What Matchings Can Be Stable? The Testable Implications of Matching Theory [1]

Federico Echenique [2]

October 22, 2007

[1]I thank the Area Editor, Eilon Solan, and two anonymous referees, for their detailed comments on a previous draft. I am also grateful to David Ahn, Chris Chambers, Geoffroy De Clippel, Alekos Kechris, Hideo Konishi, Jay Sethuraman, Tayfun Sönmez and seminar audiences at UC Berkeley, Boston College, and audiences at the Wallis/William Thomson Conference in Rochester, the Caltech SISL retreat, the SUNY Stony Brook Game Theory Conference, and the matching conference at Universidad Autónoma in Barcelona. Special thanks are due to Lozan Ivanov for carefully proof-reading the whole manuscript.

[2]Division of the Humanities and Social Sciences, California Institute of Technology. `fede@caltech.edu`

Abstract

This paper studies the falsifiability of two-sided matching theory when agents' preferences are unknown. A collection of matchings is *rationalizable* if there are preferences for the agents involved so that the matchings are stable. We show that there are non-rationalizable collections of matchings; hence, the theory is falsifiable. And we characterize the rationalizable collections of matchings, which leads to a test of matching theory in the spirit of revealed-preference tests of individual optimizing behavior.

1. Introduction

This paper presents results on the positive implications of two-sided matching theory, as first developed by Gale and Shapley (1962). The existing applications of the theory have mostly dealt with normative questions of market design. We develop the testable implications of the theory, and present results akin to the *revealed preference* tests of consumer theory. The tests are formulated in abstract settings, but can be taken to real data by using identifying assumptions already employed by applied researchers, and can hopefully serve as basis for a statistical theory of testing in matching markets.

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 over potential partners from the opposite set. The theory studies matchings that have the core property; the core matchings are called "stable." Matching models have been studied extensively since Gale and Shapley's (1962) seminal paper: Al Roth's online bibliography lists close to 500 papers.

The literature has focused on the structure of stable matchings when agents' preferences are given. Hence, to test the theory using existing results, one must know the agents' preferences. We study the problem of which matchings can be stable when agents' preferences are unknown. Concretely, given a collection of matchings, $\mu_1, \mu_2 \ldots \mu_k$, we ask if there are preferences for the agents involved so that all these matchings are stable. When this is the case, we say that the set of matchings is *rationalizable*.

The problem is important because it is often difficult to infer agents' preferences, and it is important to understand the implications of the theory when preferences are unobserved. One issue is that the theory may not have testable implications—perhaps all collections of matchings can be rationalized with suitable preferences. A second issue is, if the theory has testable implications, what are they? Can we characterize the rationalizable collections of matchings in a way that is useful for empirical work?

In this paper, we show: (1) that the theory is testable, so there are non-rationalizable sets of matchings; and (2) we provide a series of results, leading up to a characterization of the rationalizable sets of matchings. The characterization is in graph-theoretic terms. A necessary condition is simply that a certain graph has no odd cycles. A necessary and sufficient condition is in terms of no odd cycles and a certain integral polynomial system.

We 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 of the term. Second, we consider the problem of when purely randomly generated matchings would be rationalizable. We show that the probability of rationalizing a fixed number of random matchings remains bounded away from zero as the number of agents grows. So in large populations, one needs large samples of matchings for the theory to have power. Third, we discuss rationalization by an alternative solution concept: von Neumann-Morgenstern stable sets (Ehlers, 2007).

The rest of the introduction presents a brief description of the nature of the results, and a discussion of how the results are related to actual empirical work on matching.

The problem of rationalizing matchings is part of a larger research program of studying refutability in economics. This program is best known for Samuelson's (1947), Richter's (1966), and Afriat's (1967) theories of *revealed preference* in individual decision-making. But revealed-preference theory does not help in 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. Refutability has also been studied in general-equilibrium theory (e.g. Brown and Matzkin (1996) and Brown and Shannon (2000)) and non-cooperative game theory (e.g. Ledyard (1986), Sprumont (2000), Lehrer and Rosenberg (2006), Lehrer, Rosenberg, and Shmaya (2006) and Shmaya (2006)), but the results are, again, not useful in matching theory.

There is a distinct source of testable implications in matching theory. The classical results on stable matchings imply a coincidence of interest within the same side of the market, and opposition of interest across the market. We show that, essentially, stability is characterized by a version of the coincidence/opposition property which holds for any pair of matchings. In the classical results, the coincidence/opposition property holds for all agents with respect to certain matchings, and for all pairs of matchings with respect to certain agents. We show that there is a coincidence/opposition property that holds for all agents in any pair of matchings; this property characterizes stability, and it is the source of testable implications in matching theory.

