m_2, m_3, m_4\}$ and $W = \{w_1, w_2, w_3, w_4\}$. Consider the matchings μ_1 and μ_2 defined as:*

	m_1	m_2	m_3	m_4
μ_1	w_1	w_2	w_3	w_4
μ_2	w_2	w_1	w_4	w_3 .

Then the matching that matches m_1 and m_2 as in μ_1, and m_3 and m_4 as in μ_2, is also stable for the preferences constructed in the proof of Proposition 2.

Propositions 1 and 2 place very rough bounds on what can be rationalized, in the rest of the paper I build up a characterization of the sets of matching that can be rationalized.

3. An Illustration.

Here I present a simple example that illustrates the ideas behind the results in the paper. Consider the following example, with three men, three women and three matchings.

	m_1	m_2	m_3	m_4
μ_1	w_1	w_2	w_3	w_4
μ_2	w_1	w_3	w_4	w_2
μ_3	w_2	w_3	w_1	w_4

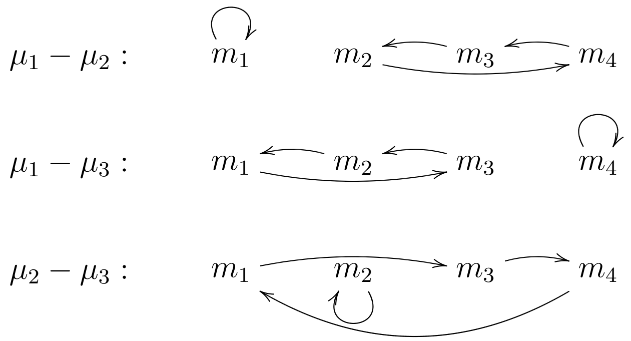
Let us construct preferences that would rationalize $\mathcal{H} = \{\mu_1, \mu_2, \mu_3\}$. We can consider all women that a man is never matched to as unacceptable. For example, set $\emptyset \; P(m_1) \; w_3$. To do this can only help in rationalizing $\mathcal{H}$. The real issue is how to specify preferences among the mens' partners in μ_1, μ_2 and μ_3.

Start with how men could rank their partners in μ_1 and μ_2. For m_1, the rank is trivial because $\mu_1(m_1) = \mu_2(m_1)$. Consider m_2. Let us say (arbitrarily) that $w_3 = \mu_2(m_2) \; P(m_2) \; \mu_1(m_2) = w_2$. Next, consider m_3. Could we have that $\mu_1(m_3) \; P(m_3) \; \mu_2(m_3)$? No, because it would imply that μ_1 and μ_2 cannot both be stable: (m_3, w_3) blocks μ_2 if $m_3 \; P(w_3) \; m_2$, and (m_2, w_3) blocks μ_1 if $m_2 \; P(w_3) \; m_3$. Hence, to set $\mu_1(m_3) \; P(m_3) \; \mu_2(m_3)$ presents a problem, regardless of what we assume about $P(w_3)$. So, if we are to rationalize $\mathcal{H}$, we have that $\mu_2(m_2) \; P(m_2) \; \mu_1(m_2)$ implies $\mu_2(m_3) \; P(m_3) \; \mu_1(m_3)$.

Suppose then that $\mu_2(m_2)\, P(m_2)\, \mu_1(m_2)$ and $\mu_2(m_3)\, P(m_3)\, \mu_1(m_3)$. Now $\mu_2(m_3) = \mu_1(m_4)$, so m_3 and m_4 are in the same situation as m_2 and m_3. Hence $\mu_2(m_3)P(m_3)\mu_1(m_3)$ implies that $\mu_2(m_4)P(m_4)\mu_1(m_4)$, by the same argument as in the previous paragraph. So the men m_2, m_3 and m_4 must agree on how they compare their partners in μ_1 and μ_2.

What if we had started with $\mu_1(m_2)P(m_2)\mu_2(m_2)$? Then m_4 and m_2 are in the same situation as m_2 and m_3 under the previous assumption: $\mu_2(m_4) = \mu_1(m_2)$. By the same argument, $\mu_1(m_2)P(m_2)\mu_2(m_2)$ implies $\mu_1(m_2)\, P(m_2)\, \mu_2(m_2)$. Repeating the argument, we obtain that m_2, m_3 and m_4 must agree on how they compare their partners in μ_1 and μ_2.

The general result is: For any two matchings, μ_i and μ_j, all the men (m, m') who stand in the relation "m's partner in μ_i is m''s partner in μ_j" must agree on how they rank their partners in μ_i and μ_j. The following diagram presents a graph among the men for each pair of matchings in $\mathcal{H}$. In the first graph, there is a directed edge $m_2 \to m_4$ because $\mu_1(m_2) = \mu_2(m_4)$; there is an edge $m_3 \to m_2$ because $\mu_1(m_3) = \mu_2(m_2)$, and so on.



The graph corresponding to $\mu_1 - \mu_2$ has two connected components, $\{m_1\}$ and $C = \{m_2, m_3, m_4\}$. By our previous argument, all the men in C must agree on how they rank their partners in μ_1 and μ_2. Similarly, reading the corresponding connected components from the diagram, all the men in $C' = \{m_1, m_2, m_3\}$ must agree on μ_1 and μ_3. And all the men in $C'' = \{m_1, m_3, m_4\}$ must agree on μ_2 and μ_3.

It is clear how these arguments restrict the possible preference profiles that might rationalize $\mathcal{H}$, but it does not by itself give a criterion for deciding that $\mathcal{H}$ is not rationalizable. The criterion arises from the presence of men who have the same partner in different matchings.

Say that $\mu_2(m)\ P(m)\ \mu_1(m)$ for all $m \in C$. Since $m_2 \in C$, and $\mu_2(m_2) = \mu_3(m_2)$, we must have that $\mu_3(m_2)\, P(m_2)\, \mu_1(m_2)$. But $m_2 \in C'$ so $\mu_3(m)\ P(m)\ \mu_1(m)$ for all $m \in C'$. Similarly, $m_4 \in C$ with $\mu_1(m_4) = \mu_3(m_4)$. So $\mu_2(m_4)P(m)\mu_1(m_2)$ now implies that $\mu_2(m)P(m)$ $\mu_3(m)$ for all $m \in C''$. The problem is that $m_1 \in C' \cap C''$, so we would need that $\mu_2(m_1)\ P(m_1)\ \mu_3(m_1)\ P(m_1)\ \mu_1(m_1)$. This is a violation of the antisymmetry of $P(m_1)$, as $\mu_2(m_1) = \mu_1(m_1)$. Hence $\mathcal{H}$ is not rationalizable.

