JOURNAL OF ECONOMIC THEORY 66, 352–369 (1995)

Stable Matchings and Rematching-Proof Equilibria in a Two-Sided Matching Market*

JINPENG MA

Department of Economics, Faculty of Arts and Sciences, Rutgers University, Camden, New Jersey 08102

Received October 30, 1992; revised July 6, 1994

In this paper we introduce the notion of a rematching-proof equilibrium for a two-sided matching market to resolve Roth's open question: What kind of equilibria of the game induced by any stable mechanism with respect to misreported profiles produce matchings that are stable with respect to the true profile. We show that the outcome of a rematching-proof equilibrium is stable with respect to the *true* profile, even though the equilibrium profile may contain *misreported* preferences. We also show that a rematching-proof equilibrium exists. Moreover, the Nash equilibria in Roth (*J. Econ. Theory* **34** (1984), 383–387) are shown to be rematching-proof equilibria. *Journal of Economic Literature* Classification Numbers: C78, D71.

1. INTRODUCTION

In a typical matching market, there are two disjoint sets of players: a set of men and a set of women. Each man (woman) has strict preferences over the women (men). In these preferences, a player may prefer to be single rather than marry someone of his or her opposite sex. A collection of preferences of all the men and the women is called a profile. Clearly, a matching market can be identified with its profile.

A matching marries[1] men and women to each other, possibly leaving some of them single. Such a matching is called stable with respect to a profile, if (a) no man (woman) prefers being single to his (her) spouse, and if (b) one cannot find a man and a woman, who are not married to each other, but who prefer each other to their spouses.

* I am grateful to an anonymous referee for many fruitful comments. I also thank Robert Aumann, Pradeep Dubey, John Hillas, Abraham Neyman and Alvin E. Roth for their comments, Barry Feldman, Dobrin Kolev, and Raman Murty Mandapaka for their help in exposition. The support of NSF Grant 8922610 from the Institute for Decision Sciences at Stony Brook and of the fellowship from the Center for Rationality and Interactive Decision Theory at the Hebrew University of Jerusalem is gratefully acknowledged.

[1] Neither polygamy nor polyandry is allowed.

0022-0531/95 $12.00

A map from profiles to matchings is called a (matching) mechanism. A mechanism is called stable if the matching it assigns to any profile is stable with respect to that profile. As Gale and Shapley [5] show, a stable mechanism exists (see also Roth and Sotomayor [14]). Indeed, they describe a men-proposing deferred acceptance (MDA) algorithm which explicitly constructs a stable matching for any profile.

Once any mechanism is instituted, a classical incentive problem arises. Could it be that players may sometimes have incentive to strategically misrepresent their true preferences? This motivates the search for "strategy-proof" mechanisms. A given mechanism is *strategy-proof* for an individual if there does not exist any profile in which the individual obtains a better partner by unilaterally misrepresenting his (or her) preference rather than revealing the truth. A mechanism is called strategy-proof if it is strategy-proof for all players.

Roth's impossibility result [10] is that no stable mechanism is also strategy-proof.[2] In Section 3, we show that this negative conclusion can be somewhat sharpened. For any stable mechanism, the set of individuals for whom the mechanism is strategy-proof will never contain both men and women. Thus, the MDA and the WDA (where W stands for women) mechanisms, which are known to be strategy-proof for men and women respectively, go as far as possible in the direction of strategy-proofness.

We now turn to Nash equilibria in order to capture the behavior of strategically free players. Thus, given a stable mechanism, consider the following induced game. Each player reports a preference and obtains the mate that the mechanism assigns on the basis of the reported profile. The question is: are Nash equilibrium outcomes stable with respect to the true profile (even if players misrepresent the truth in the Nash equilibrium profiles)? It is easy to see that the answer is negative (see Examples 8 and 9, and Section 5). This leads to the second question, first articulated in Roth [13, p. 245]:[3] is there a meaningful refinement of the Nash equilibria which will yield a positive result?

To answer the question, we define the notion of a rematching-proof equilibrium. This is similar to a strong Nash equilibrium (see Aumann [2]), except that we restrict coalitions to couples (i.e., one man and one woman). In other words, a rematching-proof equilibrium is a Nash equilibrium in which no pair of a man and a woman can jointly change their reported

[2] Note that a strategy-proof mechanism exists. Such a mechanism will surely produce matchings that are often unstable with respect to the true profile. It was observed in reality that a strategy-proof mechanism might have very unsatisfactory performances; see Example 4.3 in Roth and Sotomayor [14]. Roth's [9, 12] noted empirical work also showed that if a mechanism is not a stable one, there will exist some market failures associated with such a mechanism.

[3] See also open problem 5 in Roth and Sotomayor [14, p. 246].

preferences (keeping other players' reports fixed) so as to each be better off. If a Nash equilibrium is not rematching-proof, then there exists at least one couple who can make each better off by coordinating their behavior. Therefore such a Nash equilibrium is not so convincing because a couple may well coordinate their behavior if coordination can indeed make both better off. The power of the rematching-proof equilibrium is to exclude exactly this possibility. Our main results show that, given any stable mechanism, there exist rematching-proof equilibria of the induced game, and that the equilibria outcomes are stable with respect to the underlying true profile. Moreover, the Nash equilibria constructed in Roth [11] for the MDA and WDA mechanisms are shown to be rematching-proof.

The model investigated in this paper has been extensively studied in the literature since Gale and Shapley [5] published their distinguished work in 1962. A comprehensive account of this literature can be found in Roth and Sotomayor [14]. The issues addressed here are motivated by Roth [11, 13]. They are also related to Roth and Vande Vate [16] in which the truncated strategy equilibria were defined. Roth and Vande Vate [16] showed that unstable matchings could arise in their equilibria (see Section 5, and Example 9 in Section 4); they resolved this negative result by using a dynamic matching procedure, which is quite distinct from our approach.

The paper is structured as follows. Section 2 introduces some notation and definitions. Section 3 discusses both the impossibility theorems of Roth *et al.* and the relation of this work to the impossibility theorems in the social choice literature. Section 4 studies Nash equilibria and stable matchings. Section 5 shows the main results and Section 6 concludes the paper.