The coincidence/opposition property implies that there are non-rationalizable sets of matchings. We show that these sets must involve some agents who are matched to the same partner in more than one matching. For this reason, in empirical tests of the theory, it is crucial to be able to identify some individuals in different matchings as the same agent. 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 which 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. So one treats the firms with the same observable characteristics as identical agents (and make an analogous assumption for the buyers).

A well-known result in matching theory is that the set of stable matchings forms a distributive lattice (see Knuth (1976, p. 56), who attributes the result to John Conway). Our results are related to a problem posed by Knuth on the universe of lattices that can be stable sets of matching markets. Blair (1984) gave the first and seemingly definitive answer to the problem. Blair proves that, for any distributive lattice L, there are sets of men and women, and a preference profile, so that the resulting set of stable matchings is lattice isomorphic to L. 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. But 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 some of our results (see the remark after Lemma 5) as a characterization of the matching lattices that are stable. Our results imply that matching lattices have other properties, in addition to distributivity.

2. Statement of the problem.

2.1. Preliminary definitions.

In this paper, we 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 $(v', v) \in E$; i.e. 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 $v' = v_K$ and 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. A *cycle* is a path $v_1, v_2, \ldots, v_K$ with $v_1 = v_K$.

2.2. The Model

Let M and W be disjoint, finite sets. We call men the elements of M and women the elements of W. A *matching* is a function $\mu : M \cup W \to M \cup W \cup \{\emptyset\}$ 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 is unmatched in μ (she/he is single). 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 usually 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 $\mathcal{H} = \{\mu_1, \ldots \mu_K\}$ be a set of matchings ($\mathcal{H} \subseteq \mathcal{M}$). The problem we study is: *When is there a preference profile P such that $\mathcal{H} \subseteq S(P)$*. We shall say that $\mathcal{H}$ can be *rationalized* when this is the case, and that P *rationalizes* $\mathcal{H}$. In the introduction we relate rationalizability to actual empirical tests of matching theory.

Note that we assume the same sets of agents are involved in each of the matchings in $\mathcal{H}$. In Echenique (2006), we discuss 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 $\mu \in \mathcal{H}$. This assumption is 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. Note that this model allows agents to be single (as is standard in matching theory), we are only assuming that the given matchings in $\mathcal{H}$ have no single agents.

We 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 $\mu_W = \bigwedge S(P)$ be the man-optimal and women-optimal stable matchings, respectively (Gale and Shapley, 1962). 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 $\mu' \notin S(P)$, which contradicts that $\mathcal{M} \subseteq S(P)$. □

PROPOSITION 2: *If, for all m, $\mu_i(m) \neq \mu_j(m)$ for all $\mu_i, \mu_j \in \mathcal{H}$ with $i \neq j$, 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 $i < j$; set $\emptyset\, P(m)\, w$ if $w \neq \mu(m)$, for all $\mu \in \mathcal{H}$, and order arbitrarily these w with $\emptyset\ P(m)\ w$.

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 $i > j$; set $\emptyset \, P(w) \, m$ if $m \neq \mu(w)$, for all $\mu \in \mathcal{H}$, and order arbitrarily these w with $\emptyset \, P(m) \, w$.

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 preferences constructed in the proof of Proposition 2 do not imply $\mathcal{H} = S(P)$. A rationalizing preference profile will typically give $\mathcal{H}$ as a proper subset of $S(P)$. Example 7 presents a more subtle instance of a $\mathcal{H}$ which is a proper subset of $S(P)$, for any rationalizing P.[1]

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.

Note that, in fact, there are no rationalizing preferences for which μ_1 and μ_2 are the only stable matchings: the cases not covered by Proposition 2 by re-labeling the matchings are the cases where m_1 and m_2 prefer one of the two matchings while m_3 and m_4 prefer the other; for example m_1 and m_2 prefer their partner in μ_1 over μ_2 while m_3 and m_4 prefer their partner in μ_2

[1] The question of which $\mathcal{H}$ satisfy $\mathcal{H} = S(P)$ for some P is also interesting, but seems to require different arguments than we have used here.

over μ_1. In that case, however, the matching described above, matching m_1 and m_2 as in μ_1, and m_3 and m_4 as in μ_2, is also stable.

Propositions 1 and 2 say that some collections of matchings are not rationalizable, and that failures of rationalizability arise from having some agents match with the same partner in more than one matching. But there is too much slack between the cases covered by Propositions 1 and 2: Everything cannot be rationalized and matchings where all agents have unique partners can be rationalized. As bounds on what can be rationalized, these are too coarse. In the rest of the paper, we present increasingly tighter results, building up to a characterization of the sets of matching that can be rationalized. The next section presents an example illustrating why one may fail to rationalize a set of matchings.

