# One-to-One Matching with Interdependent Preferences[1]

Ayşe Mumcu

*Department of Economics, Bogazici University and University of Pennsylvania*

and

Ismail Saglam[2]

*Department of Economics, Bogazici University and Massachusetts Institute of Technology*

November 2, 2006

In this paper, we introduce *interdependent* preferences to a classical one-to-one matching problem that allows for the prospect of being single, and study the existence and properties of stable matchings. We obtain the relationship between the stable set, the core, and the Pareto set, and give a sufficiency result for the existence of the stable set and the core. We also present several findings on the issues of gender optimality, lattices, strategy-proofness, and rationalizability.

*Keywords:* One-to-one matching, externalities.

*JEL Classification Numbers:* C71, C78, D62.

[1]The authors thank Jan Eeckhout and Muhamet Yıldız for helpful discussions. The second author acknowledges the support of Turkish Academy of Sciences, in the framework of Distinguished Young Scientist Award Program (TUBA-GEBIP). The usual disclaimer applies.

[2]*Corresponding author.* Department of Economics, Massachusetts Institute of Technology, 50 Memorial Drive, E52-251D, Cambridge, MA 02142, USA. E-mail: saglam@mit.edu

# 1 Introduction

The standard assumption in matching models pioneered by Gale and Shapley (1962) is that agents belonging to one side of the market have preferences over the agents in the opposite side. However, in many social settings people care about not only whom they are matched with but also the partners of the others. That is, in a marriage market a person may prefer to be single if majority of people in the society is single, but to be married if the majority is married. In school assignment, parents may prefer to send their kids to a particular school only if it is mostly populated by kids belonging to a certain group. In a labor market, firms would be very much interested in what subset of workers is hired by their competitors or in the case of couples searching for jobs each mate cares about which firm his/her mate is matched with. In all these cases, individuals' preferences depend on the realized matching in the society, hence are *interdependent*.

In this paper, we introduce *interdependent* preferences to a classical one-to-one matching problem, and study the existence and properties of stable matchings. The issue of *interdependent* preferences has received due attention from at least three works in matching theory. Echenique and Yenmez (2006) study the assignment of students to colleges where students have preferences over the other students who would attend the same college. They introduce an algorithm that finds all matchings in the core, whenever it is nonempty. Klaus and Klijn (2005) study the matching in a job market with couples where individuals do not only care about the firm they are matched with but also the firm their mates are matched with. They show that a stable matching exists only when couples' preferences exhibit weak responsiveness; i.e., the unilateral improvement of one partner's job is considered beneficial for the couple as well.

Definitely, the paper which is closest to ours is by Sasaki and Toda (1996), who study the same formulation of *interdependent* preferences in one-to-one matching problems. However, our model differs from theirs as we allow the prospect of being single in matchings while their stability notion includes ours as a special case. Sasaki and Toda (1996) reasonably use the idea of conjectural equilibrium to define stability which requires a deviating pair to prefer all members of the set of conjectured matches that are likely to occur to the current one. They show that the general existence of the stable set can be guaranteed if and only if the collection of conjectured matchings coincides with the set of all conceivable matchings for any agent considering deviation.[3]

Our stability definition is a natural extension of Gale and Shapley's stability notion to the environment with *interdependent* preferences. We assume that for any given initial matching the set of conjectured matchings due to any deviation is always a singleton, consisting of a matching that differs from the initial one only in the marital status of the deviating agents and their divorced initial partners, if any.[4] We show that with *interdependent* preferences, the existence of stable matchings cannot be guaranteed. For every society, there will be a preference profile at which no stable matching exists. Besides, the stable set does not need to coincide with the core (when they both exist), except for societies in which either the set of men or the set of women is a singleton.

Because of a stronger notion of stability and a richer set of possible matches (allowing for the possibility of staying single) than what are assumed by Sasaki and Toda (1996), our model is not always compatible with the general existence of equilibrium matchings. Despite this fact, we have abstained from employing the extreme (weakest) form of Sasaki and Toda's stability that requires deviating agents always to worry about the worst case. In order to see how weak such a notion may indeed be, one can consider the following analogy bearing in mind the close parallel between Gale and Shapley's stability (and core) notion in the matching theory and the Nash (and strong Nash) solution in the non-cooperative game theory, where preferences over the admissible set of strategies are typically *interdependent*, which actually makes a situation a strategic game.[5] In a two-person game called as 'matching pennies', there exists no Nash equilibrium in pure strategies, since at each strategy pair one agent has an incentive to deviate if the other agent

[3] An important self-criticism made by Sasaki and Toda (1996) for their own approach is that the conjectural valuations (expectations) are exogeneously given. Hafalir (2006) perfects Sasaki and Toda's model by endogenizing the set of matchings that a deviating pair considers possible on the preferences of the other agents in the society.

[4] We use the common jargon of marriage models for convenience.

[5] This conventional *interdependency* in game theory must not be confused with *interdependency* of preferences modeled in many different setting in economics, in which an outcome, whenever defined over the strategies, specifies an allocation for each distinct player. This formulation allows one to define utility payoffs that are *interdependent* over the individual allocations. See Pollak (1976), Postlewaite (1998), Ok and Kockesen (2000), Charness and Rabin (2002), Gul and Pesendorfer (2005), Li (2005), and Sobel (2005) for a comprehensive examination of this new literature.

keeps his or her strategy unchanged.[6] With a weaker equilibrium concept requiring each deviating agent to consider all possible reactions of the opponents, all of the four possible outcomes may arise in equilibrium. Yet, such a 'productive' solution concept is hardly recognizable in game theory, for the induced positive result is entirely illusory. We believe that after choosing a proper stability notion under *interdependent* preferences, one should welcome the 'annoying' non-existence of stable matchings with the same composure that is shown by any game theorist in facing the frequent non-existence of Nash equilibrium (in pure strategies) in many non-cooperative games. However, we should also admit that the justification of any stable outcome by rational acts and beliefs of individuals remains to be an essential issue, which can undoubtedly be studied in isolation.

Acknowledging the possible nonexistence of stable matchings, we give a sufficiency result for the existence of a nonempty stable set and a nonempty core, imposing some restrictions on the preference profiles. We call a proper subset of all available matchings as a *top-matching collection* if each element of which is preferred by any individual in the society to any matching outside the collection. We say that two machings are *unconnected* by an agent if he or she has distinct mates under these two matchings. We then define that a matching is *reachable* from another matching by some coalition of individuals if this coalition contains anyone who unconnects the two matchings. Clearly, a top-matching collection, whenever singleton, is equal to the core of the associated marriage market, and is contained by the stable set. Moreover, a non-singleton top-matching collection, where any two of its elements are unconnected by a coalition of individuals that contains at each of the matchings either a married couple or a single individual opposing to the formation of the other matching, is equal to the core, associated with the preference profile that induces this top-matching collection. We are able to relax this sufficiency condition in an existence result for the stable set, where one has to deal with only blocking coalitions of size not exceeding two.

In this study, we also show that for any society with a nonempty core, the Pareto set has to coincide with neither the stable set nor the core, unlike in the case of *independent* preferences. While all stable matchings are Pareto optimal in any society where one of the genders has a unique representative,

[6] The game is played between two players. Each player shows either heads or tails from a coin. If both are heads or both are tails then player one receives one dollar from player two, otherwise player two wins one dollar from player one.

the stable set and the Pareto set may in general have an empty intersection.

Another finding of ours is that the properties of stable and core matchings with respect to gender-optimality under *independent* preferences cannot be replicated for a society with *interdependent* preferences. There may exist gender-optimal stable matchings that are not in the core as a direct result of the non-equivalence of the stable set and the core. Moreover, the existence of men-optimal stable and men-optimal core matchings does not guarantee the existence of women-optimal stable and women-optimal core matchings. Furthermore, even when the core is non-empty, there may exist neither a gender-optimal stable matching nor a gender-optimal core matching. As opposed to the case with *independent* preferences, the common preferences of the two sides of the market are not always opposed on the core. Consequently, the set of people who are matched may not be the same for all core matchings. Besides, the core and the stable set do not always exhibit a lattice structure.