2. Notation and Definitions

There are two disjoint sets: a set $M=\{m_1, ..., m_n\}$ of men and a set $W=\{w_1, ..., w_r\}$ of women. Assume that both n and r are larger than 1 for nontriviality. Denote $N=M \cup W$. Let Ω_m (Ω_w) denote the set of all strict orderings over the set $W \cup \{m\}$ ($M \cup \{w\}$) for $m \in M$ ($w \in W$). Thus an element in Ω_m (Ω_w) defines a strict preference P_m (P_w) for a man m (woman w). For example, $P_m=(w_2, w_1, m, w_3, \cdots)$ means that the man m prefers w_2 to w_1, written $w_2 P_m w_1$, and prefers w_1 to remaining single, and prefers remaining single to w_3, and so forth. Also let $\Omega=\prod_{m_i \in M} \Omega_{m_i} \times \prod_{w_j \in W} \Omega_{w_j}$. Thus an element P in Ω defines a (preference) profile $P=(P_{m_1}, ..., P_{m_n}; P_{w_1}, ..., P_{w_r})$. Once N is given, a matching market is completely determined by a profile.

The map $\mu: N \to N$ is called a *matching* if $\mu^2(k)=k$ for all $k \in N$, $\mu(m) \in W$ if $\mu(m) \neq m$ for all $m \in M$, and $\mu(w) \in M$ if $\mu(w) \neq w$ for all $w \in W$. The set of all matchings is denoted by $\mathcal{M}$. Given a profile P, a matching

μ is called (a) *individually rational* if $\neg mP_m\mu(m)$ for all $m \in M$ and $\neg wP_w\mu(w)$ for all $w \in W$; (b) *pairwise stable* if $\nexists(m, w) \in M \times W$ such that $wP_m\mu(m)$ and $mP_w\mu(w)$; and (c) *stable* if it is both individually rational and pairwise stable. The sets of all matchings that are individually rational, pairwise stable, and stable with respect to the profile P are denoted by $IR(P)$, $PS(P)$ and $S(P)$ respectively. Thus by definition, $S(P) = IR(P) \cap PS(P)$.

A player i in N with P_i is said to like a matching μ at least as much as the matching ν if and only if $\neg\nu(i)\, P_i\mu(i)$. Given a profile P, a stable matching in $S(P)$ is M-optimal (W-optimal) if every man m with P_m (woman w with P_w) likes it at least as much as any other matching in $S(P)$.

Gale and Shapley [5] showed that for any profile P in Ω, the set $S(P)$ of stable matchings is nonempty and there exists a unique M-optimal (W-optimal) stable matching in $S(P)$.

A (matching) mechanism is a map $\varphi: \Omega \to \mathcal{M}$ from profiles in Ω to matchings in $\mathcal{M}$. Denote the set of all mechanisms by Ω^*. Let S^* denote the set of all stable mechanisms. Formally, $S^* = \{\varphi \in \Omega^* \mid \varphi(Q) \in S(Q)$ for all $Q \in \Omega\}$. For example, if φ^M (φ^W) is the MDA (WDA) algorithm, then φ^M and φ^W are two elements in S^*.

A stable mechanism φ in S^* induces a game form where each player $i \in N$ reports a (strict) preference Q_i in Ω_i as a strategy[4] and the resulting outcome is the matching $\varphi(Q)$ assigned by the mechanism φ to the reported profile Q. Given an underlying true profile P and a stable mechanism φ, we have an induced game $\Gamma(\varphi, P)$, which is the auxiliary game studied in this paper. In fact, the auxiliary game itself is not of direct interest since the underlying true profile P is not known to us. But by means of this auxiliary game, we will define a few useful notions that may guide us in the implementation of the mechanism φ.

3. Impossibility Theorems

This section details the impossibility theorems in Roth and Sotomayor [14] for completeness. Some additional interpretations (Propositions 3 and 5) are provided. They are only implicitly included in the proof of Theorem 4.6 in Roth and Sotomayor [14]. To our best knowledge, explicit discussions are not available. Propositions 1 to 5 provide motivations for the next two sections.

We start with some additional notation. Recall $N = M \cup W$. Denote the set of all nonempty subsets of N by $\mathcal{T}$. For $T \in \mathcal{T}$, let $\Omega^T = \prod_{i \in T} \Omega_i$. Define $Q_T = (Q_i)_{i \in T} \in \Omega^T$ and $Q_{-T} = (Q_v)_{v \in N \setminus T}$; thus $(Q_{-T}, Q_T) \in \Omega$. If

[4] Players are allowed to choose only pure strategies here; an assumption which is common for the literature on matching models.

$T=\{k\}$ is a one-point set, Q_{-T} and Q_T will be denoted Q_{-k} and Q_k respectively. For a matching mechanism $\varphi: \Omega \to \mathcal{M}$, $Q \in \Omega$, and $k \in N$, define $\varphi_k(Q) = \varphi(Q)(k)$ as the mate that k receives in the induced game when preferences Q are reported to the mechanism φ.

Next we define a subclass $\Phi(k, P)$ of stable mechanisms S^* in which it is a dominant strategy for player k to reveal his or her true preference P_k at profile P no matter what others report, that is, for all Q_{-k}. Thus by definition, an element in $\bigcap_{P \in \Omega} \Phi(k, P)$ represents a stable mechanism that is strategy-proof for player k, that is, it is always in player k's best interest to report his or her true preference P_k for any profile P once a stable mechanism in $\bigcap_{P \in \Omega} \Phi(k, P)$ is used. Formally, $\Phi(k, P)$ is given by

$$\Phi(k, P) = \{\varphi \in S^* \mid \neg \varphi_k(Q_{-k}, Q_k)\, P_k\, \varphi_k(Q_{-k}, P_k) \text{ for all } Q \in \Omega\}.$$

For $T \in \mathcal{T}$, define $\Phi(T, P) = \bigcap_{k \in T} \Phi(k, P)$ to be the subset of S^* such that each stable mechanism in $\Phi(T, P)$ induces each player in T to reveal his or her true preference at profile P no matter what the other players report. Similarly, an element in $\bigcap_{P \in \Omega} \Phi(T, P)$ represents a stable mechanism that is strategy-proof for all players in T. Therefore a stable mechanism is strategy-proof if it is strategy-proof for all players in N.