3. An Illustration.

Here we present a simple example that illustrates the ideas behind the results in the paper. Consider the following example, with four men, four 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

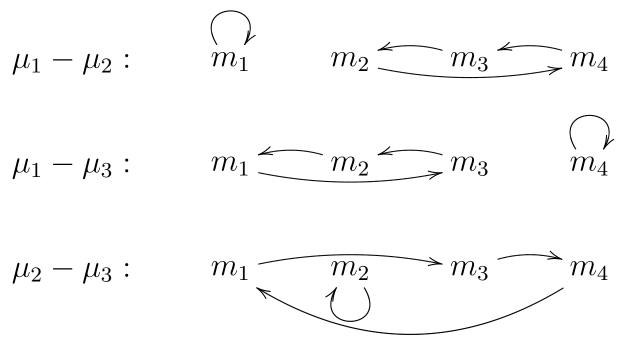
Let us construct preferences that would rationalize $\mathcal{H} = \{\mu_1, \mu_2, \mu_3\}$. We can consider all women that a man is never matched to as unacceptable. For example, set $\emptyset \, P(m_1) \, w_3$ and $\emptyset \, P(m_1) \, w_4$. To do this can only help in rationalizing $\mathcal{H}$: it eliminates the need to check for blocks by agents who are not matched in any of the matchings in $\mathcal{H}$. The issue, then, is how to specify preferences among the men's' 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 $\mu_1(m_1) = \mu_2(m_1)$. Next, consider m_2, and 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, saying that $\mu_1(m_3)\, P(m_3)\, \mu_2(m_3)$ 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. Note that the result would be the same if we had started with $\mu_1(m_2)\ P(m_2)\ \mu_2(m_2)$ instead of $\mu_2(m_2)\ P(m_2)\ \mu_1(m_2)$.

More generally, the lattice structure on $S(P)$ lies behind agreement of any two men who are related by relation "m's partner in μ_1 is m''s partner in μ_2." Note that $\mu_1 \vee \mu_2$, obtained by giving each men his best partner in μ_1 and μ_2, in $S(P)$ is a matching. So $\mu_2(m_2)P(m_2)\mu_1(m_2)$ implies $\mu_2(m_3)P(m_3)\mu_1(m_3)$, or both m_2 and m_3 would be assigned $\mu_2(m_2)$ as partner in $\mu_1 \vee \mu_2$, and then $\mu_1 \vee \mu_2$ would not be a matching. 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 $\mathcal{H}$. For example, the graph corresponding to μ_1 and μ_2 has M as vertex-set and (directed) edges given by the relation that one man's partner in μ_1 is the related man's partner in μ_2. So there is an 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 this argument restricts 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.

Assume 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 we denote by $\mathbf{C}$ below. The second is the relation between connected components in $\mathbf{C}$, derived from having agents with the same partners in two different matchings. In this section we 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 $(M, E(\mu_i, \mu_j))$. See Section 3 for examples of these.

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 (its proof is trivial and thus omitted).

Lemma 4: *The following statements are equivalent:*

1. *C is a connected component of $(M, E(\mu_i, \mu_j))$*
2. *$\mu_i(C)$, the image of C through μ_i, 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 C is a cycle, and $\mu_j(C) = \mu_i(C)$.

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:*

$$\mu_i(m)\ P(m)\ \mu_j(m) \text{ for all } m \in C$$
$$\text{and } \mu_j(w)\ P(w)\ \mu_i(w) \text{ for all } w \in \mu_i(C); \tag{1}$$
$$\mu_j(m)\ P(m)\ \mu_i(m) \text{ for all } m \in C$$
$$\text{and } \mu_i(w)\ P(w)\ \mu_j(w) \text{ for all } w \in \mu_i(C). \tag{2}$$

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

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

then P rationalizes $\mathcal{H}$.

Remark: The first statement in Lemma 5 is a refinement of the classical results on opposition and coincidence of interest in matching markets. The classical results say 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. There may still be men, for example, who disagree on the ranking of two matchings, but they must be in different components (see Example 7 where certain men must disagree in any rationalizing preference profile).

The first part of Lemma 5 says that this coincidence/opposition holds for any pair of matchings within the connected components of the corresponding graph. Using the lattice structure on $S(P)$ (Knuth, 1976), it can be restated as follows. If P rationalizes $\mathcal{H}$, then for any $C \in \mathbf{C}(\mu_i, \mu_j)$, either (3) or (4) must hold:

$$(\mu_i \wedge \mu_j)|_C = \mu_i|_C \text{ and } (\mu_i \vee \mu_j)|_C = \mu_j|_C \tag{3}$$

$$(\mu_i \wedge \mu_j)|_C = \mu_j|_C \text{ and } (\mu_i \vee \mu_j)|_C = \mu_i|_C.. \tag{4}$$

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 we show in the rest of the paper, to construct such preferences is not trivial.