The idea—which is formalized below—is that the presence of men with the same partner in different matchings gives a relation between objects such as C, C' and C''. These relations must satisfy a consistency condition for $\mathcal{H}$ to be rationalizable.

4. Preferences over Partners in Pairs of Matchings

The discussion in Section 3 suggests that two objects are important in studying rationalizability. The first is the set of connected components obtained from pairs of matchings in $\mathcal{H}$, which I denote by $\mathbf{C}$ below. Second are the relations between connected components in $\mathbf{C}$ derived from having agents with the same partners in two different matchings. In this section I describe the connected components, and show how these capture the essence of stability.

Fix a pair of matchings μ_i and μ_j in $\mathcal{H}$. Consider the (directed) graph for which M is the vertex-set, and $E(\mu_i, \mu_j)$ is the set of edges, defined by: $(m, m') \in E(\mu_i, \mu_j)$ if and only if $\mu_i(m) = \mu_j(m')$. Denote by $\mathbf{C}(\mu_i, \mu_j)$ the set of all connected components of $(M, E(\mu_i, \mu_j))$.

There is an analogous graph with the women as vertexes: Let $(W, F(\mu_i, \mu_j))$ be the graph for which the vertex-set is the set of women, and where $(w, w') \in F(\mu_i, \mu_j)$ if $\mu_j(w) = \mu_i(w)$. A first result relates the women's graph and the men's graph.

Lemma 4. *The following statements are equivalent:*

(1) *C is a connected component of $(M, E(\mu_i, \mu_j))$*
(2) *$\mu_i(C)$ is a connected component of $(W, F(\mu_i, \mu_j))$*

In addition, if C is a connected component of $(M, E(\mu_i, \mu_j))$, then $\mu_j(C) = \mu_i(C)$.

Proof. I first prove that $(m, m') \in E(\mu_i, \mu_j)$ if and only if $(\mu_i(m), \mu_i(m')) \in F(\mu_i, \mu_j)$, which establishes the equivalence of (1) and (2) in the lemma. First, $\mu_i(m) = \mu_j(m)$ if and only if $\mu_j(\mu_i(m)) = \mu_j(\mu_j(m')) = m'$, as μ_j is one-to-one. Hence, $(m, m') \in E(\mu_i, \mu_j)$ if and only if $\mu_j(\mu_i(m)) = m'$. Second, $(\mu_i(m), \mu_i(m')) \in F(\mu_i, \mu_j)$ if and only if $\mu_j(\mu_i(m)) = \mu_i(\mu_i(m'))$. But $m' = \mu_i(\mu_i(m'))$, so $(\mu_i(m), \mu_i(m')) \in F(\mu_i, \mu_j)$ if and only if $m' = \mu_j(\mu_i(m))$.

To prove the second statement in the lemma, note that $w \in \mu_i(C)$ if there is $m \in C$ with $w = \mu_i(m)$. Since $m \in C$ there is $m' \in C$ with $(m, m') \in E(\mu_i, \mu_j)$. Then $w = \mu_j(m')$ and therefore $w \in \mu_j(C)$. Similarly, if $w \in \mu_j(C)$ then $w \in \mu_i(C)$. □

Lemma 5. *Let $\mathcal{H}$ be rationalized by preference profile P. If $\mu_i, \mu_j \in \mathcal{H}$, and $C \in \mathbf{C}(\mu_i, \mu_j)$, then either* (1) *or* (2) *hold:*

$$\begin{aligned} &\mu_i(m)\; P(m)\; \mu_j(m) \textit{ for all } m \in C \\ &\textit{and } \mu_j(w)\; P(w)\; \mu_i(w) \textit{ for all } w \in \mu_i(C); \end{aligned} \tag{1}$$

$$\begin{aligned} &\mu_j(m)\; P(m)\; \mu_i(m) \textit{ for all } m \in C \\ &\textit{and } \mu_i(w)\; P(w)\; \mu_j(w) \textit{ for all } w \in \mu_i(C). \end{aligned} \tag{2}$$

Further, if P is a preference profile such that: for all $\mu_i, \mu_j \in \mathcal{H}$, and $C \in \mathbf{C}(\mu_i, \mu_j)$, either (1) *or* (2) *hold, and in addition*

$$\emptyset\; P(m)\; w \textit{ if and only if } w \notin \{\mu(m) : \mu \in \mathcal{H}\}$$
$$\emptyset\; P(w)\; m \textit{ if and only if } m \notin \{\mu(w) : \mu \in \mathcal{H}\},$$

then P rationalizes $\mathcal{H}$.

Remark. In part, Lemma 5 is a refinement of the classical result on opposition and coincidence of interest in matching markets. The classical result says that the agents on the same side of the market agree, and agents on opposite sides disagree, on their preferences among certain pairs of matchings. The first part of Lemma 5 says that the coincidence/opposition holds for *any* pair of matchings, but it holds within the connected components of the corresponding graph.

The second part of the lemma says that this opposition and coincidence is all that stability requires—up to the ability to construct well-defined preferences with the opposition and coincidence property. As I show in the rest of the paper, to construct such preferences is not trivial.

Lemma 5 is a variation on well-known results. For example, Knuth (1976) contains a weaker statement in his Theorem 3 of Chapter 2, and then a stronger statement in his Corollary 1. But the idea of studying the components of the graphs $(M, E(\mu_i, \mu_j))$ is new, and, as we shall see, crucial to studying refutability. It is clear from the second part of the lemma that the existing results on conflict/coincidence are either too weak or too strong as a characterization of stability.

Proof. I prove first the first statement. If C is a singleton there is nothing to prove. Assume, then that C has two or more elements. Note that C is a cycle, $C = \left\{m^1, \dots m^L\right\}$, with $(m^l, m^{l+1}) \in E(\mu_i, \mu_j)$ (modulo L) for $l = 1, \dots L$. This is because, for each $m \in M$ there is a unique $m' \in C$ with $(m', m) \in E(\mu_i, \mu_j)$ and a unique $m'' \in C$ with $(m, m'') \in E(\mu_i, \mu_j)$.