We know that a stable mechanism always exists. It may be asked whether there exists a stable mechanism that is also strategy-proof. Roth [10], via an example, gave a negative answer. Proposition 2 is a generalization of Proposition 1.

PROPOSITION 1 (ROTH [10]). *There does not exist any stable mechanism that is also strategy-proof, i.e.,* $\bigcap_{P \in \Omega} \Phi(N, P) = \emptyset$.

PROPOSITION 2 (ROTH AND SOTOMAYOR [14, THEOREM 4.6]). *For any stable mechanism* φ *and any profile* P *with* $|S(P)| \geqslant 2$, *these exist a player* i *in* N *and a (misreported) preference* Q_i *in* Ω_i *such that* $\varphi_i(P_{-i}, Q_i)\, P_i\, \varphi_i(P_{-i}, P_i)$.

Proposition 2 shows that at least one player would be better off by misrepresenting his or her true preference P_i, assuming all other players are telling the truth, P_{-i}. Thus, a stable mechanism fails to induce all players to tell the truth in general. In other words, the true profile cannot be strategically revealed by utilizing a stable mechanism.

By definition, the set $\bigcap_{P \in \Omega} \Phi(M, P)$ includes all stable mechanisms that are strategy-proof for the men M. $\bigcap_{P \in \Omega} \Phi(M, P)$ is nonempty because $\varphi^M \in \bigcap_{P \in \Omega} \Phi(M, P)$ by Dubins and Freedman [3] and Roth [10] (see Theorem 4.7 in Roth and Sotomayor [14]). Similarly, $\bigcap_{P \in \Omega} \Phi(W, P)$ is nonempty. Do the sets $\bigcap_{P \in \Omega} \Phi(M, P)$ or $\bigcap_{P \in \Omega} \Phi(M, P)$ include more than one member? Proposition 3 gives a negative answer.

PROPOSITION 3. *$\bigcap_{P\in\Omega}\Phi(M,P)=\{\varphi^M\}$ and $\bigcap_{P\in\Omega}\Phi(M,P)=\{\varphi^W\}$.*

Proof. See Appendix. ▮

Proposition 3 shows that φ^M (φ^W) is the *unique* stable mechanism that always induces the men (women) to reveal their true preferences. Proposition 2 can be strengthened by Proposition 3. Formally,

COROLLARY 4. *The following statements are equivalent.*

(a) $\bigcap_{P\in\Omega}\Phi(N,P)\neq\varnothing$.

(b) $\varphi^M=\varphi^W$.

(c) $|S(Q)|=1$ for all $Q\in\Omega$.

Proof. It is sufficient to prove that (a) implies (b). Since $N=M\cup W$, $\bigcap_{P\in\Omega}\Phi(N,P)\neq\varnothing$ implies both $\bigcap_{P\in\Omega}\Phi(M,P)\neq\varnothing$ and $\bigcap_{P\in\Omega}\Phi(W,P)\neq\varnothing$. By Proposition 3, $\varphi^M=\Phi^W$.

Corollary 4 and Theorem 3.19 in Roth and Sotomayor [14] give a non-example proof of Proposition 1. Theorem 3.19 in Roth and Sotomayor [14] shows that there exists a profile Q such that $|S(Q)|>1$ under our nontriviality assumption: $n, r>1$. By Corollary 4, $|S(Q)|>1$ for some profile Q implies that $\bigcap_{P\in\Omega}\Phi(N,P)=\varnothing$, which leads to Proposition 1.

By Propositions 1 and 2, $\bigcap_{P\in\Omega}\Phi(N,P)$ is empty. Yet $\bigcap_{P\in\Omega}\Phi(T,P)$ is nonempty for all $T\subset M$ because $\bigcap_{P\in\Omega}\Phi(M,P)$ is nonempty. Symmetrically, $\bigcap_{P\in\Omega}\Phi(T,P)$ is nonempty for all $T\subset W$. Can we say anything about $\bigcap_{P\in\Omega}\Phi(T,P)$ when T contains a proper subset of M and a proper subset of W? Specifically, let T contain only one man and one woman. Is the set $\bigcap_{P\in\Omega}\Phi(T,P)$ nonempty as well? Surprisingly, the answer is no.

PROPOSITION 5. *For all T in $\mathcal{T}$ with both $T\cap M$ and $T\cap W$ nonempty, $\bigcap_{P\in\Omega}\Phi(T,P)=\varnothing$.*

Proof. See Appendix. ▮

Proposition 5 generalizes Proposition 1 in another direction. It simply says that there exists no stable mechanism that is strategy-proof for all players in a coalition T if the coalition T contains men and women. Perhaps such "two-sided" incentive aspects of stable mechanisms are unique features of our matching market.

A remark is in order. Proposition 2 generalizes Proposition 1. However, Proposition 2 is not a generalization of Proposition 5. Proposition 2 holds on some subsets of Ω where Proposition 5 fails. The domain of Proposition 5 must be Ω.

EXAMPLE 6. Let $n = r = 3$, $\hat{\Omega}_{m_1} = \{P_{m_1}\}$ and $\hat{\Omega}_{w_1} = \{P_{w_1}\}$ where $P_{m_1} = (w_1, m_1, w_2, w_3)$ and $P_{w_1} = (m_1, w_1, m_2, m_3)$. Also let $\hat{\Omega}_{m_i} = \Omega_{m_i}$ and $\hat{\Omega}_{m_i} = \Omega_{m_i}$ for $i = 2, 3$. Define $\hat{\Omega} = \prod_{m_i \in M} \hat{\Omega}_{m_i} \times \prod_{w_i \in W} \hat{\Omega}_{w_i}$. Note that $\hat{\Omega} \subset \Omega$. There exists $Q \in \hat{\Omega}$ such that $|S(Q)| > 1$ by Theorem 3.19 in Roth and Sotomayor [14]. Thus Proposition 2 applies to $\hat{\Omega}$. But if $T = \{m_1, w_1\}$, Proposition 5 fails on $\hat{\Omega}$.