These components of $(M, E(\mu_i, \mu_j))$ are also used by Irving and Leather (1986) (see also Roth and Sotomayor (1990), Section 3.2), and in a recent paper on the assignment game by Nuñez and Rafels (2006). Irving and Leather construct certain graphs from given preference profiles, and use the resulting cycles to find new stable matchings. When a stable matching is found, Irving and Leather's components coincide with ours. Nuñez and Rafels use them to study the dimension of the core of the assignment game.

Proof: We prove 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, $C = \{m^1, \dots m^L\}$, 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. We shall prove that $\mu_i(m)\,P(m)\,\mu_j(m)$ for all $m \in C$. Now, $S(P)$ has a lattice structure (Knuth, 1976), and $(\mu_i \vee \mu_j)$ is obtained by letting $(\mu_i \vee \mu_j)(m)$ be the best, according to $P(m)$, of $\mu_i(m)$ and $\mu_j(m)$. Then, $(\mu_i \vee \mu_j)(m^l) = \mu_i(m^l)$. Now 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

$$(\mu_i \vee \mu_j)(m^{l+1}) = \mu_j(m^{l+1}) = \mu_i(m^l) = (\mu_i \vee \mu_j)(m^l),$$

and $\mu_i \vee \mu_j$ would not be a matching. 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 is in would be a singleton and would not coincide with $\mu_i(C)$ (Lemma 4). Now we 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 that $\mu_j(m^l)\ P(m^l)\ \mu_j(m^l)$ for some l implies statement (2) is analogous.

We 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 exist such an 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, $\mu_j(w)\ P(w)\ \mu_i(w) = m$. 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 matched 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 those relations. Here we 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 non-singleton elements of $\mathbf{C}(\mu_i, \mu_j)$, for any two distinct $\mu_i, \mu_j \in \mathcal{H}$ with $i < j$. That is,

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

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 $\mathbf{C}(\mu_h, \mu_k)$, we abuse notation and regard each "copy" of C as a different element of $\mathbf{C}$. As a result, for each $C \in \mathbf{C}$ there is a unique pair (μ_i, μ_j) such that $C \in \mathbf{C}(\mu_i, \mu_j)$. This abuse does not, we believe, confuse, and makes the notation lighter.

We define two binary relations on the elements of $\mathbf{C}$, and denote them by $\vartriangle$ and $\triangledown$. The meaning of the relations is as follows. We can regard C and C' (in $\mathbf{C}$) as equivalent if they share one matching in the pair generating the graphs of which they are components, *and* there is an agent in $C \cap C'$ with the same partner in the differing matchings. We write $C \vartriangle C'$ or $C \triangledown C'$ depending on which of the matchings is the same, and which has an agent with the same partner: $\vartriangle$ checks whether starting from i, the other two indexes j and k are on the same side of i, i.e. whether both $j < i$ and $k < i$ or both $j > i$ and $k > i$; $\triangledown$ checks whether i is in between j and k.

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

- *either $C \in \mathbf{C}(\mu_i, \mu_j)$ and $C' \in \mathbf{C}(\mu_i, \mu_k)$*
  or $C \in \mathbf{C}(\mu_j, \mu_i)$ and $C' \in \mathbf{C}(\mu_k, \mu_i)$,
- *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 $\{1, 2, \ldots K\}$, such that*

- *either $C \in \mathbf{C}(\mu_i, \mu_j)$ and $C' \in \mathbf{C}(\mu_k, \mu_i)$*
  or $C \in \mathbf{C}(\mu_j, \mu_i)$ and $C' \in \mathbf{C}(\mu_i, \mu_k)$,
- *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 notation for the binary relation $\triangle$ and $\mathbf{E}_\bigtriangledown$ is the binary relation $\bigtriangledown$. This duplicate notation is useful.

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

> 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 $\bigtriangledown$ edges.*

Theorem 6 follows from Lemma 10 below. The idea behind the theorem is simple: each $C \in \mathbf{C}$ is "oriented" by whether the men in C prefer the first or the second matching in the pairwise graph from which C is taken. The $\triangle$ relation preserves the orientation while $\bigtriangledown$ reverses it. Hence there cannot be a cycle with an odd number of $\bigtriangledown$s.