Now, say that $\mu_i(m^l)\; P(m^l)\; \mu_j(m^l)$ for some l. I shall prove that $\mu_i(m)\; P(m)\; \mu_j(m)$ for all $m \in C$. We must have $\mu_i(m^{l+1})\; P(m^{l+1})$

$\mu_j(m^{l+1})$ because $\mu_j(m^{l+1})\ P(m^{l+1})\ \mu_i(m^{l+1})$ would imply that μ_i and μ_j are not both stable: $(m^l, m^{l+1}) \in E(\mu_i, \mu_j)$, so $\mu_i(m^l) = \mu_j(m^{l+1})$; thus $(m^l, \mu_i(m^l))$ blocks μ_j if $m^l\ P(\mu_i(m^l))\ m^{l+1}$ and $(m^{l+1}, \mu_i(m^l))$ blocks μ_i if $m^{l+1}\ P(\mu_i(m^l))\ m^l$. The result that $\mu_i(m)\ P(m)\ \mu_j(m)$ for all $m \in C$ follows by induction.

Let $w \in \mu_i(C)$. We must have that $\mu_i(w) \neq \mu_j(w)$ or the component of $(W, F(\mu_i, \mu_j))$ that w lives in would be a singleton and would not coincide with $\mu_i(C)$ (Lemma 4). Now I show that $\mu_j(w)\, P(w)\, \mu_i(w)$: if we instead have $\mu_i(w)\ P(w)\ \mu_j(w)$, then $(\mu_i(w), w)$ would block μ_j, as $\mu_i(w) \in C$ and thus $w\ P(\mu_i(w))\ \mu_j(\mu_i(w))$.

So we have established that $\mu_i(m^l)\ P(m^l)\ \mu_j(m^l)$ for some l implies statement (1) of the lemma. The argument hat $\mu_j(m^l)\ P(m^l)\ \mu_j(m^l)$ for some l implies statement (2) is analogous.

I now prove the second part of the lemma. Let $\mu \in \mathcal{H}$. It is clear that μ is individually rational by the requirement on P. Let w and m be such that $w\ P(m)\ \mu(m)$. Let i and j be such that $w = \mu_i(m)$ and $\mu = \mu_j$. There must exists such a i because $\emptyset\ P(m)\ w$ if w is not m's partner in some matching in $\mathcal{H}$. Let $C \in \mathbf{C}(\mu_i, \mu_j)$ with $m \in C$. Then $w \in \mu_i(C)$ and, by statement (1) of the lemma, $\mu_j(w)\, P(w)\, \mu_i(w0 = m$. Hence (m, w) is not a blocking pair. Since (m, w) was arbitrary, μ is stable. □

5. Relations Between Components, and a Necessary Condition for Rationalization

The discussion in Section 3 suggests that there are relations between components of the pairwise graphs, relations that come from the presence of some agents who are with the same partner in two (or more) matchings. The discussion also suggests that the rationalizability of $\mathcal{H}$ depends on the restrictions imposed by these relations. Here I define the relations and show how they give a simple necessary condition for $\mathcal{H}$ to be rationalizable.

Let $\mathbf{C}$ be the set of all elements of $\mathbf{C}(\mu_i, \mu_j)$, for any two distinct $\mu_i, \mu_j \in \mathcal{H}$ with $i < j$. That is,

$$\mathbf{C} = \cup \{C \subseteq M : |C| \geq 2 \text{ and } \exists(\mu_i, \mu_j) \text{ s.t. } i < j \text{ and } C \in \mathbf{C}(\mu_i, \mu_j)\}.$$

Note that a set may be a connected component of more than one graph $(M, E(\mu_i, \mu_j))$. If a set C is in $\mathbf{C}(\mu_i, \mu_j)$ and in $\mathbf{C}(\mu_h, \mu_k)$ I abuse notation and regard each "copy" of C as a different element of $\mathbf{C}$. As a result, for each $C \in C$ there is a unique pair (μ_i, μ_j) such that $C \in \mathbf{C}(\mu_i, \mu_j)$. This abuse does not, I believe, confuse, and it makes the notation lighter.

I define two binary relations on the elements of $\mathbf{C}$, and denote them by $\triangle$ and $\bigtriangledown$.

Definition ($\triangle$)**.** Let $C, C' \in \mathbf{C}$. Say that $C \triangle C'$ if there are three distinct numbers, i,j, and k, in $\{1, 2, \dots K\}$, such that

- either $C \in \mathbf{C}(\mu_i, \mu_j)$, $C' \in \mathbf{C}(\mu_i, \mu_k)$
  or $C \in \mathbf{C}(\mu_j, \mu_i)$, $C' \in \mathbf{C}(\mu_k, \mu_i)$, and
- there is $m \in C \cap C'$ with $\mu_j(m) = \mu_k(m)$.

Definition ($\bigtriangledown$)**.** Let $C, C' \in \mathbf{C}$. Say that $C \bigtriangledown C'$ if there are three distinct numbers, i,j, and k, in $\{1, 2, \dots K\}$, such that

- either $C \in \mathbf{C}(\mu_i, \mu_j)$, $C' \in \mathbf{C}(\mu_k, \mu_i)$
  or $C \in \mathbf{C}(\mu_j, \mu_i)$, $C' \in \mathbf{C}(\mu_i, \mu_k)$, and
- there is $m \in C \cap C'$ with $\mu_j(m) = \mu_k(m)$.

Let $\mathbf{E}_\triangle$ be the set of pairs (C, C') with $C \triangle C'$ and $\mathbf{E}_\bigtriangledown$ be the set of pairs (C, C') with $C \bigtriangledown C'$. So $\mathbf{E}_\triangle$ is another way of writing the binary relation $\triangle$ and $\mathbf{E}_\bigtriangledown$ is just the binary relation $\bigtriangledown$. This duplicate notation is useful.

Now, $(\mathbf{C}, \mathbf{E}_\triangle \cup \mathbf{E}_\bigtriangledown)$ represents the (loop-free and undirected) graph with vertex-set $\mathbf{C}$ and where there is an edge between C and C' if either $C \triangle C'$ or $C \bigtriangledown C'$.

Theorem 6. *If $\mathcal{H}$ is rationalizable then $(\mathbf{C}, \mathbf{E}_\triangle \cup \mathbf{E}_\bigtriangledown)$ can have no cycle with an odd number of $\bigtriangledown$.*

Theorem 6 follows from Theorem 7 below.

In a sense, the necessary condition in Theorem 6 is the content of the theory of stable matchings. I will show in the rest of this section, and in Section 6, that as long as the necessary condition in Theorem 6 is compatible with a specification of well-behaved preferences, then $\mathcal{H}$ can be rationalized.