Consider maximal sets S of players such that $\bigcap_{P \in \Omega} \Phi(S, P)$ is nonempty. By Proposition 5 and Theorem 4.7 in Roth and Sotomayor [14], S either equals M or W. Hence, there are no stable mechanisms which can do better than the MDA and WDA algorithms in the direction of strategy-proofness. Thus, Propositions 1 to 5 provide a complete picture of the incentive aspects of stable mechanisms.

Before we leave this section, we should mention that the negative results stated above are similar to many impossibility theorems in social choice literature under the universal domain assumptions originated by Arrow [1]. The impossibility theorem in Gibbard [7] and Satterthwaite [17] showed that, with a minor assumption on the alternatives, a strategy-proof (voting) mechanism that is also Pareto optimal is dictatorial under the universal domain assumptions. To define an analog for the matching market to the universal domain in the social choice situations, one can define a player j's admissible preference $\mathcal{R}_j$ satisfying Arrow's two axioms namely *completeness*[5] and *transitivity*, on all matchings $\mathcal{M}$ as follows: $\mu \mathcal{R}_j \nu$ if and only if $\mu(j) \mathcal{R}_j \nu(j)$ for all $j \in N$, all $\mu, \nu \in \mathcal{M}$. Denote all such preferences of player j by $\mathcal{U}_j$. Thus, the universal domain of our matching model could be defined by $\mathcal{U} = \prod_{j \in N} \mathcal{U}_j$. Clearly, Ω is a subset of $\mathcal{U}$ where *weak* preferences[6] are not allowed. Because stable mechanisms are nondictatorial, Proposition 1 shows that an analog to the impossibility theorem in Gibbard [7] and Satterthwaite [17] holds on the more restricted domain Ω in our model. Nevertheless, Propositions 3 and 5 show that stable mechanisms may possess some incentive features that may be "peculiar" to some social choice circumstances such as voting.

4. NASH EQUILIBRIA

In this section we study the relations between Nash equilibrium outcomes and stable matchings. Two examples show that a Nash equilibrium outcome may be unstable with respect to the true profile. Recall that we only consider pure strategy equilibria in this paper.

[5] The *connected axiom* in [1].

[6] See Roth and Sotomayor [14] for more detail discussions about weak preferences and the related issues.

DEFINITION 7. Define the *Nash equilibrium* correspondence $N: S^* \times \Omega \to \Omega$ by $N(\varphi, P) = \{Q \mid \neg\varphi_k(Q_{-k}, Q'_k) P_k \varphi_k(Q_{-k}, Q_k)$ for all $k \in N$, all $Q'_k \in \Omega_k.\}$

EXAMPLE 8. Let $Q_k = (k, \cdots) \in \Omega_k$ for all $k \in N$. Then Q is a Nash equilibrium in $N(\varphi, P)$ for all stable mechanisms φ and all underlying true profiles P. Thus, for any stable mechanism φ, there exist a profile P and a Nash equilibrium Q in $N(\varphi, P)$ such that $\varphi(Q)$ is not stable for P.

Given a game $\Gamma(\varphi, P)$, we say that a strategy Q_i of player i with P_i dominates another strategy Q'_i for player i if no matter what the others choose, Q_{-i}, player i likes $\varphi_i(Q_i, Q_{-i})$ at least as much as $\varphi_i(Q'_i, Q_{-i})$. An equilibrium is called a dominated strategy equilibrium if there are some players whose equilibrium strategies are dominated. It is called an undominated strategy equilibrium if every player's strategy in the equilibrium is not dominated. Observe that the equilibrium of Example 8 is a dominated strategy equilibrium when P does not coincide with Q for some players; see Lemma 1 in Roth and Vande Vate [16]. It may be wondered whether some more "sensible" type of Nash equilibria play the same role. For example, it may be asked whether the undominated strategy equilibrium notion could have enough bite to circumvent the negative result. Here is an example, borrowed from Roth and Vande Vate [16], which gives an answer to both.

EXAMPLE 9. (ROTH AND VANDE VATE [16, THEOREM 3(iv)]). Let $n = r = 2$, $P_{m_i} = (w_1, w_2, m_i)$ and $P_{w_i} = (m_1, m_2, w_i)$ for $i = 1, 2$. Construct a profile Q as follows: $Q_{m_i} = (w_1, m_i, w_2)$ and $Q_{w_i} = (m_1, w_i, m_2)$ for $i = 1, 2$. Now the matching $\mu = [(m_1, w_1), (m_2, m_2), (w_2, w_2)]$ is not stable for the true profile P but it is a unique stable matching for the profile Q. Thus, $\varphi(Q) = \mu$ for all stable mechanisms $\varphi \in S^*$. To see that Q is a Nash equilibrium in $N(\varphi, P)$, it is sufficient to observe that w_2 cannot do better by unilaterally listing m_2 and m_2 cannot do better by unilaterally listing w_2. Further, this equilibrium is undominated by Lemma 2 in Roth and Vande Vate [16].

Roth [11] proved that a subset of $N(\varphi^M, P)$ has nice properties. Any Nash equilibrium in which the men reveal their true preferences (the women may misrepresent their true preferences in the equilibrium) will yield a matching that is stable for the true profile P. Such type of Nash equilibrium profiles are specifically denoted by $Q = (P_M, Q_W)$. Certainly, Proposition 10 applies to φ^W symmetrically.

PROPOSITION 10 (ROTH [11]). *For any underlying true profile P and any Nash equilibrium* $Q = (P_M, Q_W) \in N(\varphi^M, P)$, $\varphi^M(Q)$ *is stable for P.*

For the sake of completeness we will give a proof of this proposition. The proof is identical to the proof of Theorem 4.16 in Roth and Sotomayor [14]. We shall use this idea (Theorem 2.25) in the proof twice to show that the Nash equilibria in Proposition 10 are in fact rematching-proof (see Section 5 for the formal definition).