Remark: The necessary condition in Theorem 6 can be checked in polynomial time. Note that the pairwise graphs $(M, E(\mu_i, \mu_j))$ can be constructed in time polynomial in the number of agents. The absence of odd cycles is equivalent to the graph (of the equivalence classes of $\triangle$) being bipartite, and this can be checked in linear time.

The following example illustrates the use of the pairwise graphs and relations $\triangle$ and $\bigtriangledown$. It also presents an instance of an $\mathcal{H}$ for which there is no rationalizing preference profile P with $\mathcal{H} = S(P)$.[2]

Example 7: Let $M = \{m_1, m_2, m_3, m_4, m_5\}$ and $W = \{w_1, w_2, w_3, w_4, w_5\}$. Let $\mathcal{H} = \{\mu_1, \mu_2, \mu_3\}$ defined as:

	m_1	m_2	m_3	m_4	m_5
μ_1	w_1	w_2	w_3	w_4	w_5
μ_2	w_1	w_3	w_2	w_5	w_4
μ_3	w_3	w_2	w_5	w_1	w_4 .

Note that

$$\mathbf{C}(\mu_1, \mu_3) = \{\{m_2\}, \{m_1, m_3, m_4, m_5\}\}$$
$$\mathbf{C}(\mu_2, \mu_3) = \{\{m_5\}, \{m_1, m_2, m_3, m_4\}\}$$
$$\mathbf{C}(\mu_1, \mu_2) = \{\{m_1\}, \{m_2, m_3\}, \{m_4, m_5\}\}.$$

Write $C_{1,3}$ and $C_{2,3}$ for the non-singleton elements of $\mathbf{C}(\mu_1, \mu_3)$ and $\mathbf{C}(\mu_2, \mu_3)$, respectively. Write $C^1_{1,2}$ for the $\{m_2, m_3\}$ element, and $C^2_{1,2}$ for $\{m_4, m_5\}$, of $\mathbf{C}(\mu_1, \mu_2)$.

Then $\mu_1(m_1) = \mu_2(m_1)$ implies that $C_{1,3} \triangle C_{2,3}$, $\mu_1(m_2) = \mu_3(m_2)$ implies $C_{2,3} \bigtriangledown C^1_{1,2}$, and $\mu_2(m_5) = \mu_3(m_5)$ implies $C_{1,3} \triangle C^2_{1,2}$. These are all the relations among components. Note that there are no cycles:

$$C^2_{1,2} \triangle C_{1,3} \triangle C_{2,3} \bigtriangledown C^1_{1,2}.$$

[2]Example 3 is another instance, but it is not very subtle, since it essentially involves two separate two-men, two-women, matchings.

Let P be a rationalizing profile. Since $C_{1,3} \bigtriangleup C_{2,3}$, all men must either prefer their partner in μ_1 over μ_3 and μ_2 over μ_3, or prefer μ_3 over μ_1 and μ_3 over μ_2. Without loss of generality, suppose the first case holds. Then $C^2_{1,2} \bigtriangleup C_{1,3}$ implies that m_4 and m_5 prefer their partner in μ_1 over μ_2. And $C_{2,3} \bigtriangledown C^1_{1,2}$ implies that m_2 and m_3 prefer their partner in μ_2 over μ_1. But then one can check that the matching μ' must be stable for P, where μ' is:

	m_1	m_2	m_3	m_4	m_5
μ'	w_1	w_3	w_2	w_4	w_5

EXAMPLE 8: Consider an example with four men and four women, and $\mathcal{H} = \{\mu_1, \mu_2\}$, where μ_1 and μ_2 are the first two matchings in Section 3. Clearly, $\mathcal{H}$ is rationalizable; the preferences where all men rank their partner in μ_1 over their partner in μ_2, and women have the opposite preferences, rationalize $\mathcal{H}$. As we have seen, though, when we add matching μ_3 in Section 3, the resulting set of matchings cannot be rationalized. One way of understanding the effect of adding μ_3 is that one cannot add it on top of μ_1 and μ_2 in the men's preferences because then $\mu_3(m_2)P(m_2)\mu_1(m_2)P(m_2)\mu_2(m_2)$, while $\mu_3(m_2) = \mu_2(m_2)$; one cannot add it below μ_1 and μ_2 because then $\mu_1(m_4)\ P(m_4)\ \mu_2(m_4)\ P(m_4)\ \mu_3(m_4)$, while $\mu_1(m_4) = \mu_3(m_4)$; and so on.

In light of Lemma 5, Theorem 6 expresses the requirement that the coincidence/opposition of interest property be consistent with the connections across components implied by the agents for whom two matchings are the same. The theorem does not guarantee that the components and the relations between components are compatible with well-defined preferences.