A first requirement of the compatibility with well-behaved preferences is that $\mathbf{C}$, $\mathbf{E}_\triangle$ and $\mathbf{E}_\bigtriangledown$ cannot imply intransitiveness. I express this requirement by making $\triangle$ a larger relation: I define a monotone increasing sequence $\left\{\mathbf{E}^k_\triangle\right\}$, and work with the larger binary relation $\mathbf{D}_\triangle = \cup_{k=1}^\infty \mathbf{E}^k_\triangle$. Let $\mathbf{E}^0_\triangle = \mathbf{E}_\triangle$. Given $\mathbf{E}^k_\triangle$, for $k \geq 0$, let $\mathbf{E}^{k+1}_\triangle$ be those edges (C, C') between elements in $\mathbf{C}$ such that either $(C, C') \in \mathbf{E}^k_\triangle$ and/or there is i, j, h and $\tilde{C} \in \mathbf{C}$ with $C \cap \tilde{C} \cap C' \neq \emptyset$ such that $C \in \mathbf{C}(\mu_i, \mu_j)$ and either 1 or 2 hold:

(1) $i < j < h$, $\tilde{C} \in \mathbf{C}(\mu_j, \mu_h)$, $C' \in \mathbf{C}(\mu_i, \mu_h)$, and C and $\tilde{C}$ are connected in $\left(\mathbf{C}, \mathbf{E}^{k-1}_\triangle\right)$

(2) $i < h < j$, $\tilde{C} \in \mathbf{C}(\mu_h, \mu_j)$, $C' \in \mathbf{C}(\mu_i, \mu_h)$, and there is a path in $\left(\mathbf{C}, \mathbf{E}^{k-1}_{\triangle} \cup \mathbf{E}_{\bigtriangledown}\right)$ between C and $\tilde{C}$ with an odd number of $\bigtriangledown$s.

Let $\mathbf{D}_{\triangle} = \cup_{k=1}^{\infty} \mathbf{E}^k_{\triangle}$. Note that $\mathbf{D}_{\triangle} = \mathbf{E}^L_{\triangle}$, for some $L \geq 1$, as the sequence of $\mathbf{E}^k_{\triangle}$ is monotone increasing and $\mathbf{C}$ is finite.

Theorem 7. *If $\mathcal{H}$ is rationalizable then $(\mathbf{C}, \mathbf{D}_{\triangle} \cup \mathbf{E}_{\bigtriangledown})$ can have no cycle with an odd number of $\bigtriangledown$.*

The proof of Theorem 7 requires Lemmas 5 and 8.

Let $\mathcal{H}$ be rationalizable. Define the function $d : \mathbf{C} \to \{-1, 1\}$ as follows. For each $C \in \mathbf{C}$, let i, j be such that $C \in \mathbf{C}(\mu_i, \mu_j)$. Say that $d(C) = 1$ if $(\forall m \in C)(\mu_i \, P(m) \, \mu_j)$ and -1 otherwise. Note that Lemma 5 says that all $m \in C$ must agree on their preferences over $\mu_i(m)$ and $\mu_j(m)$.

Lemma 8. *Let $\mathcal{H}$ be rationalizable and $(C_1, \ldots C_N)$ be a cycle in $(\mathbf{C}, \mathbf{E}_{\triangle} \cup \mathbf{E}_{\bigtriangledown})$. For each n and L, mod N,*

$$d(C_n) = \Pi_{l=n}^{L} (-1)^{\mathbf{1}_{\{C_l \bigtriangledown C_{l+1}\}}} d(C_L) \tag{3}$$

Proof. Let P rationalize $\mathcal{H}$. I only prove the case $L = n + 1$; the result then follows by induction. Let $C_n \triangle C_{n+1}$. There are i, j and k such that (say) $C_n \in \mathbf{C}(\mu_i, \mu_j)$ and $C_{n+1} \in \mathbf{C}(\mu_i, \mu_k)$. There is $m^* \in C_n \cap C_{n+1}$ with $\mu_j(m^*) = \mu_k(m^*)$, so $\mu_i(m^*) \, P \, (m^*) \mu_j(m^*)$ if and only if $\mu_i(m^*) \, P \, (m^*) \mu_k(m^*)$. Since $m^* \in C_n \cap C_{n+1}$, Lemma 5 implies

$$(\forall m \in C_n) \left(\mu_i(m) \, P(m) \, \mu_j(m)\right) \text{ iff } (\forall m \in C_{n+1}) \left(\mu_i(m) \, P(m) \, \mu_k(m)\right).$$

Hence $d(C_n) = d(C_{n+1})$. Similarly when $C_n \in \mathbf{C}(\mu_j, \mu_i)$ and $C_{n+1} \in \mathbf{C}(\mu_k, \mu_i)$.

On the other hand, when $C_n \bigtriangledown C_{n+1}$ and i, j and k are such that $C_n \in \mathbf{C}(\mu_i, \mu_j)$ and $C_{n+1} \in \mathbf{C}(\mu_k, \mu_i)$, the existence of $m^* \in C_n \cap C_{n+1}$ with $\mu_j(m^*) = \mu_k(m^*)$ implies (Lemma 5) that $d(C_n) = 1$ if and only if $d(C_{n+1}) = -1$. □

Proof of Theorem 7. Let $\mathcal{H}$ be rationalizable by preference profile P. First note that Lemma 8 implies Theorem 6 because any cycle $C_1, \ldots C_N$ with an odd number of $\bigtriangledown$s implies that $d(C_1) = (-1) d(C_1)$.

We prove Theorem 7 by induction. In the previous paragraph we proved that $(\mathbf{C}, \mathbf{E}_{\triangle} \cup \mathbf{E}_{\bigtriangledown}) = \left(\mathbf{C}, \mathbf{E}^0_{\triangle} \cup \mathbf{E}_{\bigtriangledown}\right)$ can have no cycle with an odd number of $\bigtriangledown$, and Lemma 8 implies that the formula (3) holds in $\left(\mathbf{C}, \mathbf{E}^0_{\triangle} \cup \mathbf{E}_{\bigtriangledown}\right)$. Suppose this statement is true of $\left(\mathbf{C}, \mathbf{E}^k_{\triangle} \cup \mathbf{E}_{\bigtriangledown}\right)$; if we prove that it is true of $\left(\mathbf{C}, \mathbf{E}^{k+1}_{\triangle} \cup \mathbf{E}_{\bigtriangledown}\right)$ then the proof of the theorem is done.