Regarding strategic issues, we show that there are no core mechanisms that are strategy-proof, replicating the well-known impossibility theorem in the case of *independent* preferences. Besides, when a core (stable matching) mechanism is applied to a marriage market, under certain restrictions of preferences, there is always at least one agent who wants to misrepresent his/her preferences. Moreover, the core (stable matching) mechanism is also prone to successful manipulation by coalitions of men and women.

It is promising for empirical research that the matching model we consider is refutable, since for any society facing at least two different matchings there exists at least one collection of matchings, e.g. the set of all conceivable matchings, that is not rationalizable either for the stable set or the core. On the other hand, we also have a sufficiency result showing that any set containing no pairs of connected matchings is rationalizable for the stable set, and also rationalizable for the core provided that the number of pairwise unconnected matchings in this set does not exceed the number of individuals in the society. However, our refutable matching model of *interdependent* preferences is not always exactly *identifiable*, as there may exist many different preference profiles that rationalize some collections of matchings.

The organization of the paper is as follows: Section 2 introduces the model. Section 3 presents our results, and Section 4 concludes.

## 2 The Model

There are two nonempty, finite and disjoint sets of agents: a set of men, $M$ and a set of women, $W$. The society is denoted by $N = M \cup W$. There exist at least three agents in the society; i.e., $|M||W| \geq 2$. We denote a generic man by $m$, a generic woman by $w$, and a generic agent in the society by $i$.

A matching is a one-to-one function, $\mu$, from $N$ to itself, such that for each $m \in M$ and for each $w \in W$ we have $\mu(m) = w$ if and only if $\mu(w) = m$, $\mu(m) \notin W$ implies $\mu(m) = m$, and similarly $\mu(w) \notin M$ implies $\mu(w) = w$. If $\mu(m) = w$, then $m$ and $w$ are matched to one another. If $\mu(i) = i$, then $i$ remains single. When denoting a matching $\mu$, we list the mates of men $m_1, m_2, m_3, \ldots$. For example, $\mu = w_2, m_2, w_1, \ldots$ denotes a matching where $m_1$ is matched to $w_2$, $m_2$ to himself, and $m_3$ to $w_1$. Any woman not listed in $\mu$ is single. We denote by $\mu_{m,w}$ a matching obtained from $\mu$ by marrying $m$ and $w$ after divorcing them from their mates, if any, under $\mu$; i.e., $\mu_{m,w}(m) = w$, $\mu_{m,w}(\mu(m)) = \mu(m)$ if $\mu(m) \neq m$, $\mu_{m,w}(\mu(w)) = \mu(w)$ if $\mu(w) \neq w$, and $\mu_{m,w}(i) = \mu(i)$ for all $i \notin \{m, w, \mu(m), \mu(w)\}$. We denote by $\mathcal{M}^N$ the set of all possible matchings in society $N$.

Each agent has a complete, transitive, and strict preference relation over the matchings in $\mathcal{M}^N$. Men's preferences are represented by $P^m$ and women's preferences by $P^w$. The profile of all agents' preferences is denoted by $P = (P^i)_{i \in N}$. A marriage market is a triple $(M, W, P)$. The list of all agents' preferences excluding the preference of agent $i$ is denoted by $P^{-i}$. For any preference profile $P$, we denote by $P^i[k]$ the $k$th-ranked matching from top in the preference ordering $P^i$ of agent $i$. We write $\mu >_m \mu'$ to mean $m$ prefers $\mu$ to $\mu'$, and $\mu \geq_m \mu'$ to mean $m$ likes $\mu$ at least as well as $\mu'$. We also write $\mu >_M \mu'$ to denote that all men like $\mu$ at least as well as $\mu'$, with at least one man strictly preferring $\mu$ to $\mu'$. We denote by $\mu \geq_M \mu'$ that either $\mu >_M \mu'$ or that all men are indifferent between $\mu$ and $\mu'$. Similarly, we write $\mu >_w \mu'$, $\mu \geq_w \mu'$, $\mu >_W \mu'$, and $\mu \geq_W \mu'$.

In the classical one-to-one matching literature, each individual's preferences (Gale and Shapley's preferences), over the prospect of being single and the feasible mates of the opposite gender, are *independent* of the realized matching in the society. In other words, from the viewpoint of every individual, the ordering of any two matchings boils down to the ordering of the corresponding mates achieved at these two matchings. Clearly, Gale and Shapley's *independent* preferences are contained in our space of *interdependent* preferences as special cases, whenever we allow for indifferences. In what follows, we give some basic definitions.

**Definition 1.** A matching is *acceptable* to agent $i$ if

$$\mu \geq_i \mu_{i,i}.$$

A matching is called *unaccaptable* to an agent if it is not acceptable to him or her. We also say that an agent *individually blocks* matching $\mu$ (via $\mu_{i,i}$) if $\mu$ is unacceptable to him or her.

**Definition 2.** A matching is *individually rational* if it is acceptable to each agent.

**Definition 3.** For a given matching $\mu$, $(m, w)$ is a *blocking pair* if they are not matched to one another but prefer one another to their matches at $\mu$; i.e., $\mu(m) \neq w$ and

$$\mu_{m,w} >_m \mu \quad \text{and} \quad \mu_{m,w} >_w \mu.$$

**Definition 4.** A matching is *stable* if it is individually rational and if there are no blocking pairs.

We denote the set of stable matchings (the stable set) for the marriage market $(M, W, P)$ by $S(M, W, P)$.

**Definition 5.** A matching $\mu'$ *dominates* another matching $\mu$ via a (blocking) coalition $M' \cup W'$ of men and women such that $\mu'(M' \cup W') = M' \cup W'$, $\mu'(\mu(m')) = \mu(m')$ for any $m' \in M'$ if $\mu(m') \notin W' \cup \{m'\}$, $\mu'(\mu(w')) = \mu(w')$ for any $w' \in W'$ if $\mu(w') \notin M' \cup \{w'\}$, $\mu'(i) = \mu(i)$ for any $i \notin M' \cup W' \cup \mu(M' \cup W')$, and

$$\mu' >_i \mu$$

for all $i \in M' \cup W'$.

We assume in the above definition that the members of the blocking coalition seek marriage within the coalition. Moreover, the previous mate, if exists, of any agent in the blocking coalition becomes single under the new matching if he or she is not inside the blocking coalition, while the marital status of all other agents are unchanged.

**Definition 6.** The *core*, $C(M, W, P)$, is the set of all matchings dominated by no other matchings.

For most part of our results in the following section, we will make generalizations from findings obtained for some 'small' marriage markets. For that reason, we will extend the matchings and preference profiles in a convenient way to keep the marriage relationships in a given society $N$ preserved in a *larger* society $\tilde{N} \supset N$.

**Definition 7.** For any society $N$, define the extension operator $E^N$ as a function that maps each matching $\mu \in \mathcal{M}^N$ to an extended matching $E^N[\mu] \in \mathcal{M}^{\tilde{N}}$ as follows:

$$E^N[\mu](j) = \begin{cases} \mu(j) & \text{if } j \in N \\ j & \text{otherwise} \end{cases}$$

**Definition 8.** Given any marriage market $(M, W, P)$ associated with the society $N = M \cup W$, a new profile $\tilde{P}$ of a larger society $\tilde{N} = \tilde{M} \cup \tilde{W} \supset N$ is called a *society-respecting extension* of $P$ if

(i) for all $k \in \{1, 2, \ldots, |\mathcal{M}^N|\}$

$$\tilde{P}^i[k] = \begin{cases} E^N[P^i[k]] & \text{if } i \in N \\ E^N[P^m[k]] & \text{if } i \in \tilde{M} \backslash M \\ E^N[P^w[k]] & \text{if } i \in \tilde{W} \backslash W \end{cases}$$

for some $m \in M$ and $w \in W$;

(ii) for all $\mu' \in \mathcal{M}^{\tilde{N}}$, for all $i \in \tilde{N} \backslash N$ and for all $k \in \{1, 2, \ldots, |\mathcal{M}^N|\}$ such that $\mu' \neq \tilde{P}^i[k]$ and $\mu' \neq \mu'_{i,i}$,

$$\mu'_{i,i} >_i \mu'$$

under the profile $\tilde{P}$.