A first requirement of the compatibility with well-behaved preferences is that $\mathbf{C}$, $\mathbf{E}_\bigtriangleup$ and $\mathbf{E}_\bigtriangledown$ cannot imply intransitivity. We express this requirement by making $\bigtriangleup$ a larger relation: we define a monotone increasing sequence $\left\{\mathbf{E}^k_\bigtriangleup\right\}$, and work with the larger binary relation $\mathbf{D}_\bigtriangleup = \cup_{k=1}^{\infty} \mathbf{E}^k_\bigtriangleup$. Let $\mathbf{E}^0_\bigtriangleup =$

$\mathbf{E}_\triangle$. Given $\mathbf{E}_\triangle^k$, for $k \geq 0$, let $\mathbf{E}_\triangle^{k+1}$ be those edges (C, C') between elements in $\mathbf{C}$ such that either $(C, C') \in \mathbf{E}_\triangle^k$ and/or there are 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 $\left(\mathbf{C}, \mathbf{E}_\triangle^{k-1}\right)$

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 $\left(\mathbf{C}, \mathbf{E}_\triangle^{k-1} \cup \mathbf{E}_\triangledown\right)$ between C and $\tilde{C}$ with an odd number of $\triangledown$s.

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

Theorem 9: *If $\mathcal{H}$ is rationalizable, then $(\mathbf{C}, \mathbf{D}_\triangle \cup \mathbf{E}_\triangledown)$ can have no cycle with an odd number of $\triangledown$ edges.*

The proof of Theorem 9 is below.

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 $(\forall m \in C)(\mu_i(m)\ P(m)\ \mu_j(m))$ 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 10: *Let $\mathcal{H}$ be rationalizable and $(C_1, \ldots, C_N)$ be a cycle in $(\mathbf{C}, \mathbf{E}_\triangle \cup \mathbf{E}_\triangledown)$. Then, for each n and L, mod N,*

$$d(C_n) = \Pi_{l=n}^{L}(-1)^{\mathbf{1}\{C_l \triangledown C_{l+1}\}} d(C_L) \qquad (5)$$

Proof: Let P rationalize $\mathcal{H}$. We 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

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

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 (Proof of Theorem 6): Lemma 10 implies Theorem 6 because any cycle $C_1, \ldots C_N$ with an odd number of $\bigtriangledown$s implies that $d(C_1) = (-1)d(C_1)$.□

Proof (Proof of Theorem 9): Let $\mathcal{H}$ be rationalizable by the preference profile P. We prove Theorem 9 by induction. By Theorem 6, $(\mathbf{C}, \mathbf{E}_\bigtriangleup \cup \mathbf{E}_\bigtriangledown) = \left(\mathbf{C}, \mathbf{E}^0_\bigtriangleup \cup \mathbf{E}_\bigtriangledown\right)$ can have no cycle with an odd number of $\bigtriangledown$. Lemma 10 implies that the formula (5) holds in $\left(\mathbf{C}, \mathbf{E}^0_\bigtriangleup \cup \mathbf{E}_\bigtriangledown\right)$. Suppose this statement is true of $\left(\mathbf{C}, \mathbf{E}^k_\bigtriangleup \cup \mathbf{E}_\bigtriangledown\right)$; if we prove that it is true of $\left(\mathbf{C}, \mathbf{E}^{k+1}_\bigtriangleup \cup \mathbf{E}_\bigtriangledown\right)$ then the proof of the theorem is done.

Let $(C, C') \in \mathbf{E}^{k+1}_\bigtriangleup \backslash \mathbf{E}^k_\bigtriangleup$. We shall prove that $d(C) = d(C')$. Let i, j, h and $\tilde{C} \in \mathbf{C}$ with $C \cap \tilde{C} \cap C' \neq \emptyset$ be 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 $\left(\mathbf{C}, \mathbf{E}^{k-1}_\bigtriangleup\right)$, by Lemma 10, 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 we are in 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 $\left(\mathbf{C}, \mathbf{E}_{\triangle}^{k} \cup \mathbf{E}_{\bigtriangledown}\right)$, (5) holds in $\left(\mathbf{C}, \mathbf{E}_{\triangle}^{k+1} \cup \mathbf{E}_{\bigtriangledown}\right)$. Then $\left(\mathbf{C}, \mathbf{E}_{\triangle}^{k+1} \cup \mathbf{E}_{\bigtriangledown}\right)$ 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 well-defined preferences, but not all of them. In this section we 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 Cs. 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 12) that, when $\mathcal{H}$ is rationalizable, the rationalizing P will generally not be unique.[3]

A 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 $C \in \mathcal{C}$

[3] Another interesting question (posed by an anonymous referee) is if there is a rationalizing P such that $S(P)$ is minimal.

and $C' \in \mathcal{C}'$ with $C \bigtriangledown C'$.