Proof. Let $P \in \Omega$ and $Q = (P_M, Q_W) \in N(\varphi^M, P)$. Clearly, $\varphi^M(Q) \in IR(P)$. Suppose $\varphi^M(Q) \notin PS(P)$. Then $\exists (m, w) \in M \times W$ such that $mP_w \varphi^M_w(Q)$ and $wP_m \varphi^M_m(Q)$. Now construct a strategy $R_w = (m, w, \cdots)$ for the woman w. Let $R = (Q_{-w}, R_w)$. Consider $\varphi^M_w(R)$: If $\varphi^M_w(R) = m$, then $Q \notin N(\varphi^M, P)$ because $\varphi^M_w(R)\, P_w \varphi^M_w(Q)$, a contradiction. Thus, we have $\varphi^M_w(R) = w$. If this is the case, then consider $\varphi^M_m(R)$: If $wP_m \varphi^M_m(R)$, then $\varphi^M(R) \notin S(R)$ because $mR_w w$ and $\varphi^M_w(R) = w$, a contradiction. Thus, we have $\varphi^M_m(R)\, P_m w$. Because $wP_m \varphi^M_m(Q)$ by the assumption, we then have $\varphi^M_m(R)\, P_m \varphi^M_m(Q)$. But by Roth and Sotomayor [14, Theorem 2.25], we have $\neg \varphi^M_m(R)\, P_m \varphi^M_m(Q)$, since $\varphi^M(R)$ restricted to $N \backslash \{w\}$ is clearly the M-optimal stable matching for the market R'_{-w}, where R'_{-w} is the restriction of R_{-w} to $N \backslash \{w\}$. ▮

5. Main Results

In Section 3, we saw that a stable mechanism may be vulnerable to its participants' misrepresentation of preferences. In Section 4, we saw that unstable matchings could arise in some Nash equilibria. In fact, the definition of the Nash equilibrium is so loose that every individually rational matching could be an equilibrium outcome. To show this, given any underlying true profile P, let μ be an individually rational matching in $IR(P)$, and let $Q_i = (\mu(i), i, \cdots)$ if $\mu(i) \neq i$ and $Q_i = (i, \cdots)$ if $\mu(i) = i$ for all players i in N. Then one can easily show that Q is a Nash equilibrium in $N(\varphi, P)$ for all stable mechanisms φ in S^* (where $\varphi(Q) = \mu$; see Roth and Vande Vate [16]). Nevertheless, we also know that there are equilibria whose outcomes are stable matchings for the true profile. The issue addressed here is to see whether some strengthening of the equilibrium notion beyond Nash will exclude unstable matchings. Note that the strengthened equilibrium notion should apply to all stable mechanisms as required in Roth's open question. In this section, we introduce such an equilibrium notion, namely the rematching-proof equilibrium, and show that the defined equilibrium does satisfy the needs and possesses a few more desirable features.

Rematching-proof equilibrium is based on Aumann's [2] strong equilibrium. Recall Aumann's [2] definition of a strong equilibrium: A Nash equilibrium is a strong equilibrium if no coalition of players can make all its members better off by jointly deviating from the equilibrium, taking all others' strategies as given.[7] A rematching-proof equilibrium is similar to Aumann's strong equilibrium in that it also considers coalition formation. It differs from Aumann's definition by the fact that it is restricted to coalitions of couples. Formally, a rematching-proof equilibrium is a Nash equilibrium if there is no couple which can make each better off by jointly deviating from the equilibrium when all others stay put. In other words, players are assumed to be bilaterally rational, in contrast to being unilaterally rational in the Nash equilibrium. Certainly, Nash equilibrium has merits in many economic situations. But it is not the case in the matching markets where almost every matching can arise as a Nash equilibrium. Recall that an outcome of a matching market is a matching that consists of couples and singletons. A couple may well coordinate their behavior if coordination can make both better off.

DEFINITION 11. Define the *rematching-proof equilibrium* correspondence $\tilde{N}: S^* \times \Omega \twoheadrightarrow \Omega$ by $\tilde{N}(\varphi, P) = \{Q \in N(\varphi, P) |$ if $\varphi_m(Q_{-\{m,w\}}, Q'_m, Q'_w)\, P_m \varphi_m(Q)$, then $\neg \varphi_w(Q_{-\{m,w\}}, Q'_m, Q'_w)\, P_w \varphi_w(Q)$, for all $m \in M$, all $w \in W$, all $Q'_m \in \Omega_m$, all $Q'_w \in \Omega_w\}$.

Proposition 12 shows that the Nash equilibria as defined in Proposition 10 are rematching-proof equilibria.

PROPOSITION 12. *For any profile P, if $Q = (P_M, Q_W)$ is a Nash equilibrium in $N(\varphi^M, P)$, then Q is a rematching-proof equilibrium in $\tilde{N}(\varphi^M, P)$.*

An outline of the proof may be helpful. Let P be the underlying true profile and suppose $Q = (P_M, Q_W)$ is a Nash equilibria in $N(\varphi^M, P)$. Assume that Q is not a rematching-proof equilibrium in $\tilde{N}(\varphi^M, P)$. Then there exist a man m and a woman w with a pair of strategies $(R_m, R_w) \in \Omega_m \times \Omega_w$, such that the man m prefers $j = \varphi^M_m(\tilde{Q})$ to $\varphi^M_m(Q)$ and the woman w prefers $i = \varphi^M_w(\tilde{Q})$ to $\varphi^M_w(Q)$, where $\tilde{Q} = (Q_{-\{m,w\}}, R_m, R_w)$. Let $R'_w = (i, w, \cdots) \in \Omega_w$ and $\check{Q} = (Q_{-w}, R'_w)$. We first show that $\varphi^M_w(\check{Q}) = w$; otherwise Q is not a Nash equilibrium in $N(\varphi^M, P)$. But if $\varphi^M_w(\check{Q}) = w$, we show that the man m likes $\varphi^M_m(\check{Q})$ at least as $\varphi^M_m(\check{Q})$. Then Theorem 2.25 in Roth and Sotomayor [14] shows that when a "new woman" w enters the market $\tilde{Q}$, no man is hurt at the M-optimal matching $\varphi^M(\tilde{Q})$ of $\tilde{Q}$ compared to the M-optimal matching $\varphi^M(\check{Q})$ of $\check{Q}$. It follows that it is impossible that the man m prefers $\varphi^M_m(\check{Q})$ to $\varphi^M_m(\tilde{Q})$. Therefore, we have

[7] Motivated by Aumann's definition, coalition proof equilibrium has been discussed in the literature. In matching markets, strong equilibria for all men (i.e., no coalitions of men can profitably deviate) or all women (i.e., no coalitions of women can profitably deviate) were studied by Gale and Sotomayor [6]. For further general discussion about the strong equilibrium, the reader is referred to Fudenberg and Tirole [4, Chap. 1] and the references cited there.