The extension described above respects the stable marriages in the initial society. Condition (i) keeps the preference orderings of the members of the initial society essentially unchanged while at the same time making all of the new members entering the society comply with the preferences of one of the initial members of the same gender over the set of extended initial matchings.[7] Condition (ii) along with condition (i) ensures that the new members entering the society will not *steal* the mates of the former members.

In order to clarify our basic structures and notions, we present two examples which will also be used in some of our results in the next section.

**Example 1.** Consider a society $N$ with $M = \{m_1, m_2\}$ and $W = \{w_1, w_2\}$. The seven possible matchings are given by

$$\begin{aligned} \mu_1 &= w_1, w_2 \\ \mu_2 &= m_1, w_2 \\ \mu_3 &= w_2, m_2 \\ \mu_4 &= m_1, w_1 \\ \mu_5 &= w_2, w_1 \\ \mu_6 &= w_1, m_2 \\ \mu_7 &= m_1, m_2. \end{aligned}$$

Every single agent under a given matching can induce some other matching at which he or she is married upon the mutual consent of a prospective partner. Conversely, every married agent can change (by an individual action of divorcing his or her mate) the prevailing matching to a one at which he or she is single. As a third possibility, every married agent can change a given matching by replacing his or her mate upon the consent of a prospective partner.

The ability of the agents to change the seven matchings listed above are illustrated by the directed graphs in Figures 1-4, which will also be useful in following some of our results in the next section. In these graphs, a circled $k$ stands for matching $\mu_k$. The labels $d, m$ and $dm$ on the arrows in the directed graphs respectively stand for the actions of 'divorcing', 'marrying', and 'divorcing and re-marrying'.

An agent can individually block a matching $\mu_k$ to induce some other matching $\mu_l$ only if there exists an immediate directed arrow with the label $d$ from node $k$ to node $l$ in the ability map of this agent. Moreover, a man and a woman can pairwise block matching $\mu_k$ to induce some other matching

[7] In condition (i), a new member's adopting the preference ordering of an individual of the same gender in the initial society simply aims to preserve the gender polarization, if exists, in the set of stable matchings, which will be the subject of Theorems 8-10.

$\mu_l$ only if the ability maps of these two agents both contain an immediate directed arrow with any of the labels $m$ or $dm$ from node $k$ to node $l$.

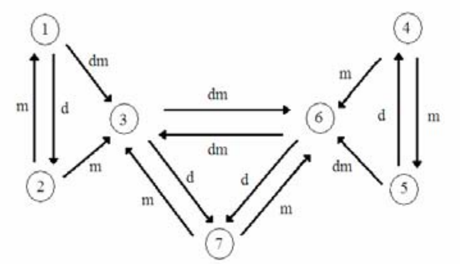


Figure 1: Ability map of $m_1$

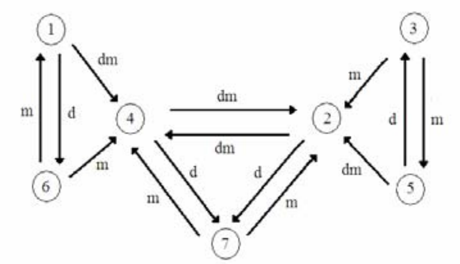


Figure 2: Ability map of $m_2$

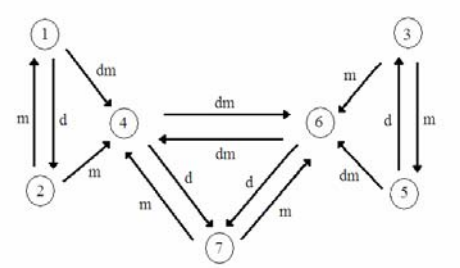


Figure 3: Ability map of $w_1$

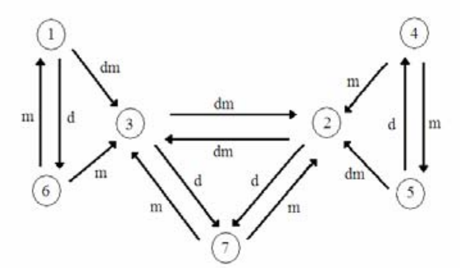


Figure 4: Ability map of $w_2$

Consider the following preferences for the individuals:

$$
\begin{aligned}
P^{m_1} &= \mu_1\mu_7\mu_5\mu_2\mu_3\mu_4\mu_6 \\
P^{m_2} &= \mu_1\mu_5\mu_7\mu_2\mu_3\mu_4\mu_6 \\
P^{w_1} &= \mu_1\mu_7\mu_5\mu_2\mu_3\mu_4\mu_6 \\
P^{w_2} &= \mu_1\mu_5\mu_7\mu_2\mu_3\mu_4\mu_6
\end{aligned}
$$

The set of stable matchings $S(M, W, P)$ is $\{\mu_1, \mu_5, \mu_7\}$ while the core $C(M, W, P)$ consists of $\mu_1$. Although for every individual $\mu_1$ dominates $\mu_5$ and $\mu_7$, all of the three matchings are stable as the above figures illustrate that any of the three nodes corresponding to these three (stable) matchings cannot be reached from the other two nodes by any individual or any pair.

Now, suppose that society $N$ enlarges to $\tilde{N} = \tilde{M} \cup W$ where $\tilde{M} = M \cup \{m_3\}$. There are thirteen possible matchings, of which seven can directly be obtained by applying the extension operator $E^N$ on $\mathcal{M}^N$ as

$$E^N[\mu_k] \;=\; \mu_k, m_3$$

for $k \in \{1, ..., 7\}$. A preference profile $\tilde{P}$ for society $\tilde{N}$ can be obtained using a *society-respecting extension* from the profile $P$. Then, it is easy to check that $S(\tilde{M}, W, \tilde{P}) = S(M, W, P)$ and $C(\tilde{M}, W, \tilde{P}) = C(M, W, P)$.

**Example 2.** Consider a society with $M = \{m_1, m_2\}$ and $W = \{w_1\}$. The three possible matchings are

$$\begin{aligned} \mu_1 &= w_1 m_2 \\ \mu_2 &= m_1 w_1 \\ \mu_3 &= m_1 m_2. \end{aligned}$$

Let the preference profile $P$ be given by

$$\begin{aligned} P^{m_1} &= \mu_2 \mu_1 \mu_3 \\ P^{m_2} &= \mu_1 \mu_2 \mu_3 \\ P^{w_1} &= \mu_1 \mu_2 \mu_3. \end{aligned}$$

Each man is most happy when the other man is married to the unique woman in the society. Nevertheless, each man prefers to being married to this woman to being single. It is easy to check that $S(M, W, P) = C(M, W, P) = \{\mu_1, \mu_2\}$.

# 3 Results

A well-known theorem by Gale and Shapley (1962) shows the existence of a stable matching for every marriage market with *independent* preferences. This result does not extend to our framework.

**Theorem 1.** *For any society with $|M||W| \geq 2$, there exists a preference profile $P$ such that $S(M, W, P) = \emptyset$.*

**Proof.** Consider a society $N$ with $M = \{m_1, m_2\}$ and $W = \{w_1\}$. The three possible matchings are as listed in Example 2. Let the preference profile be given by

$$
\begin{aligned}
P^{m_1} &= \mu_1\mu_3\mu_2 \\
P^{m_2} &= \mu_3\mu_2\mu_1 \\
P^{w_1} &= \mu_2\mu_1\mu_3.
\end{aligned}
$$

Clearly, no stable matching exists for the above marriage market. By simply renaming $m_1, m_2, w_1$ as $w_1, w_2, m_1$, respectively, in the matchings and the above preference profile, one can obtain a similar argument for the market with $M = \{m_1\}$ and $W = \{w_1, w_2\}$. In each case, the result can be generalized to $|M||W| \geq 2$ as we can extend the preference profile for $N$ to a new profile $\tilde{P}$ for any larger society $\tilde{N} \supset N$ through a *society-respecting extension.* Then, one can easily check that any matching $\mu' \in \mathcal{M}^{\tilde{N}}$ is not stable (being acceptable to none of the new members) if $\mu' \neq E^N[\mu_i]$ for some $i \in \{1, 2, 3\}$, while for any $i \in \{1, 2, 3\}$ matching $E^N[\mu_i]$ is not stable by our previous argument. ∎

Since $C(M, W, P) \subset S(M, W, P)$, a direct corollary of the above theorem is that the existence of the core cannot be guaranteed, either.