If $(\mathbf{C}, \mathbf{D}_{\bigtriangleup} \cup \mathbf{E}_{\bigtriangledown})$ 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 $(\mathbf{C}, \mathbf{D}_{\bigtriangleup})$ 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 are 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 $C \cap C' \cap C'' \neq \emptyset$. A triple $(\mathcal{C}, \mathcal{C}', \mathcal{C}'')$ stands in relation B if its components have non-empty intersection, and correspond to three pairwise graphs, with indexes i, j, j, h and i, h, and $i < j, h$.

> THEOREM 11: *$\mathcal{H}$ is rationalizable if and only if $(\mathbf{C}, \mathbf{D}_{\bigtriangleup} \cup \mathbf{E}_{\bigtriangledown})$ has no cycle with an odd number of $\bigtriangledown$ edges, 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: We only prove the "if" statement; "only if" is straightforward given the results in the previous section. Let $(\mathbf{C}, \mathbf{D}_{\bigtriangleup} \cup \mathbf{E}_{\bigtriangledown})$ 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. We 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 $\mu_j(m)\, P(m)\, \mu_h(m)$. We 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 (5) 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 (5) 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, $d(C) = d(C') = 1$ implies $d(\mathcal{C}) = d(\mathcal{C}') = 1$, 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,

$$\left.\begin{array}{l} i < j < h \\ \mu_j(m)\, P(m)\, \mu_i(m) \\ \mu_h(m)\, P(m)\, \mu_j(m) \end{array}\right\} \text{ implies } \mu_h(m)\, P(m)\, \mu_i(m). \tag{6}$$

This gives us $\mu_i(m)\, P(m)\, \mu_h(m)$ in the case $h < j < i$ by applying (6) 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 $C \bigtriangleup C''$ we have $d(C) = d(C'')$ so there is nothing to prove. Suppose then that $C \bigtriangleup C''$ is false. It cannot be that $C' \bigtriangleup C''$, since that would imply $C' \bigtriangleup C$ by the construction of $\mathbf{D}_\bigtriangleup$, and $d(C') \neq d(C)$ implies that C' and C are disconnected in $(\mathbf{C}, \mathbf{D}_\bigtriangleup)$. So it must be the case that all of C, C' and C'' are disconnected in $(\mathbf{C}, \mathbf{D}_\bigtriangleup)$. 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 $C' \bigtriangleup C''$, then $d(C'') = 1$ so there is nothing to prove. Second, it cannot be that $C \bigtriangleup C''$, since that would imply $C \bigtriangleup C'$ by the construction of $\mathbf{D}_\bigtriangleup$, and $d(C') \neq d(C)$ implies that C' and C are disconnected in $(\mathbf{C}, \mathbf{D}_\bigtriangleup)$. 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.

We 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 any 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, we show that matching theory is generally 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_a be the number of agents that a is not matched to in any matching in $\mathcal{H}$. Note that u_a counts men if a is a woman and women if a is a man. 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 12: *If $\mathcal{H}$ is rationalizable, then it is rationalizable by at least*

$$|M|^{2|M|}\left(\Pi_{m\in M}u_m\Pi_{w\in W}u_w\right)$$

different preference profiles, of which at least

$$\Pi_{m\in M}u_m(|M|-|u_m|)\Pi_{w\in M}u_w(|M|-|u_w|)$$

are essentially different.

For example, suppose 20 men and women, and that each agent is matched to 10 agents in some matching in $\mathcal{H}$. If $\mathcal{H}$ is rationalizable, it is rationalizable by at least 10^{80} essentially different preference profiles.

PROOF: Let P rationalize $\mathcal{H}$ such that any unmatched agents are considered unacceptable. Fix a man m. For each w that m is not matched to in any matching in $\mathcal{H}$, we 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 (or $|W|-u_m$ places in the ranking among m acceptable partners for the second calculation). 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. □

7. Rationalization as a von Neumann-Morgenstern stable set

We have restricted attention to the rationalization of the matchings in $\mathcal{H}$ as stable matchings, but one could study other solution concepts as well. We discuss briefly the rationalization by von Neumann-Morgenstern stable sets; the first study of these in matching theory is Ehlers (2007).[4]

Fix a set of men, M, and women, W, and a preference profile P. Say that a matching μ' *dominates* a matching μ if there is a pair $(m, w) \in M \times W$ with $w = \mu'(m)$, $w\, P(m)\, \mu(m)$, and $m\, P(w)\, \mu(w)$. A set of matchings $V \subseteq \mathcal{M}$ is a *von Neumann-Morgenstern stable set* if (a) no matching in V dominates another matching in V, and (b) if $\mu \in \mathcal{M} \setminus V$, then there is $\mu' \in V$ which dominates μ.