Let $(C, C') \in \mathbf{E}_{\triangle}^{k+1} \backslash \mathbf{E}_{\triangle}^{k}$. I shall prove that $d(C) = d(C')$. Let i, j, h and $\tilde{C} \in \mathbf{C}$ with $C \cap \tilde{C} \cap C' \neq \emptyset$ such that $C \in \mathbf{C}(\mu_i, \mu_j)$ is in the situation described by Item 1 or Item 2. Suppose that they are in the situation described by Item 1. Since C and $\tilde{C}$ are connected in $\big(\mathbf{C}, \mathbf{E}_{\triangle}^{k-1}\big)$, by Lemma 8, we have $d(C) = d(\tilde{C})$. Suppose, without loss of generality, that $d(C) = 1$. Let $m \in C \cap C' \cap \tilde{C}$; then $d(C) = d(\tilde{C}) = 1$ implies $\mu_i(m) P(m) \mu_j(m)$ and $\mu_j(m) P(m) \mu_h(m)$. So $\mu_i(m) P(m) \mu_h(m)$ and we must have $d(C') = d(C)$ Suppose now the situation described by Item 2. The existence of a path with an odd number of $\bigtriangledown$s connecting C and $\tilde{C}$ implies that $d(C) \neq d(\tilde{C})$. Suppose, without loss of generality, that $d(C) = 1$. Let $m \in C \cap C' \cap \tilde{C}$; then $1 = d(C) \neq d(\tilde{C})$ implies $\mu_i(m)\, P(m)\, \mu_j(m)$ and $\mu_j(m)\, P(m)\, \mu_h(m)$. So $\mu_i(m)\, P(m)\, \mu_h(m)$ and we must have $d(C') = d(C)$

Now, since $d(C') = d(C)$ for all $(C, C') \in \mathbf{E}_{\triangle}^{k+1} \backslash \mathbf{E}_{\triangle}^{k}$, and holds in $\big(\mathbf{C}, \mathbf{E}_{\triangle}^{k} \cup \mathbf{E}_{\bigtriangledown}\big)$, (3) holds in $\big(\mathbf{C}, \mathbf{E}_{\triangle}^{k+1} \cup \mathbf{E}_{\bigtriangledown}\big)$. Then $\big(\mathbf{C}, \mathbf{E}_{\triangle}^{k+1} \cup \mathbf{E}_{\bigtriangledown}\big)$ has no cycles with an odd number of $\bigtriangledown$s. □

6. A Necessary and Sufficient Condition for Rationalization

The graph $(\mathbf{C}, \mathbf{D}_{\triangle} \cup \mathbf{E}_{\bigtriangledown})$ captures some of the requirements put by transitivity of preferences, but not all. In this section I express the remaining requirements as a system of polynomial inequalities. The idea is that $C \in \mathbf{C}(\mu_i, \mu_j)$ be assigned a value of 1 if all $m \in C$ prefer μ_i over μ_j and value -1 if they prefer μ_j. It is then simple to control the transitivity of preferences by controlling the values one can assign to the different C. The result is a characterization of the $\mathcal{H}$ that can be rationalized.

The characterization poses the question of when the rationalizing P is unique; in econometrics such a situation is called *(exact) identification.* It is easy to show (Proposition 10) that, when $\mathcal{H}$ is rationalizable, the rationalizing P will generally not be unique.

One first step in the characterization is that all C and C' that are connected in $(\mathbf{C}, \mathbf{D}_{\triangle})$ must have the same value, so we can treat them as the same object. Let $\mathbb{C}$ be the set of all connected components of $(\mathbf{C}, \mathbf{D}_{\triangle})$. Let $(\mathbb{C}, \mathbb{D})$ be the graph that has $\mathbb{C}$ as vertex-set, and where $(\mathcal{C}, \mathcal{C}') \in \mathbb{D}$ if there is $C \in \mathcal{C}$ and $C' \in \mathcal{C}'$ with $C \bigtriangledown C'$.

If $(\mathbf{C}, \mathbf{D}_{\triangle} \cup \mathbf{E}_{\bigtriangledown})$ has no cycle with an odd number of $\bigtriangledown$s, $(\mathbb{C}, \mathbb{D})$ is a well-defined loop-free graph: For any two C and C' in the same component $\mathcal{C} \in \mathbb{C}$ it cannot be that $C \bigtriangledown C'$, as there is a path from C to C' in $(\mathbf{C}, \mathbf{D}_{\triangle})$ and $C \bigtriangledown C'$ would imply a cycle with exactly one $\bigtriangledown$.

Let B be a ternary relation on $\mathbb{C}$ defined as follows: $(\mathcal{C}, \mathcal{C}', \mathcal{C}'') \in B$ if there is i, j, and h, $i < j < h$, and $C \in \mathcal{C} \cap \mathbf{C}(\mu_i, \mu_j)$ $C' \in \mathcal{C}' \cap \mathbf{C}(\mu_j, \mu_h)$ and $C'' \in \mathcal{C}'' \cap \mathbf{C}(\mu_i, \mu_h)$ with $C \cap C' \cap C'' \neq \emptyset$.

Theorem 9. *$\mathcal{H}$ is rationalizable if and only if $(\mathbf{C}, \mathbf{D}_{\triangle} \cup \mathbf{E}_{\bigtriangledown})$ has no cycle with an odd number of $\bigtriangledown$s, and for the resulting graph $(\mathbb{C}, \mathbb{D})$, there is a function $d : \mathbb{C} \to \{-1, 1\}$ that satisfies:*

(1) $\mathcal{C} \bigtriangledown \mathcal{C}' \Rightarrow d(\mathcal{C}) + d(\mathcal{C}') = 0$,
(2) $(\mathcal{C}, \mathcal{C}', \mathcal{C}'') \in B \Rightarrow (d(\mathcal{C}) + d(\mathcal{C}'))\, d(\mathcal{C}'') \geq 0$.

Further, there is a rationalizing preference profile for each function d satisfying (1) *and* (2)*.*

Proof of Theorem 9. I only prove the "if" statement; "only if" is straightforward given the results in the previous section. Let $(\mathbf{C}, \mathbf{D}_{\triangle} \cup \mathbf{E}_{\bigtriangledown})$ have no cycle with an odd number of $\bigtriangledown$s, and d be a function in the conditions of the theorem. Abusing notation, interpret d as defined on $\mathbf{C}$ by letting $d(\mathcal{C}) = d(C)$ for all $C \in \mathcal{C}$. Note that, for all C there is some $\mathcal{C} \ni C$.