When preferences are *independent*, the set of stable matchings always equals the core. We have a similar result with *interdependent* preferences when there are only one man or only one woman in the society.

**Theorem 2.** *For any society with* $min\{|M|, |W|\} = 1$, $S(M, W, P) = C(M, W, P)$ *for all* $P$.

**Proof.** The inclusion $C(M, W, P) \subset S(M, W, P)$ is true by definition. To show the converse, suppose first $S(M, W, P) = \emptyset$. Then, $S(M, W, P) \subset C(M, W, P)$ trivially holds. Next, suppose $S(M, W, P) \neq \emptyset$. For any matching $\mu \in \mathcal{M}^N$ it is true that either every individual is single or there is only one married couple. In both situations, the smallest coalition of individuals that can block $\mu$ contains at most one man and one woman. Therefore, $\mu \in C(M, W, P)$ if $\mu \in S(M, W, P)$. ∎

A distinction between the core and the stable set is found by Sasaki and Toda (1996) in a model with *interdependent* preferences. However, the matchings they consider are bijections and the weaker stability notion they use requires each blocking pair to become better off under all conjectured matchings.[8] The following theorem states that the non-equivalence of the core and stable set also arises in our framework when there are at least two members from each gender in the society.

**Theorem 3.** *For any society with* $min\{|M|, |W|\} \geq 2$*, there exists a preference profile* $P$ *such that* $C(M, W, P) \neq S(M, W, P) \neq \emptyset$.

**Proof.** Consider the market described in Example 1. Recall that $S(M, W, P) = \{\mu_1, \mu_5, \mu_7\} \neq \{\mu_1\} = C(M, W, P)$. Evidently, the result follows for any larger market by a *society-respecting extension* of preferences. ∎

A natural question is whether we can always have a nonempty core or a nonempty stable set on a restricted domain of preferences.

**Definition 9.** For any society $N$ with the preference profile $P$, a nonempty proper subset $\mathcal{V}$ of matchings $\mathcal{M}^N$ is called a *top-matching collection* if for all $i \in N$, $P^i[k] \in \mathcal{V}$ for all $k \in \{1, 2, \ldots, |\mathcal{V}|\}$.

In Example 1, the collections $\{\mu_1\}$, $\{\mu_1, \mu_5, \mu_7\}$, $\{\mu_1, \mu_5, \mu_7, \mu_2\}$, $\{\mu_1, \mu_5, \mu_7, \mu_2, \mu_3\}$, and $\{\mu_1, \mu_5, \mu_7, \mu_2, \mu_3, \mu_4\}$ are all top-matching collections. However, not all of these collections are contained by the stable set or the core.

**Definition 10.** Given a society $N$ and an agent $i \in N$, two matchings $\mu, \mu' \in \mathcal{M}^N$ are called *connected by agent* $i$ if $\mu(i) = \mu'(i)$ and *unconnected by agent* $i$ otherwise.

**Definition 11.** Given a society $N$ and a coalition $T$ of agents in $N$, a matching $\mu'$ is *reachable by* $T$ from another matching $\mu$ if the set of all individuals that connect $\mu$ to $\mu'$ is $N \backslash T$. Let $R(\mu, \mu')$ denote the coalition by which $\mu'$ is reachable from $\mu$.

A matching $\mu$ can be blocked by a coalition $T$ of individuals via some other matching $\mu'$ only if $T = R(\mu, \mu')$, i.e., the coalition $T$ can reach $\mu'$ from $\mu$. Clearly, the converse of the statement is not always true, i.e., a coalition that can reach a matching $\mu'$ from a given matching $\mu$ may not be able to

[8] Sasaki and Toda (1996) obtains this result when the set of conjectured matchings differs from the set of all conceivable matchings.

block $\mu$ via $\mu'$ if the members of the coalition do not agree upon that $\mu'$ is superior.

**Theorem 4.** *Let $N$ be a society with the marriage market $(M, W, P)$ satisfying $|M||W| \geq 2$. Suppose $N$ has a top-matching collection $\mathcal{V}$ with $|\mathcal{V}| \geq 1$ such that whenever $|\mathcal{V}| > 1$ it is true that for any $\mu, \mu' \in \mathcal{V}$, $R(\mu, \mu')$ contains at least one married couple or one unmarried individual under $\mu$, preferring $\mu$ to $\mu'$. Then $C(M, W, P) = \mathcal{V}$.*

**Proof.** No coalition of individuals can block any matching in $\mathcal{V}$ via any other matching in $\mathcal{M}^N \backslash \mathcal{V}$, since $\mathcal{V}$ is a top-matching collection. Moreover, no matching in $\mathcal{M}^N \backslash \mathcal{V}$ can be in the core since it can be blocked by the grand coalition $N$ via any matching in $\mathcal{V}$. Therefore, $C(M, W, P) \subset \mathcal{V}$. It is obvious that $C(M, W, P) = \mathcal{V}$ if $|\mathcal{V}| = 1$. Suppose $|\mathcal{V}| \geq 2$. Consider any two matchings $\mu, \mu' \in \mathcal{V}$. The matching $\mu$ cannnot be blocked by $R(\mu, \mu')$ via $\mu'$, by the assumption that there exists in $R(\mu, \mu')$ a married couple or a single individual under $\mu$, preferring $\mu$ to $\mu'$. Since $\mu$ and $\mu'$ were arbitrary, we have $\mathcal{V} \subset C(M, W, P)$. Together with $C(M, W, P) \subset \mathcal{V}$, this implies $C(M, W, P) = \mathcal{V}$. ∎

Apparently, a singleton top-matching collection, like the set $\{\mu_1\}$ in Example 1, is equal to the core of the marriage market. In the same example, one can check that any top-matching collection that contains $\{\mu_1, \mu_5, \mu_7\}$ (hence differs from the core) does not satisfy the sufficiency condition of the above theorem, since $\mu_1$ dominates any other matching for any individual. One can also verify that the unique top-matching collection $\{\mu_1, \mu_2\}$ in Example 2 satisfies the hypothesis in Theorem 4, hence equals the core.

We should notice that the condition in Theorem 4 that a non-singleton top-matching collection must satisfy in order to be equal to the core of the market requires the existence of either a married couple or a single individual resisting against the formation of a potentially capable coalition to change the current matching. We can surely relax this condition to characterize preference profiles that yield a nonempty stable set, inspiring from the observation that if any two matchings in a top-matching collection are unconnected by a sufficiently large number of individuals (e.g. the collection $\{\mu_1, \mu_5, \mu_7\}$ in Example 1), then no individual or pair can block any matching in the collection even in cases it is to the benefit of all individuals to collectively do so.

**Theorem 5.** *Let $N$ be a society with the marriage market $(M, W, P)$ satisfying $|M||W| \geq 2$. Suppose $N$ has a top-matching collection $\mathcal{V}$ with $|\mathcal{V}| \geq 1$ such that whenever $|\mathcal{V}| > 1$ it is true that for any $\mu, \mu' \in \mathcal{V}$ at least one of the following is met: i) $|R(\mu, \mu')| \leq 4$ and $R(\mu, \mu')$ contains at least one married couple or one unmarried individual under $\mu$, preferring $\mu$ to $\mu'$; ii) $|R(\mu, \mu')| = 4$ and individuals in $R(\mu, \mu')$ are all married or all single under any of the two matchings; iii) $|R(\mu, \mu')| > 4$. Then $\mathcal{V} \subset S(M, W, P)$.*

**Proof.** No pair or singleton can block any matching in $\mathcal{V}$ via any other matching in $\mathcal{M}^N \backslash \mathcal{V}$, since $\mathcal{V}$ is a top-matching collection. It is obvious that $S(M, W, P) \supset \mathcal{V}$ if $|\mathcal{V}| = 1$. Suppose $|\mathcal{V}| \geq 2$. Consider any two matchings $\mu, \mu' \in \mathcal{V}$ that satisfy one of the three conditions in the theorem. First note that a pair of man and woman can block $\mu$ via $\mu'$ only if $|R(\mu, \mu')| \in \{2, 3, 4\}$. So, if condition (iii) holds, $\mu$ cannot be blocked via $\mu'$ by coalitions of size not exceeding two. The conclusion also remains to be true if condition (ii) holds, since one needs the approvals of all four individuals in $R(\mu, \mu')$ to reach $\mu'$ from $\mu$. Finally, the argument is clear if condition (i) is the case. Thus, we have $S(M, W, P) \supset \mathcal{V}$. ∎

One can check in Example 1 that the top-matching collection $\mathcal{V} = \{\mu_1, \mu_5, \mu_7\}$ satisfies condition (ii) (but not the other two conditions) in the above theorem as for any two matchings $\mu, \mu' \in \mathcal{V}$, $|R(\mu, \mu')| = 4$, and all agents are married under $\mu_1$ and $\mu_5$ while all agents are single under $\mu_7$. Hence $\mathcal{V} \subset S(M, W, P)$ becomes true. When $\min\{|M|, |W|\} = 1$, i.e., one of the genders has a unique member in the society, a top-matching collection can only satisfy condition (i) in the above theorem, since any two matchings can be unconnected by at most three agents. We simply observe that condition (i) is met in Example 2 by the unique top-matching collection $\{\mu_1, \mu_2\}$ that has already been checked to satisfy a similar hypothesis in Theorem 4.