$\varphi_m^M(\check{Q}) = \varphi_m^M(\tilde{Q})$. But if this is the case, we can use Theorem 2.25 again to conclude that the man m likes $\varphi_m^M(Q)$ at least as much as $\varphi_m^M(\check{Q})$ $(=\varphi_m^M(\tilde{Q}))$, which is impossible since the man m prefers $j = \varphi_m^M(\tilde{Q})$ to $\varphi_m^M(Q)$ by the assumption.

Proof. Let $P \in \Omega$ and $Q = (P_M, Q_W) \in N(\varphi^M, P)$. By Proposition 10, $\varphi^M(Q) \in S(P)$. Suppose that Q is not in $\tilde{N}(\varphi^M, P)$. Then by Definition 11, there exist a couple $T = (m, w) \in M \times W$ and a pair of strategies $(R_m, R_w) \in \Omega_m \times \Omega_w$ such that

$$\varphi_m^M(Q_{-T}, R_m, R_w)\, P_m\, \varphi_m^M(Q_{-T}, P_m, Q_w), \tag{1}$$

$$\varphi_w^M(Q_{-T}, R_m, R_w)\, P_w\, \varphi_w^M(Q_{-T}, P_m, Q_w). \tag{2}$$

Denote $\tilde{Q} = (Q_{-T}, R_m, R_w)$. Observe that $w \neq \varphi_m^M(\tilde{Q})$. Otherwise $w = \varphi_m^M(\tilde{Q})$ and thus $m = \varphi_w^M(\tilde{Q})$. Then by (1) and (2), we have $wP_m\varphi_m^M(Q)$ and $mP_w\varphi_w^M(Q)$, which contradicts that $\varphi^M(Q) \in S(P)$.

Now let $R'_w = (\varphi_w^M(\tilde{Q}), w, \cdots) \in \Omega_w$. Denote $\bar{Q} = (Q_{-T}, R_m, R'_w)$. We now show that $\varphi^M(\bar{Q}) = \varphi^M(\tilde{Q})$. Clearly, $\varphi^M(\tilde{Q}) \in IR(\bar{Q})$. Suppose that $\varphi^M(\tilde{Q})$ is not $PS(\bar{Q})$, then $\exists i \in M$ and $\exists j \in W \backslash \{w\}$ such that $jP_i\varphi_i^M(\tilde{Q})$ and $iQ_j\varphi_j^M(\tilde{Q})$ for $i \neq m$ or $jR_m\varphi_m^M(\tilde{Q})$ and $mQ_j\varphi_j^M(\tilde{Q})$ for $i = m$. Both cases imply that $\varphi^M(\tilde{Q}) \notin S(\tilde{Q})$, a contradiction. Thus, $\varphi^M(\tilde{Q}) \in S(\bar{Q})$. Clearly, $\varphi^M(\tilde{Q})$ is the M-optimal stable matching in $S(\bar{Q})$. Thus, we have $\varphi^M(\bar{Q}) = \varphi^M(\tilde{Q})$. Then from (1) and (2), it follows that

$$\varphi_m^M(\bar{Q})\, P_m\, \varphi_m^M(Q), \tag{3}$$

$$\varphi_w^M(\bar{Q})\, P_w\, \varphi_w^M(Q). \tag{4}$$

If $R_m = P_m$, then Q is not in $N(\varphi^M, P)$ by (4), a contradiction. So we have $R_m \neq P_m$. Let $\check{Q} = (Q_{-T}, P_m, R'_w)$. Consider $\varphi^M(\check{Q})$: If $\varphi_w^M(\check{Q}) = \varphi_w^M(\bar{Q})$, by (4) we know that Q is not in $N(\varphi^M, P)$, a contradiction. Then it must be that $\varphi_w^M(\check{Q}) = w$. But if this is the case, we first claim that

$$\neg\varphi_m^M(\tilde{Q})\, P_m\, \varphi_m^M(\check{Q}). \tag{5}$$

If (5) is not true, i.e.,

$$\varphi_m^M(\tilde{Q})\, P_m\, \varphi_m^M(\check{Q}), \tag{6}$$

then P_m is not a dominant strategy for the man m since

$$\varphi_m^M(\tilde{Q}) = \varphi_m^M(\bar{Q})\, P_m\, \varphi_m^M(\check{Q}). \tag{7}$$

Thus (5) is shown.

Next we use Theorem 2.25 in Roth and Sotomayor [14] twice to lead to a contradiction.

From (5), we have either

$$\varphi_m^M(\check{Q}) = \varphi_m^M(\tilde{Q}), \tag{8}$$

or

$$\varphi_m^M(\check{Q})\, P_m\, \varphi_m^M(\tilde{Q}). \tag{9}$$

First, from (1) and (9), we have

$$\varphi_m^M(\check{Q})\, P_m\, \varphi_m^M(\tilde{Q})\, P_m\, \varphi_m^M(Q). \tag{10}$$

Let Q'_{-w} be the restriction of Q_{-w} to $N\backslash\{w\}$. Since $\varphi^M(\check{Q})$ restricted to the set $N\backslash\{w\}$ is the M-optimal stable matching at Q'_{-w} (note that only the woman w changes preferences from Q_w to R'_w), [14, Theorem 2.25] shows that

$$\neg\varphi_m^M(\check{Q})\, P_m\, \varphi_m^M(Q). \tag{11}$$

Thus (10) contradicts with (11).