For each $m \in M$, construct preferences $P(m)$ by setting $\emptyset\; P(m)\; w$ for all $w \notin \{\mu(m) : \mu \in \mathcal{H}\}$, $w\, P(m)\, \emptyset$ for all $w \in \{\mu(m) : \mu \in \mathcal{H}\}$, and $\mu_i(m)\; P(m)\; \mu_j(m)$ if either $i < j$ and $d(C) = 1$ for $C \in \mathbf{C}(\mu_i, \mu_j)$ with $C \ni m$, or $j < i$ and $d(C) = -1$ for $C \in \mathbf{C}(\mu_j, \mu_i)$ with $C \ni m$.

For each $w \in W$, define $P(w)$ by $\emptyset P(w) m$ for all $m \notin \{\mu(w) : \mu \in \mathcal{H}\}$, $m\, P(w)\, \emptyset$ for all $m \in \{\mu(w) : \mu \in \mathcal{H}\}$, and $\mu_i(w)\; P(m)\; \mu_j(w)$ if either $i < j$ and $d(\mu_i(C)) = -1$ for $\mu_i(C) \in \mathbf{C}(\mu_i, \mu_j)$ with $\mu_i(C) \ni \mu_i(w)$ or $j < i$ and $d(\mu_i(C)) = 1$ for $\mu_i(C) \in \mathbf{C}(\mu_j, \mu_i)$ with $\mu_i(C) \ni \mu_i(m)$. Extend $P(m)$ and $P(w)$ arbitrarily to pairs of agents that are ranked below $\emptyset$.

Note that $P(m)$ and $P(w)$ are antisymmetric. I show that $P(m)$ is transitive. The proof that $P(w)$ is transitive is analogous. Let $\mu_i(m) P(m) \mu_j(m)$ and $\mu_j(m) P(m) \mu_h(m)$. I shall prove that $\mu_i(m) P(m)$ $\mu_h(m)$.

Case 1. Let $i < j < h$, $m \in C \in \mathbf{C}(\mu_i, \mu_j)$, $m \in C' \in \mathbf{C}(\mu_j, \mu_h)$ and $m \in C'' \in \mathbf{C}(\mu_i, \mu_h)$. Note that $\mu_i(m)\; P(m)\; \mu_j(m)$ implies $d(C) = 1$ and $\mu_j(m)\, P(m)\, \mu_h(m)$ implies $d(C') = 1$. If C and C' are connected in $(\mathbf{C}, \mathbf{D}_{\triangle})$, then, by the construction of $\mathbf{D}_{\triangle}$, C and C'' are also connected. So (3) implies that $d(C'') = d(C) = 1$; thus $\mu_i(m)\; P(m)\; \mu_h(m)$. Now let C and C' not be connected in $(\mathbf{C}, \mathbf{D}_{\triangle})$. If C and C'' are connected then there is nothing to prove, as (3) gives $d(C'') = d(C) = 1$ and $\mu_i(m)\; P(m)\; \mu_h(m)$. Similarly, we obtain $\mu_i(m)\; P(m)\; \mu_h(m)$ if C' and C'' are connected. Suppose then that C, C' and C'' are not connected in $(\mathbf{C}, \mathbf{D}_{\triangle})$. Let $\mathcal{C}, \mathcal{C}', \mathcal{C}'' \in \mathbb{C}$ be such that $C \in \mathcal{C}$, $C' \in \mathcal{C}'$, and $C'' \in \mathcal{C}''$;

$\mathcal{C}$, $\mathcal{C}'$, and $\mathcal{C}''$ are all different because C, C' and C'' are disconnected. Since $m \in \mathcal{C} \cap \mathcal{C}' \cap \mathcal{C}''$, $(\mathcal{C}, \mathcal{C}', \mathcal{C}'') \in B$. Now, $d(C) = d(C') = 1$ imply $d(\mathcal{C}) = d(\mathcal{C}') = 1$, so Item (2) of the theorem requires that $2d(C'') \geq 0$, i.e. $d(C'') = 1$. Hence $\mu_i(m)\, P(m)\, \mu_h(m)$.

The argument in Case 1 also yields that,

$$\left.\begin{array}{l} i < j < h \\ \mu_j(m)\, P(m)\, \mu_i(m) \\ \mu_h(m)\, P(m)\, \mu_j(m) \end{array}\right\} \text{ implies } \mu_h(m)\, P(m)\, \mu_i(m). \tag{4}$$

This gives us $\mu_i(m)\, P(m)\, \mu_h(m)$ in the case $h < j < i$ by applying (4) to (i', j', h') defined as $i' = h$, $j' = j$ and $h' = i$.

Case 2. Let $i < h < j$, $m \in C \in \mathbf{C}(\mu_i, \mu_j)$, $m \in C' \in \mathbf{C}(\mu_h, \mu_j)$ and $m \in C'' \in \mathbf{C}(\mu_i, \mu_h)$. So $d(C) = 1$ and $d(C') = -1$.

First, if $C \triangle C''$ we have $d(C) = d(C'')$ so there is nothing to prove. Suppose then that $C \triangle C''$ is false. It cannot be that $C' \triangle C''$, since that would imply $C' \triangle C$ by the construction of $\mathbf{D}_\triangle$, and $d(C') \neq d(C)$ implies that C' and C are disconnected in $(\mathbf{C}, \mathbf{D}_\triangle)$. So it must be the case that all of C, C' and C'' are disconnected in $(\mathbf{C}, \mathbf{D}_\triangle)$. Let $\mathcal{C}, \mathcal{C}', \mathcal{C}'' \in \mathbb{C}$ be as in Case 1. Then $(C'', C', C) \in B$. By Item (2) of the theorem, $d(C'')$ must satisfy $(d(C'') - 1) \geq 0$. So $d(C'') = 1$ and $\mu_i(m)\, P(m)\, \mu_h(m)$.

The argument in Case 2 also covers the case $h < i < j$, by a reasoning similar to the one for $h < j < i$ at the end of Case 1.

Case 3. Let $j < i < h$, $m \in C \in \mathbf{C}(\mu_j, \mu_i)$, $m \in C' \in \mathbf{C}(\mu_j, \mu_h)$ and $m \in C'' \in \mathbf{C}(\mu_i, \mu_h)$. Now we have $d(C) = -1$ and $d(C') = 1$. First, if $C' \triangle C''$, then $d(C'') = 1$ so there is nothing to prove. Second, it cannot be that $C \triangle C''$, since that would imply $C \triangle C'$ by the construction of $\mathbf{D}_\triangle$, and $d(C') \neq d(C)$ implies that C' and C are disconnected in $(\mathbf{C}, \mathbf{D}_\triangle)$. Let $\mathcal{C}, \mathcal{C}', \mathcal{C}'' \in \mathbb{C}$ be as in Case 1. Then $(C, C'', C') \in B$. By Item (2) of the theorem, $d(C')$ must satisfy $(d(C'') - 1) \geq 0$. So $d(C'') = 1$ and $\mu_i(m)\, P(m)\, \mu_h(m)$.