Now, we shall consider the relationship between the notions of optimality and stability.

**Definition 12.** For a given marriage market $(M, W, P)$, a matching $\mu$ is *Pareto optimal* if there is no $\mu'$ such that

$$\mu' >_i \mu$$

for all $i \in N$. Let $PO(M, W, P)$ denote the set of all Pareto optimal matchings (the Pareto set).

When preferences are *independent*, any stable matching is also Pareto optimal. But, with *interdependent* preferences, the stable set and the core differ from the Pareto set as stated by the next theorem.

**Theorem 6.** *For any society with $|M||W| \geq 2$, there exists a preference profile $P$ such that $C(M,W,P) \neq \emptyset$, and*

$$PO(M,W,P) \neq S(M,W,P) \text{ and } PO(M,W,P) \neq C(M,W,P).$$

**Proof.** First consider a society with $M = \{m_1, m_2\}$ and $W = \{w_1\}$. The three possible matchings are as defined in Example 2. Let the preference profile $P$ be given by

$$\begin{aligned} P^{m_1} &= \mu_3\mu_1\mu_2 \\ P^{m_2} &= \mu_1\mu_3\mu_2 \\ P^{w_1} &= \mu_1\mu_3\mu_2. \end{aligned}$$

Clearly, $S(M,W,P) = C(M,W,P) = \{\mu_3\}$ whereas $PO(M,W,P) = \{\mu_1, \mu_3\}$. The result symmetrically obtains for a society with $M = \{m_1\}$ and $W = \{w_1, w_2\}$, by changing the identities $m_1, m_2, w_1$ with $w_1, w_2, m_1$ respectively in Example 2. In each case, the result simply generalizes to any larger society by a *society-respecting extension* of preferences. ∎

A similar result as to the distinction between the stable set and the Pareto set is reported by Sasaki and Toda (1996). For societies involving only one agent in one side of the market, we have an immediate Corollary to Theorem 2 stating the optimality of every stable matching.

**Corollary 1.** *For any society with $min\{|M|, |W|\} = 1$, $S(M,W,P) \subset PO(M,W,P)$ for all $P$.*

The above result follows from the obvious fact that the core is always contained by the Pareto set. However, the converse is not always true as implied by Theorem 6. Indeed, the relation between the stable set and the Pareto set is even weaker. Below, we will show that for societies involving at least two members from each gender a nonempty stable set does not always include a Pareto optimal matching. This result is in contrast with Sasaki and Toda (1996) proving that in a model with *interdependent* preferences and bijective matchings there always exists a stable matching that is Pareto optimal.

**Theorem 7.** *For any society with* $min\{|M|, |W|\} \geq 2$*, there exists a preference profile* $P$ *such that* $S(M, W, P) \neq \emptyset$ *and* $PO(M, W, P) \cap S(M, W, P) = \emptyset$*.*

**Proof.** Consider a society with $M = \{m_1, m_2\}$ and $W = \{w_1, w_2\}$. The seven possible matchings are as defined in Example 1. Consider the preference profile given by

$$\begin{aligned} P^{m_1} &= \mu_3\mu_1\mu_4\mu_5\mu_2\mu_6\mu_7 \\ P^{m_2} &= \mu_1\mu_4\mu_5\mu_2\mu_3\mu_6\mu_7 \\ P^{w_1} &= \mu_1\mu_4\mu_5\mu_2\mu_3\mu_6\mu_7 \\ P^{w_2} &= \mu_3\mu_1\mu_4\mu_5\mu_2\mu_6\mu_7. \end{aligned}$$

One can easily check that $S(M, W, P) = \{\mu_4\}$, $(C(M, W, P) = \emptyset)$ and $PO(M, W, P) = \{\mu_1, \mu_3\}$. The result immediately follows for any larger society by a *society-respecting extension* of preferences. ∎

Now, we turn our attention to stable matchings that are optimal for one side of the marriage market.

**Definition 13.** For a given marriage market $(M, W, P)$, a stable matching $\mu$ is $M$*-optimal* if every man likes it at least as well as any other stable matching; i.e.,

$$\mu \geq_M \mu'$$

for every other stable matching $\mu'$. Similarly, a stable matching $\nu$ is $W$-*optimal* if every woman likes it at least as well as any other stable matching; i.e.,

$$\nu \geq_W \nu'$$

for every other stable matching $\nu'$. Let $\mu^M$ and $\mu^W$ denote $M$*-optimal stable matching* and $W$*-optimal stable matching*, respectively. By applying the above definition on the core matchings, one can similarly define $M$*-optimal core matching* $\mu^{M,C}$ and $W$*-optimal core matching* $\mu^{W,C}$.

Gale and Shapley (1962) showed that when all men and women have strict *independent* preferences, there always exist an $M$-optimal stable matching and a $W$-optimal stable matching. Moreover, $\mu^{M,C} = \mu^M$ and $\mu^{W,C} = \mu^W$ since the stable set always equals the core.

In our case, where preferences are *interdependent*, the core does not need to coincide with the stable set; therefore for a given gender (men or women) the optimal matchings over the stable set and the core (whenever they are nonempty) may not be the same. Besides, their existence are not guaranteed, either.

**Theorem 8.** *For any society with $|M||W| \geq 2$ there exists a preference profile $P$ such that $C(M,W,P) \neq \emptyset$ and either $(\mu^M, \mu^{M,C})$ do not exist or $(\mu^W, \mu^{W,C})$ do not exist.*

**Proof.** Consider the marriage market described in Example 2. We have $S(M,W,P) = C(M,W,P) = \{\mu_1, \mu_2\}$. Obviously, $\mu^W = \mu^{W,C} = \mu_1$ while $\mu^M$ and $\mu^{M,C}$ do not exist. By simply renaming $m_1, m_2, w_1$ as $w_1, w_2, m_1$, respectively, in the matchings and the preference profile in Example 2, we can show that for a society with $M = \{m_1\}$ and $W = \{w_1, w_2\}$, we have $\mu^M = \mu^{M,C} = \mu_1$ while $\mu^W$ and $\mu^{W,C}$ do not exist. Both results follow for any larger society using a *society-respecting extension* of preferences. ∎

In fact, one can further show that for societies containing at least two men and two women, there are preference profiles at which neither $M$-optimal matching nor $W$-optimal matching exists in the stable set and in the core.