Second, if (8) is true, by [14, Theorem 2.25] again, then

$$\varphi_m^M(Q)\, P_m\, \varphi_m^M(\check{Q}) \tag{12}$$

or

$$\varphi_m^M(Q) = \varphi_m^M(\check{Q}) \tag{13}$$

must hold.

By (1), (8), and (12), we have

$$\varphi_m^M(Q)\, P_m\, \varphi_m^M(\check{Q}) = \varphi_m^M(\tilde{Q})\, P_m\, \varphi_m^M(Q), \tag{14}$$

a contradiction.

And by (1), (8), and (13), we obtain

$$\varphi_m^M(Q) = \varphi_m^M(\check{Q}) = \varphi_m^M(\tilde{Q})\, P_m\, \varphi_m^M(Q), \tag{15}$$

a contradiction as well. ∎

Next we prove that a rematching-proof equilibrium exists. Proposition 12 and the existence theorem of Nash equilibria in Zhou [18] show the existence of rematching-proof equilibria for the original MDA (WDA) algorithm. Proposition 12 and the existence theorem in Gale and Sotomayor [6] show the existence of rematching-proof equilibria for a generalized MDA (WDA) algorithm. It is, however, not clear if there exists a rematching-proof equilibrium for any stable mechanism.

The truncated strategy equilibria in Roth and Vande Vate [16] play an important role in the following. We first introduce truncated strategies. Let P_i be player i's true preference. An l-truncated strategy Q_i for player i is a strategy with the property that Q_i has common elements with P_i from its top to its lth element ($l \geqslant 0$) in the same order with P_i; the $(l+1)$-th element of Q_i is i, and after the $(l+1)$-st element Q_i could be in any order provided $Q_i \in \Omega_i$.[8] Therefore, a truncated strategy equilibrium is defined to be a Nash equilibrium in which players are restricted to choose their truncated strategies only.

We now show a result that is stronger than existence.

PROPOSITION 13. *For any underlying true profile P, $\bigcap_{\varphi \in S^*} \tilde{N}(\varphi, P) \neq \varnothing$.*

Proof. By Gale and Shapley [5] or Theorem 2.8 in Roth and Sotomayor [14], $S(P) \neq \varnothing$ for any profile $P \in \Omega$. Let $\mu \in S(P)$. Construct a profile $Q = (Q_{m_1}, ..., Q_{m_n}; Q_{w_1}, ..., Q_{w_r})$ as follows:

$$Q_i = \begin{cases} (\overbrace{\cdots\cdots, \mu(i)}^{\text{truncation of } P_i}, i, \cdots) & \text{if} \quad \mu(i) \neq i \\ P_i & \text{if} \quad \mu(i) = i \end{cases}$$

for all $i \in N$.

By Theorem 3(i)(ii) in Roth and Vande Vate [16], Q is a truncated strategy equilibrium for all stable mechanisms $\varphi \in S^*$ since $S(Q) = \{\mu\}$. We now prove that $Q \in \tilde{N}(\varphi, P)$ for all $\varphi \in S^*$. Suppose not for some φ. Then by Definition 11, there exists a pair $T = (m, w) \in M \times W$, with strategies $(R_m, R_w) \in \Omega_m \times \Omega_w$, such that

$$\varphi_m(Q_{-T}, R_m, R_w) \; P_m \, \varphi_m(Q), \tag{16}$$

$$\varphi_w(Q_{-T}, R_m, R_w) \; P_w \, \varphi_w(Q). \tag{17}$$

Observe that $\varphi_m(Q_{-T}, R_m, R_w) \neq w$ and $\varphi_w(Q_{-T}, R_m, R_w) \neq m$ by the same proof as in Proposition 12. Denote $j = \varphi_m(Q_{-T}, R_m, R_w)$. Note that

[8] Note that for a given l, and l-truncated strategy may not exist for some P_i. For example, if $l = 2$ and P_i prefers being single to all possible marriage, then an l-truncated strategy is not well defined. But this is too strong for our purpose. What we need in the context is that for any P_i, there exists at least one l-truncated strategy corresponding to P_i. For the above example, the only truncated strategy is the preference P_i, that is, P_i is a 0-truncated strategy for player i.

$j \in W$ because $jP_m m$ by (16) (since $\varphi(Q) = \mu \in IR(P)$, i.e., $\neg mP_m \varphi_m(Q)$). For the woman j, we claim that

$$\varphi_j(Q)\ P_j m P_j j. \tag{18}$$

First, because $j \neq w$ and $\varphi_j(Q_{-T}, R_m, R_w) = m$ by the definition of j, m must be in Q_j such that $mQ_j j$, which implies $mP_j j$ by the definition of Q_j. Second, $\varphi_j(Q)\ P_j m$. Otherwise, $mP_j \varphi_j(Q)$, but by (16) we also have $jP_m \varphi_m(Q)$, a contradiction to $\varphi(Q) = \mu \in S(P)$. Thus (18) is proved. But (18) contradicts the definition of Q_j. ∎

We immediately obtain existence.

COROLLARY 14. *For any profile P and any stable mechanism* φ, *there exists at least one rematching-proof equilibrium.*

COROLLARY 15. *For any profile P and any matching* μ *that is stable with respect to P, there exists a rematching-proof equilibrium for all stable mechanisms whose outcome is the matching* μ.

The truncated strategy equilibria are a class of undominated strategy equilibria; see Roth and Vande Vate [16]. It may be of interest to note that this class of equilibrium strategies are actually utilized by players in some real matching markets; see Mongell and Roth [8]. But one should be aware of the fact that a truncated strategy equilibrium may result in an unstable matching, as shown by Example 9 and Mongell and Roth [8]

The next proposition shows that a rematching-proof equilibrium achieves a matching that is stable with respect to the true profile even if the equilibrium profile contains misreported preferences. Proposition 16, in conjunction with Proposition 12 and Corollary 14, extends Roth [11] to any stable mechanism. Roth's [13] open question is resolved.

PROPOSITION 16. *Given any stable mechanism* φ *and any underlying true profile P,* $\varphi(Q)$ *is stable for P for all rematching-proof equilibria* $Q \in \tilde{N}(\varphi, P)$.