The argument in Case 3 also covers the case $j < h < i$ by a reasoning similar to the one in Case 1.

Finally, I show that all $\mu \in \mathcal{H}$ are stable under the constructed preferences. Let $\mu \in \mathcal{H}$. It is clear that μ is individually rational. Let w and m be such that $w\, P(m)\, \mu(m)$. Let i and j be such that $w = \mu_j(m)$ and $\mu = \mu_i$. There must exists such a j because $\emptyset\, P(m)\, w$ if w is not m's partner in some matching in $\mathcal{H}$. Without loss of generality, say that $i < j$. Let $C \in \mathbf{C}(\mu_i, \mu_j)$ with $m \in C$, so $d(C) = -1$. Then $w \in \mu_i(C)$, so the construction of $P(w)$ implies that $\mu_i(w)\, P(w)\, \mu_j(w)$. So $\mu_i(m)\, P(w)\, m$, and hence (m, w) cannot block μ. $\square$

Finally, I show that, generally, matching theory is not exactly identified; if $\mathcal{H}$ is rationalizable there are generally many different preference relations that rationalize it. The source of the different preferences is that, if m is not matched to w in any matching in $\mathcal{H}$, then the data in $\mathcal{H}$ contains very little information on m's standing in w's preference relation.

Let U_m be the set of women m is not matched to in a matching in $\mathcal{H}$. Say that two preference profiles are *essentially different* if there is at least agent on which the preference for two acceptable partners is different.

Proposition 10. *If $\mathcal{H}$ is rationalizable, then it is rationalizable by at least*

$$(2\,|M|)^{|M|}\,\Pi_{m\in M}\,|U_m|$$

essentially different preference profiles.

Proof. Let P rationalize $\mathcal{H}$. For each $w \in U_m$, I can modify P by setting $\emptyset\; P(w)\; m$ and vary $P(m)$ by placing w in any of the possible $|W|$ $(= |M|)$ places in the ranking of m's preferences. This will not change the fact that all $\mu \in \mathcal{H}$ are individually rational, and the only blocking pair it could give rise to is (m, w), but having set $\emptyset\; P(w)\; m$ guarantees that (m, w) will not be a blocking pair. The same is true if I set $\emptyset\; P(m)\; w$. So we obtain $2\,|M|$ different preference relation for each $w \in U_m$. □

7. The lattice structure of stable matchings.

Here I discuss the problem of the universe of lattices that can be stable sets of matching problems; this problem is related to the question of rationalizability. Recall the classical result in matching theory that the set of stable matchings is a non-empty distributive lattice. The problem, first stated by Knuth (1976), is to characterize the distributive lattices that can be stable matchings for some instance of the matching problem.

Blair (1984) gave what seems to be both the first and definitive answer to the problem. Blair proves that, for any distributive lattice L, there is a set of men and women, and a preference profile, so that the resulting set of stable matchings is lattice isomorphic to L. Blair's proof is constructive; Gusfield, Irving, Leather, and Saks (1987) improve on his construction by requiring a smaller set of men and women to generate any given lattice.

The interpretation of Blair's result in the literature is that the lattice structure of the set of stable matchings has no properties beyond distributivity. In the words of Roth and Sotomayor (1990):

> "We might (...) hope to say something more about what kinds of lattices arise as sets of stable matchings, in order to use any additional properties thus specified to learn more about the market. (Blair's) Theorem shows that this line of investigation will not bear any further fruit."

Gusfield and Irving (1989) make a similar observation:

> "There is no special structure that holds in general for marriage lattices ... that does not also hold for general distributive lattices."

Roth and Sotomayor's, and Gusfield and Irving's, is one interpretation of Blair's result, but it may be useful to think of the result in a different way. While only distributivity is preserved by lattice homomorphism, the lattice structure of stable matchings may still have additional properties, properties that are not shared by other lattices of matchings. In fact, one can rewrite Lemma 5 as a characterization of the lattices of stable matchings. The lemma implies that these lattices have properties in addition to distributivity.

The additional properties refer to the opposition and coincide of interest property of any *pair* of stable matchings. This opposition/coincidence property is characteristic of lattices of stable matchings, and may not be present in other lattices, even in other lattices of matchings. Concretely, Lemma 5 implies that, if μ_i and μ_j are stable, then, for any $C \in \mathbf{C}(\mu_i, \mu_j)$, either (5) or (6) must hold:

$$(\mu_i \wedge \mu_j)|_C = \mu_i|_C \text{ and } (\mu_i \vee \mu_j)|_C = \mu_j|_C \tag{5}$$

$$(\mu_i \wedge \mu_j)|_C = \mu_j|_C \text{ and } (\mu_i \vee \mu_j)|_C = \mu_i|_C. \tag{6}$$

One can endow a set of matchings with $\vee$ and $\wedge$ operations so that it is a distributive lattice, but violates (5) and (6). One example is the set of matchings in Section 3; these cannot be endowed with a lattice structure that respects (5) and (6) because any such structure would involve the matchings being totally ordered, and we have seen that a total order is incompatible with stability.

I should emphasize that one can make this point using existing results. For example, if μ_i and μ_j are stable, $\vee$ cannot be such that $\mu_i(m) \vee \mu_j(m) \notin \{\mu_i(m), \mu_j(m)\}$ (see e.g. Roth and Sotomayor (1990)). The contribution here is that, by the second part of Lemma 5, (5) and (6) are also sufficient for a set of matchings to be stable.

Finally, Lemma 5 provides an answer to one interpretation of Knuth's problem. Knuth wrote "Can one obtain all distributive lattices from suitable preference matrices ?" (I refer to preference matrices as preference profiles). If we interpret the question as: given M and W, can all distributive lattices of matchings be obtained with suitable preference profiles? The answer is negative, as exemplified by the matchings in Section 3.