**Theorem 9.** *For any society with min$\{|M|, |W|\} \geq 2$, there exists a preference profile $P$ such that $C(M,W,P) \neq \emptyset$ and none of $\mu^M, \mu^W, \mu^{M,C}$, and $\mu^{W,C}$ exists.*

**Proof.** Consider a society with $M = \{m_1, m_2\}$ and $W = \{w_1, w_2\}$. The seven possible matchings are as defined in Example 1 and the preference profile is given by

$$
\begin{aligned}
P^{m_1} &= \mu_7\mu_1\mu_5\mu_2\mu_3\mu_4\mu_6 \\
P^{m_2} &= \mu_1\mu_5\mu_7\mu_2\mu_3\mu_4\mu_6 \\
P^{w_1} &= \mu_7\mu_5\mu_1\mu_2\mu_3\mu_4\mu_6 \\
P^{w_2} &= \mu_1\mu_5\mu_7\mu_2\mu_3\mu_4\mu_6.
\end{aligned}
$$

It is easy to check that $S(M, W, P) = C(M, W, P) = \{\mu_1, \mu_5, \mu_7\}$ while none of $\mu^M, \mu^W, \mu^{M,C}$, and $\mu^{W,C}$ exists. The result immediately follows for any larger society using a *society-respecting extension* of preferences. ∎

A theorem by Knuth (1976) states that when all agents have strict *independent* preferences, the common preferences of the two sides of the market are opposed on the set of stable matchings: if $\mu$ and $\mu'$ are stable matchings, then all men like $\mu$ at least as well as $\mu'$ if and only if all women like $\mu'$ at least as well as $\mu$. That is, $\mu >_M \mu'$ if and only if $\mu' >_W \mu$. But, with *interdependent* preferences this result does not prevail any longer for matchings either in the stable set or in the core.

**Theorem 10.** *For any society with $|M||W| \geq 2$, there exists a preference profile $P$ and some $\mu, \mu' \in C(M, W, P)$ such that $\mu >_M \mu'$ is not true while $\mu' >_W \mu$.*

**Proof.** Consider the marriage market described in Example 2. Recall that $C(M, W, P) = \{\mu_1, \mu_2\}$. We have $\mu_1 >_W \mu_2$ while it is not true that $\mu_2 >_M \mu_1$. The result immediately follows for any larger market using a *society-respecting extension* of preferences. ∎

Apparently, Theorem 10 can symmetrically be restated by interchanging men and women. For markets with strict *independent* preferences, a theorem (McVitie and Wilson 1970, Roth 1984) benefiting from the opposition of the common preferences of the two sides of the market proposes that the set of people who are matched is the same for all stable matchings. For strict and *interdependent* preferences, a similar statement is no longer true.

**Theorem 11.** *For any society with $|M||W| \geq 2$, there exists a preference profile $P$ such that there are at least two matchings in the core at which the set of people who are matched is not the same.*

**Proof.** Consider the society described in Example 2. The unique married pair is $(m_1, w_1)$ at the core matching $\mu_1$ whereas $(m_2, w_1)$ at the core matching $\mu_2$. The result immediately follows for any larger society using a *society-respecting extension* of preferences. ∎

The above negative result holding for the core trivially applies for the stable set, as well. Next, we study whether the stable set or the core, whenever nonempty, has an algebraic structure, called *lattice*. We borrow the following definition from Roth and Sotomayor (1990).

**Definition 14.** A *lattice* is a partially ordered set $L$ any two of whose elements $x$ and $y$ have a "sup", denoted by $x \vee y$ and an "inf", denoted by $x \wedge y$.[9]

For any two matchings $\mu$ and $\mu'$, and for all men, define $\mu \vee_M \mu'$ (join) as the function that assigns each man to his mate at the most preferred of the two matchings, and $\mu \wedge_M \mu'$ (meet) as the function that assigns each man to his mate at the least preferred of the two matchings:

$$\mu \vee_M \mu'(m) = \begin{cases} \mu(m) & \text{if } \mu >_m \mu' \\ \mu'(m) & \text{otherwise} \end{cases}$$

and

$$\mu \wedge_M \mu'(m) = \begin{cases} \mu'(m) & \text{if } \mu >_m \mu' \\ \mu(m) & \text{otherwise} \end{cases}$$

for all $m \in M$. Define $\mu \vee_W \mu'$ and $\mu \wedge_W \mu'$ analogously.

The well-known lattice theorem (Conway) states that when all preferences are strict and independent, the set of stable matchings is a (distributive) lattice under the common preference order of the men, dual to the common preference order of the women. When preferences are *interdependent*, the core (hence the stable set) does not always have a lattice structure with respect to the common ordering of men or women.

**Theorem 12.** *For any society with $|M||W| \geq 2$ and for any $X \in \{M, W\}$, there exists a preference profile $P$ such that there are two matchings $\mu, \mu' \in C(M, W, P)$ for which the functions $\mu \vee_X \mu'$ and $\mu \wedge_X \mu'$ are not stable matchings.*

**Proof.** Consider first the society described in Example 2. For the two core matchings $\mu_1$ and $\mu_2$, $\mu_1 \vee_M \mu_2$ is not a stable matching since $\mu_1 \vee_M \mu_2(m_1) = m_1$ and $\mu_1 \vee_M \mu_2(m_2) = m_2$. On the other hand, $\mu_1 \wedge_M \mu_2$ is not a matching since $\mu_1 \wedge_M \mu_2(m_1) = w_1$ and $\mu_1 \wedge_M \mu_2(m_2) = w_1$ are contradictory. Now

[9] A binary relation is a *partial order* if it is reflexive, transitive, and antisymmetric. An order relation $R$ is transitive if $xRy$ and $yRz$ implies $xRz$; reflexive if $xRx$; antisymmetric if $xRy$ and $yRx$ implies $x = y$.

consider a society with $M = m_1$ and $W = \{w_1, w_2\}$. Let the associated matchings and the preference profile be obtained by changing the identities $m_1, m_2, w_1$ with $w_1, w_2, m_1$ respectively in Example 2. The core remains to contain $\mu_1$ and $\mu_2$, and it is obvious from our previous arguments that $\mu_1 \vee_W \mu_2$ and $\mu_1 \wedge_W \mu_2$ are not stable matchings. The result immediately follows for any larger society with $\min\{|M|, |W|\} \geq 2$ using a *society-respecting extension* of preferences. ∎

Now, we turn to some strategic issues. Specifically, we shall study the extent to which agents will be sincere about their preferences for possible matchings. Consider a marriage market where the matching of individuals is determined by a centralized clearinghouse, based on a list of preferences that individuals state. A *mechanism*, $\Gamma$, is a procedure which determines a matching for each marriage market $(M, W, P)$. If the list of preferences reported by the individuals is $Q$, the mechanism produces a matching $\Gamma[Q]$. If $\Gamma[Q]$ is always stable with respect to $Q$, the mechanism $\Gamma$ is said to be a *stable matching mechanism*. Moreover, if $\Gamma[Q]$ is always in the core with respect to $Q$, the mechanism $\Gamma$ is said to be a *core mechanism*.

**Definition 15.** A mechanism $\Gamma$ is *strategy-proof* if for all $P$ and for every $i \in N$,

$$\Gamma[P] \geq_i \Gamma[\hat{P}^i, P^{-i}]$$

for all $\hat{P}^i$.

In his seminal paper, Roth (1982) shows that when preferences are strict and independent, there is no stable matching mechanism (core mechanism) which is strategy-proof. The impossibility to design a mechanism that produces core matchings or stable matchings in terms of the reported preferences and that makes the truthful reporting a dominant strategy for every agent, remains to exist when preferences are *interdependent*, as well.

**Theorem 13.** *For any society satisfying $|M||W| \geq 2$ and having strict and interdependent preferences, there is no core mechanism which is strategy-proof.*

**Proof.** Consider the marriage market decribed in Example 2. We have $C(M, W, P) = \{\mu_1, \mu_2\}$. So, any stable mechanism must choose $\mu_1$ or $\mu_2$

when the preference report is $P$. Suppose the mechanism chooses $\mu_1$. If $m_1$ misreports his preference ordering as $Q^{m_1} = \mu_2\mu_3\mu_1$ while everyone else makes truthful revelations, then at the reported profile $(Q^{m_1}, P^{m_2}, P^{w_1})$, $\mu_2$ becomes the unique matching in the core, which is preferred by $m_1$ to $\mu_1$ at his true preferences $P^{m_1}$. So, it is not a dominant strategy for all agents to truthfully reveal their preferences. Similarly, if the mechanism chooses $\mu_2$ when the preferences $P$ are reported, then $m_2$ can profitably misrepresent his preferences as $Q^{m_2} = \mu_1\mu_3\mu_2$ to force the mechanism to select his more preferred core matching $\mu_1$. The result immediately follows for any larger society using a *society-respecting extension* of preferences. ∎

The below result, which strengthens our impossibility theorem, characterizes some restrictions on preferences under which at least one agent will behave strategically when facing a core mechanism. Obviously, both this result and the previous one hold for stable matching mechanisms, as well.