Proposition 13: *If $\mathcal{H}$ is rationalizable, then it is rationalizable by a P such that $S(P)$ is a von Neumann-Morgenstern stable set.*

Proof: Let $\mathcal{H}$ be rationalizable. Then it is rationalizable by a preference profile P in which, for any man m, a woman w is unacceptable if she is not matched to him in any matching in $\mathcal{H}$. Similarly for women. We argue that $S(P)$ is a von Neumann-Morgenstern stable set: By Theorem 1 in Ehlers (2007), a set V is a von Neumann-Morgenstern stable set if, for any $\mu \in V$, there is no pair of agents who are matched in some matching in V and who would block μ. For any matching with the constructed preferences, a block must be a block of agents who are matched in some matching in $\mathcal{H}$, as any

[4] One could also consider bargaining sets (Klijn and Massó, 2003; Echenique and Oviedo, 2006).

other pair would be mutually unacceptable. So $S(P)$ is the set of matchings which are not blocked by pairs of agents who are matched in $\mathcal{H}$. In addition, the matchings in $S(P)$ are individually rational, so with the constructed P they must have agents matched to partners they are matched with in some matching in $\mathcal{H}$. Hence $S(P)$ is a stable set by Theorem 1 in Ehlers (2007).□

Proposition 13 implies that the notion of rationalizing by a von Neumann-Morgenstern stable set is weaker than rationalization by stable matchings, and leads to another use of Theorem 11, which then provides preference profiles under which one obtains a rationalization by von Neumann-Morgenstern stable sets.

The proposition follows very obviously from setting agents as unacceptable when they are not partners in a matching in $\mathcal{H}$. Similarly, if $\mathcal{H}$ is not rationalizable by preferences with this property, it will not be a subset of a von Neumann-Morgenstern stable set for any such preferences. But it could be in a stable set for preferences where unmatched agents are acceptable.

8. 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.

We 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 $\mathcal{M}_n$. That is, $\mathcal{H}_k$ is a random set of matchings obtained by choosing k matchings from $\mathcal{M}_n$, where each possible k-tuple of matchings has the same probability of being selected.

PROPOSITION 14: *If k is fixed,*

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

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 we remarked above, the message in Proposition 14 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

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.

Brown, D. J., and R. L. Matzkin (1996): "Testable Restrictions on the Equilibrium Manifold," *Econometrica*, 64(6), 1249–1262.

Brown, D. J., and C. Shannon (2000): "Uniqueness, Stability, and Comparative Statics in Rationalizable Walrasian Markets," *Econometrica*, 68(6), 1529–1439.

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

Echenique, F. (2006): "What Matchings Can Be Stable? The Testable Implications of Matching Theory," Caltech SSWP 1252.

Echenique, F., and J. Oviedo (2006): "A Theory of Stability in Many-to-Many Matching Markets," *Theoretical Economics*, 1(2), 233–273.

Ehlers, L. (2007): "Von Neumann-Morgenstern stable sets in matching problems," *Journal of Economic Theory*, 134(1), 537–547.

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.

Irving, R. W., and P. Leather (1986): "The Complexity of Counting Stable Marriages," *SIAM Journal on Computing*, 15(5), 655–667.

Klijn, F., and J. Massó (2003): "Weak Stability and a Bargaining Set for the Marriage Model," *Games and Economic Behavior*, 42(1), 91–100.

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.

Ledyard, J. O. (1986): "The Scope of the Hypothesis of Bayesian Equilibrium," *Journal of Economic Theory*, 39(1), 59–2.

Lehrer, E., and D. Rosenberg (2006): "What restrictions do Bayesian games impose on the value of information?," *Journal of Mathematical Economics*, 42, 343–357.

Lehrer, E., D. Rosenberg, and E. Shmaya (2006): "Signaling and Mediation in Bayesian Games," mimeo, Tel Aviv University.

Nuñez, M., and C. Rafels (2006): "On the dimension of the core of the assignment game," mimeo, Universidad de Barcelona.

Richter, M. K. (1966): "Revealed Preference Theory," *Econometrica*, 34(3), 635–645.

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.

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

Shmaya, E. (2006): "The Value of Information Structures in Zero-sum Games with Lack of Information on One Side," *International Journal of Game Theory*, 34(2), 155–165.

Sprumont, Y. (2000): "On the Testable Implocations of Collective Choice Theories," *Journal of Economic Theory*, 93, 205–232.