8. Different agents for different matchings

I have assumed that the sets of agents involved in each of the matchings in $\mathcal{H}$ is the same. In empirical applications, it is likely that some agents who are present in one matching are not present in others. The results above can be modified to account for different sets of agents. The modification is straightforward but cumbersome, so I only outline how the basic argument extends. I need to emphasize, though, that the source of refutability comes from some agents having the same partner in different matchings. The looser are the ties across matchings, the more degrees of freedom one has in rationalizing the observations.

Say that M_i and W_i are the sets of men and women who are matched by μ_i. A straightforward modification of the arguments above gives that, if there is a path from m to m' in $(M_i \cap M_j, E(\mu_i, \mu_j))$, then $\mu_i(m)\ P(m)\ \mu_j(m)$ implies that $\mu_i(m')\ P(m')\ \mu_j(m')$. One can now partition each component in $(M_i \cap M_j, E(\mu_i, \mu_j))$ which is not a cycle into the paths that start at the different elements of the component. Saying that $\mu_i(m) P(m) \mu_j(m)$ for one m implies that all the paths that start at m, or at one of m's successors in $(M_i \cap M_j, E(\mu_i, \mu_j))$, agree on how they compare their partners in μ_i and μ_j. Adapting the definitions of $\triangle$ and $\bigtriangledown$ gives the appropriate versions of the results above.

9. Probability of rationalizing

The results on rationalizability have some implications for the statistical "power" of matching theory. Power refers here to how likely it is that purely random outcomes will look as if they were generated by the theory; i.e. how likely it is that one can rationalize random matchings.

I show that, for a fixed number of observed matchings, in a large population, the probability of rationalizing purely random matchings is bounded away from zero. The result says that large populations require large sample sizes, which is probably not surprising.

Let M_n be a set of men and W_n a set of women, each with n elements. Let $\mathcal{M}_n$ be the resulting set of possible matchings with no single agents. Endow $\mathcal{M}_n$ with the uniform distribution, and consider sets $\mathcal{H}_k$ of k matchings chosen independently at random from $\mathcal{M}_n$.

Proposition 11. *If k is fixed,*

$$\liminf_{n\to\infty} \boldsymbol{P}\{\mathcal{H}_k \textit{ is rationalizable }\} \geq e^{-k(k-1)/2}$$

Proof. Fix k and n. Consider the realizations of $\mathcal{H}_k$ such that, for all m, $\mu_i(m) \neq \mu_j(m)$ for all $\mu_i, \mu_j \in \mathcal{H}_k$, then $\mathcal{H}_k$ is rationalizable in (M_n, W_n) by Proposition 2. For each such realization of $\mathcal{H}_k$, form a $k \times n$ array (a_{st}) by setting $a_{st} = \mu_s(m_t)$. Then each woman will appear exactly once in each row, as the μ_s are matchings. And each woman will appear at most once in each column, by the assumption that for all m, $\mu_i(m) \neq \mu_j(m)$ for all $\mu_i, \mu_j \in \mathcal{H}_k$. The resulting array thus forms a *latin rectangle* (see e.g. Denes and Keedwell (1974)).

Thus there are as many realizations of $\mathcal{H}_k$ in the hypothesis of Proposition 2 as there are $k \times n$ latin rectangles. In turn, Erdös and Kaplanski (1946) proved that, as $n \to \infty$, the number of $k \times n$ is asymptotic to

$$(n!)^k e^{-\binom{k}{2}}. \tag{7}$$

On the other hand, an arbitrary realization of $\mathcal{H}_k$ forms an array where each woman appears exactly once in each row, but may be repeated in columns. So each row is a permutation of the women, and there are as many $\mathcal{H}_k$ as ways of making k permutations, that is $(n!)^k$. The probability then of a draw of $\mathcal{H}_k$ in the hypothesis of Proposition 2 is asymptotic to $e^{-\binom{k}{2}}$, which gives the result. □

As I remarked above, the message in Proposition 11 is probably not surprising, but it hopefully illustrates a potential for statistical applications of the rationalizability results developed in the paper. The proof of the proposition builds on the very crude sufficient condition for rationalizability in Proposition 2 of Section 2; there is clearly potential for refining this result.

References

Abdulkadiroglu, A., P. Pathak, A. Roth, and T. Sönmez (2005): "The Boston Public School Match," *American Economic Review*, 95(2), 368–371.

Afriat, S. N. (1967): "The Construction of Utility Functions from Expenditure Data," *International Economic Review*, 8(1), 67–77.

Blair, C. (1984): "Every Finite Distributive Lattice is a Set of Stable Matchings," *Journal of Combinatorial Theory (A)*, 37, 353–356.

Choo, E., and A. Siow (2006): "Who Marries Whom and Why," *Journal of Political Economy*, 114(1), 175–201.

Dagsvik, J. K. (2000): "Aggregation in Matching Markets," *International Economic Review*, 41(1), 27–57.

Denes, J., and A. D. Keedwell (1974): *Latin Squares and their Applications*. Academic Press.

Erdös, P., and I. Kaplanski (1946): "The Asymptotic Number of Latin Rectangles," *American Journal of Mathematics*, 68(2), 230–236.

Gale, D., and L. S. Shapley (1962): "College Admissions and the Stability of Marriage," *The American Mathematical Monthly*, 69(1), 9–15.

Gusfield, D., R. Irving, P. 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(2), 304–309.

Gusfield, D., and R. W. Irving (1989): *The Stable Marriage Problem: Structure and Algorithms*. MIT Press.

Hitsch, G., A. Hortaçsu, and D. Ariely (2006): "What Makes You Click? Mate Preferences and Matching Outcomes in Online Dating," Mimeo, University of Chicago.

Knuth, D. E. (1976): *Marriages Stable*. Université de Montréal Press, Translated as "Stable Marriage and Its Relation to Other Combinatorial Problems," CRM Proceedings and Lecture Notes, American Mathematical Society.

Roth, A., T. Sönmez, and U. Ünver (2004): "Kidney Exchange," *Quarterly Journal of Economics*, 119(2), 457–488.

Roth, A., and M. Sotomayor (1990): *Two-sided Matching: A Study in Game-Theoretic Modelling and Analysis*, vol. 18 of *Econometric Society Monographs*. Cambridge University Press, Cambridge England.

Roth, A. E., and E. Peranson (1999): "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design," *American Economic Review*, 89(4), 748–780.

Samuelson, P. A. (1947): *Foundations of Economic Analysis*. Harvard University Press.

Division of the Humanities and Social Sciences, MC 228-77, California Institute of Technology, Pasadena CA 91125, USA.

E-mail address: fede@hss.caltech.edu

URL: http://www.hss.caltech.edu/~fede/