**Theorem 14.** *Consider any marriage market in which preferences are strict and interdependent, and assume that (i) there is more than one stable matching in the core and every agent is married at any core matching, (ii) every matching outside the core in which at least one pair is married is unacceptable to at least one agent, and (iii) the matching at which every agent is single is bottom ranked by at least two agents. When any core mechanism is applied to this market, then at least one agent can profitably misrepresent his or her preferences, assuming the others tell the truth.*

**Proof.** Consider a marriage market $(M, W, P)$ satisfying the hypotheses of the theorem. Let the core of this market, $C(M, W, P)$, contain at least two stable matchings by assumption. Suppose that when all agents reveal their true preferences, the core mechanism $\Gamma$ selects matching $\nu$. Let $i$ be any agent who does not top rank $\nu$ in his or her preference ordering. There exists such an agent, for otherwise no matching other than $\nu$ could be in the core. Find $l = \min\{r \mid P^i[r] \in C(M, W, P)\}$, the index of the core matching that agent $i$ prefers most. It must be true that $\mu_{i,i} \neq \mu$ for any $\mu \in C(M, W, P)$ since $i$ is not single under any core matching, by assumption (i). Now, let $i$ misrepresent his or her preferences as $\hat{P}^i$ by top-ranking $P^i[l]$ while ranking $\mu_{i,i}$ above $\mu$ for any other $\mu \in C(M, W, P)$, and keeping the position of the bottom-ranked element of $P^i$ as the same in $\hat{P}^i$. Let other agents truthfully represent their preferences. Denote the new profile by $\hat{P} = (\hat{P}^i, P^{-i})$. Then, no matching $\mu \in C(M,W,P)\backslash\{P^i[l]\}$ can be in $C(M,W,\hat{P})$, since any such matching is unacceptable to agent $i$ under $\hat{P}^i$, by construction. Let $\mu^s$ denote the matching under which every agent is single. Any $\mu \notin C(M,W,P) \cup \{\mu^s\}$ cannot be in $C(M,W,\hat{P})$ as it is unacceptable to at least one agent, by assumption (ii). Finally, $\mu^s \notin C(M,W,\hat{P})$ since there is at least one pair ready to block it, by assumption (iii). Hence, $\Gamma(\hat{P}) = C(M,W,\hat{P}) = \{P^i[r]\}$, which makes agent $i$ better-off. ∎

A result by Demange, Gale, and Sotomayor (1987) shows that when preferences are strict and *independent* no coalition of men and women can manipulate their preferences so successfully that every member of the coalition prefers one of the new outcomes to every stable outcome (with respect to the true preferences). Below, we prove that such limits on successful manipulation do not exist when preferences are *interdependent*.

**Theorem 15.** *For any society with $min\{|M|,|W|\} \geq 2$, there exist two preference profiles $P$ and $\bar{P}$, where $\bar{P}$ differs from $P$ for some coalition $C$ of men and women, such that there exists a matching $\mu$ in the core (stable set) for $\bar{P}$, which is preferred to every core (stable) matching under the preference $P$ by all members of $C$.*

**Proof.** Consider a society with $M = \{m_1, m_2\}$ and $W = \{w_1, w_2\}$. The seven possible matchings are as defined in Example 1 and the preference profile is given by

$$\begin{aligned}
P^{m_1} &= \mu_3\mu_1\mu_4\mu_5\mu_2\mu_6\mu_7 \\
P^{m_2} &= \mu_1\mu_4\mu_5\mu_2\mu_3\mu_6\mu_7 \\
P^{w_1} &= \mu_1\mu_4\mu_5\mu_2\mu_3\mu_6\mu_7 \\
P^{w_2} &= \mu_3\mu_4\mu_1\mu_5\mu_2\mu_6\mu_7.
\end{aligned}$$

It is easy to check that $S(M,W,P) = C(M,W,P) = \{\mu_4\}$. Now suppose that the singleton coalition $\{m_1\}$ misrepresents his preference ordering as

$$\bar{P}^{m_1} = \mu_1\mu_3\mu_4\mu_5\mu_2\mu_6\mu_7.$$

Define $\bar{P} = (\bar{P}^{m_1}, P^{-m_1})$. We then have $S(M,W,\bar{P}) = C(M,W,\bar{P}) = \{\mu_1,\mu_4\}$. Clearly, $m_1$ prefers $\mu_1 \in C(M,W,\bar{P})$ to $\mu_4$, the unique matching in $C(M,W,P)$. The result immediately follows for any larger society using a *society-respecting extension* of preferences. ∎

When one of the genders has a unique representative in the society, we have a weaker result.

**Theorem 16.** *For any society with* $min\{|M|, |W|\} = 1$*, there exist two preference profiles* $P$ *and* $\bar{P}$*, where* $\bar{P}$ *differs from* $P$ *for some coalition* $C$ *of men and women, such that there exists a matching* $\mu$ *in the core (stable set) for* $\bar{P}$*, which is weakly preferred to every core (stable) matching under the preference* $P$ *by all members of* $C$*.*

**Proof.** The proof is obvious from the proof of Theorem 13, which shows the existence of a preference profile under which one of the agents in the society can successfully manipulate his or her preference to make the mechanism outcome uniquely select his most preferred core matching. ∎

So far, we have dealt with whether the set of stable matchings and the core for any given society have to exist, whether they must coincide whenever they both exist, and whether they satisfy some desirable properties such as optimality, gender optimality, and strategy proofness or have some useful structures, such as lattices. Now we shall try to get beyond these traditionally investigated aspects of any meaningful matching model, by asking an interesting existential question that has very recently been raised and answered by Echenique (2006) in the framework of *independent* preferences: "Can there be any set of matchings for a given society that is incompatible with the predictions of our matching model with respect to the employed stability notions?" As Echenique (2006) points out the answer to this question is important when the preferences are unknown as it allows one to know whether a matching theory at hand has testable implications. Following his treatment, we define the *rationalizability* of any set of admissible matchings for the stable set and for the core.

**Definition 16.** For a given society $N = M \cup W$, let $\mathcal{H} \subset \mathcal{M}^N$ be a subset of available matchings. We say that $\mathcal{H}$ is *rationalizable for the stable set* if there exists a preference profile $P$ such that $\mathcal{H} \subset S(M, W, P)$. Similarly, we say that $\mathcal{H}$ is *rationalizable for the core* if there exists a preference profile $P$ such that $\mathcal{H} \subset C(M, W, P)$.

We simply note that a set $\mathcal{H} \subset \mathcal{M}^N$ is rationalizable for the core only if it is rationalizable for the stable set. Echenique (2006) shows that under *independent* preferences $\mathcal{M}^N$ is not rationalizable for the stable set (equalling the core) if the number of men and the number of women are the same and at least three. We extend this result in the following way.

**Theorem 17.** *For any society $N$ satisfying $|M||W| \geq 2$ and having strict and interdependent preferences, $\mathcal{M}^N$ is not rationalizable for the stable set (and the core).*

**Proof.** Suppose, $\mathcal{M}^N$ is rationalizable for the stable set by some preference profile $P$; i.e., $\mathcal{M}^N \subset S(M, W, P)$. Let $\mu^s$ denote the matching at which every agent is single. Pick any $(m, w) \in M \times W$. Denote by $\mu^s_{m,w}$ the matching at which $(m, w)$ is the unique married couple. Then, $\mu^s_{m,w} >_m \mu^s$ and $\mu^s_{m,w} >_w \mu^s$ by the assumed stability of $\mu^s_{m,w}$. This implies that $\mu^s$ cannot be in $S(M, W, P)$, a contradiction. ∎

Theorem 17 shows that the whole set of matchings cannot be rationalizable, hence our matching model is testable. Given the refutability of our model, the next step is to check whether any proper subset of $\mathcal{M}^N$ can be rationalizable. When the preferences are *independent*, Echenique (2006) is able to show that any set of matchings in which no agent is matched with the same partner under different matchings is rationalizable. He also shows that in general the preferences that rationalize a rationalizable set of matchings are not unique. The existence of rationalizable collection of matchings must not be suprising in our framework either, given the positive results of Theorems 4 and 5 that characterize certain restrictions on preferences that yield a nonempty core or a nonempty stable set for a given marriage market. However, as the complete characterization of rationalizable sets is beyond the scope of this paper, we will here present a simple sufficiency theorem.

**Theorem 18.** *For any society $N$ having strict and interdependent preferences, consider $\mathcal{H} \subset \mathcal{M}^N$ such that no pair of matchings $\mu_k, \mu_l \in \mathcal{H}$ are connected by any agent in $N$. Then $\mathcal{H}$ is rationalizable for the stable set and there exist at least*

$$\left(|\mathcal{H}|\,!\,(|\mathcal{M}^N| - |\mathcal{H}|)\,!\right)^N$$

*distinct preference profiles that rationalize it; moreover if $|\mathcal{H}| \leq N$, then $\mathcal{H}$*

*is rationalizable for the core and there exist at least*

$$\binom{N}{|\mathcal{H}|}|\mathcal{H}|\,!\,(|\mathcal{H}|\,!)^{(N-|\mathcal{H}|)}\left((|\mathcal{M}^N|-|\mathcal{H}|)\,!\right)^N$$

*distinct preference profiles that rationalize it.*

**Proof.** Consider any society $N$ having strict and *interdependent* preferences. Pick any $\mathcal{H}=\{\mu_0,\mu_1,\mu_2,\ldots,\mu_Z\}\subset\mathcal{M}^N$ for some $Z\in\{1,2,\ldots,|\mathcal{M}^N|-2\}$ such that no pair of matchings $\mu_k,\mu_l\in\mathcal{H}$ are connected by any agent in $N$. Consider first the preference profile $P$ such that for all $i\in N$, $P^i[k]=\mu_{k-1}$ for all $k\in\{1,2,\ldots,|\mathcal{H}|\}$. Then, it is easy to check that $\mathcal{H}\subset S(M,W,P)$. Each individual in $N$ can independently order the first $k$ matchings in $|\mathcal{H}|\,!$ distinct ways while he or she can order the remaining matchings in $(|\mathcal{M}^N|-|\mathcal{H}|)\,!$ distinct ways. Hence follows the lower bound on the number of preferences that rationalize $\mathcal{H}$ for the stable set. To prove the second part of the theorem, let $|\mathcal{H}|\le N$. Enumerate agents from 1 to $N$. Consider the preference profile $P$ such that $P^i[k]=\mu_l$ with $l=(k+i-2)\mod{}_{|\mathcal{H}|}$ for all $i\in\{1,\ldots,|\mathcal{H}|\}$ and for all $k\in\{1,2,\ldots,|\mathcal{H}|\}$ whereas $P^i[k]\in\mathcal{H}$ for all $i\in\{|\mathcal{H}|+1,\ldots,N\}$ and for all $k\in\{1,2,\ldots,|\mathcal{H}|\}$ with each $P^i[k]$ being distinct. Then, it is easy to check that $\mathcal{H}\subset C(M,W,P)$. Notice that there are $\binom{N}{|\mathcal{H}|}$ distinct ways to select $|\mathcal{H}|$ agents from the society. The first $|\mathcal{H}|$ matchings in the preference orderings of the first $|\mathcal{H}|$ agents are completely tied to each other, so there are $|\mathcal{H}|\,!$ distinct ways to represent their preference ordering as a group. Each of the remaining $N-|\mathcal{H}|$ agent can independently have any of $|\mathcal{H}|\,!$ distinct orderings of the first $|\mathcal{H}|$ matchings drawn from $\mathcal{H}$. Besides, any agent in $N$ can independently order the remaining $(|\mathcal{M}^N|-|\mathcal{H}|)$ matchings in $(|\mathcal{M}^N|-|\mathcal{H}|)\,!$ distinct ways. ∎

In Example 2, the sets $\mathcal{H}_1=\{\mu_1\}$, $\mathcal{H}_2=\{\mu_2\}$, $\mathcal{H}_3=\{\mu_3\}$, and $\mathcal{H}_4=\{\mu_1,\mu_2\}$ all satisfy the connectedness hypothesis in the above theorem. Since, $|\mathcal{H}_k|\le 3=N$ for all $k\in\{1,2,3,4\}$, any $\mathcal{H}_k$ is rationalizable for the core (hence for the stable set). Moreover, one can easily calculate that the set $\mathcal{H}_1$, contained by the core, can be rationalized for the stable set by at least 8 distinct preference profiles and for the core by at least 24 distinct preference profiles.

Theorem 18 shows that if a collection of matchings, such as the set of all matchings, is not rationalizable for the stable set (or the core even when the number of matchings in the collection is less than the number of agents in the society), then some agents must have the same mate under more than one matching. We should here emphasize that this result simply characterizes collections of matchings that are not rationalizable, However, a sufficiency result such as Theorem 18 is still valuable, as already remarked by Echenique (2006) in his framework of *independent preferences*, since it has an important implication for empirical tests of the matching theory at hand, requiring some pairs of agents to be identified under more than one matching in the available data. On the other hand, Theorem 18 also implies that our refutable matching model of *interdependent* preferences is not exactly *identifiable*, as there may exist many different preference profiles that rationalize some sets of matchings.

# 4 Conclusions

In this paper, we introduce *interdependent* pereferences to a classical one-to-one matching problem that allows for the prospect of being single and study the existence and properties of stable matchings. The results that we obtain in this paper, along with some previous findings of Sasaki and Toda (1996) for the case of bijective matchings, clearly demonstrate that the tools and results of the classical theory of matching, built under the assumption of *independent* preferences, may be of little use in models involving externalities. In fact, this observation implies that one should be extremely careful in applying the classical theory to solve some real matching problems. The assumption of *independent* preferences, which can harmlessly be made in designing a market for kidney exchange, may, on the other hand, be extremely inappropriate in modeling marriage and divorce in a given society, the school choices of families for their kids or the firm choices of couple workers.

# References

Charness, G. and Rabin, M. "Understanding Social Preferences with Simple Tests," *Quarterly Journal of Economics*, 2002, 117(3), pp. 817-869.

Demange, G., Gale, D. and Sotomayor, M. "A Further Note on the Stable Matching Problem," *Discrete Applied Mathematics*, 1987, 16(3), pp. 217-222.

Echenique, F. and Yenmez, B. "A Solution to Matching with Preferences over Colleagues," Games and Economic Behavior, 2006 *forthcoming.*

Echenique, F. "What Matchings Can Be Stable? The Testable Implications of Matching Theory," Caltech SS Working Paper 1252.

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

Gul, F. and Pesendorfer, W. "The Cannonical Type Space for Interdependent Preferences," 2005, mimeo, Princeton University.

Hafalir, I. E. "Stability of Marriage with Externalities," 2006, mimeo, Penn State University.

Klaus, B. and Klijn, F. "Stable Matchings and Preferences of Couples," *Journal of Economic Theory*, 2005, 121(1), pp. 75-106.

Knuth, D. E. *Marriages stable.* Montreal: Les Presses de l'Universite de Montreal, 1976.

Li, J. "The Power of Conventions: A Theory of Social Preferences," 2005, mimeo, University of Pennsylvania.

McVitie, D. G. and Wilson, L. B. "Stable Marriage Assignments for Unequal Sets," *BIT*, 1970(10), pp. 295-309.

Ok, E. and Kockesen, L. "Negatively Interdependent Preferences," *Social Choice and Welfare*, 2000, 17(3), pp. 533-558.

Pollak, R.A. "Interdependent Preferences," *American Economic Review*, 1976, 66(3), pp. 309-320.

Postlewaite, A. "The Social Basis of Interdependent Preferences," *European Economic Review*, 1998, 42(3-5), pp. 779-800.

Roth, A. E. "The Economics of Matching: Stability and Incentives," *Mathematics of Operations Research*, 1982, 7(4), pp. 617-628.

Roth, A. E. "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory," *Journal of Political Economy*, 1984, 92(6), pp. 991-1016.

Roth, A. E. and Sotomayor, M. *Two-sided matching: A study in game theoretic modeling and analysis.* Cambridge University Press, Cambridge, UK, 1990.

Sasaki, H. and Toda, M. "Two-Sided Matching Problems with Externalities," *Journal of Economic Theory*, 1996, 70(1), pp. 93-108.

Sobel, J. "Interdependent Preferences and Reciprocity," *Journal of Economic Literature*, 2005, 43(2), pp. 392-436.