Proof. Given $(\varphi, P) \in S^* \times \Omega$, let $Q \in \tilde{N}(\varphi, P)$. Clearly, $\varphi(Q) \in IR(P)$. Suppose $\varphi(Q) \notin PS(P)$. Then $\exists (m, w) \in M \times W$ such that $mP_w \varphi_w(Q)$ and $wP_m \varphi_m(Q)$. Let $Q'_w = (m, w, \cdots)$ be a strategy for the woman w and $Q'_m = (w, m, \cdots)$ be a strategy for the man m. also let $T = \{m, w\}$, $Q'_T = (Q'_w, Q'_m)$, and $Q' = (Q_{-T}, Q'_T)$. We claim that $\neg wP_m \varphi_m(Q')$ and $\neg mP_w \varphi_w(Q')$ by $\varphi(Q') \in IR(Q') \cap PS(Q')$ for any $\varphi \in S^*$. If the claim is not true, then $\varphi(Q')$ is not a stable matching for Q'. Thus the claim is proved. By the claim we have $\varphi_m(Q') = w$ and $\varphi_w(Q') = m$, which contradicts the assumption that $Q \in \tilde{N}(\varphi, P)$.

6. Concluding Remarks

The search for a reasonable mechanism that is also strategy-proof has been a consistent theme in the institutional literature for several decades. The main advantage of such a mechanism is that it is immune against opportunistic manipulation through misreporting of true preferences. Unfortunately, for many economic situations, a strategy-proof mechanism fails often to satisfy certain reasonable properties (e.g., nondictorship, efficiency, individual rationality, among others), as established by the large number of impossibility theorems.

Our study in this paper was slightly different. We, in fact, abandoned the strategy-proof property. However, our mechanisms remain reasonable (they produce outcomes in the core for each reported profile). When a mechanism is not strategy-proof, economists commonly believe that the outcomes arrived at by such a mechanism on the basis of the reported profile may be far from what would have been reached on the basis of the true profile. As we demonstrated in Section 4, this actually happens in some Nash equilibria in our matching model. But what we have found is that there is a class of equilibria in which misrepresentation cannot distort the objectives embodied in our mechanisms. Indeed, we defined a non-empty subset of Nash equilibria in which any stable mechanism, known not to be strategy-proof, always achieves a matching that is stable with respect to the true profile. This is quite important because our mechanisms could be implemented reliably in reality in the respective equilibria regardless of players' private information.

The findings of this paper may convey a good signal for those economic situations where a reasonable strategy-proof mechanism cannot be found. Our study illustrates that a reasonable mechanism may produce outcomes that are desirable. In contrasts, a strategy-proof mechanism that is not reasonable will, with certainty, produce outcomes that are not desirable. Accordingly, a reasonable mechanism that is not strategy-proof does not necessarily have less practical value than a strategy-proof mechanism that is not reasonable. This may explain outcomes in the American labor market for graduating medical students and in the United Kingdom labor markets for new physicians. In these markets, stable mechanisms have succeeded in resolving market failures while almost all unstable mechanisms failed to do so; see Roth [9, 12] and Roth and Sotomayor [14].

In closing, we should point out that although we followed Roth [11] to consider a simple two-sided matching model, the issues addressed here should be of equal interest for more general matching markets such as many-to-one matching problems. If one defines a "strong" equilibrium by restricting attention to coalitions of "many and one" for many-to-one matching problems, Proposition 16 is expected to hold. But such a definition may be objected on the basis that players in the "one" side are competitive in general and they may not coordinate their behavior as predicted in the equilibria. Because many-to-one matching problems are quite different from the one-to-one matching problems in many, though not in all, aspects, they deserve a separate investigation. The idea of bilateral rationality of our equilibrium notion may provide a useful hint for such an investigation.

APPENDIX

A player $i \in N$ is called single under a matching μ if $\mu(i) = i$. Denote the set of all single players under a matching μ by S_μ. In order to prove Propositions 3 and 5, we need the following.

PROPOSITION 17.[9] *For any T in $\mathscr{T}$ and any profile P in Ω, $\varphi_k^1(Q) = \varphi_k^2(Q)$ for all profiles $Q = (P_T, Q_{-T})$, all $k \in T$, all $\varphi^1, \varphi^2 \in \Phi(T, P)$.*

Proof. If $\Phi(T, P)$ is empty for some T and P, there is nothing to prove. Suppose $\Phi(T, P)$ is nonempty for all T and P. Let φ^1 and φ^2 be two elements in $\Phi(T, P)$ such that $\exists u \in T$ with $\varphi_u^1(Q) \neq \varphi_u^2(Q)$ at some $Q = (P_T, Q_{-T})$. Thus $\exists m \in T \cap M$ or $w \in T \cap W$ such that for the man m either (a) $\varphi_m^1(Q)\, P_m\, \varphi_m^2(Q)$ or (b) $\varphi_m^2(Q)\, P_m\, \varphi_m^1(Q)$, and for the woman w either (c) $\varphi_w^1(Q)\, P_w\, \varphi_w^2(Q)$ or (d) $\varphi_w^2(Q)\, P_w\, \varphi_w^1(Q)$. One among (a)–(d) must be true by the assumption that $\exists u \in T$ with $\varphi_u^1(Q) \neq \varphi_u^2(Q)$ at some $Q = (P_T, Q_{-T})$. W.L.O.G. suppose (a) is true. Then we get $\varphi_m^1(Q)\, P_m\, \varphi_m^2(Q)$ and $m \notin S_{\varphi^1(Q)} = S_{\varphi^2(Q)}$ by Theorem 2.22 in Roth and Sotomayor [14]. Because $m \in T$, P_m must look like $P_m = (\cdots, \varphi_m^1(Q), \cdots, \varphi_m^2(Q), \cdots, m, \cdots)$. Now construct a preference P'_m as follows: $P'_m = (\cdots, \varphi_m^1(Q), \cdots, m, \cdots, \varphi_m^2(Q), \cdots)$, obtained from P_m by exchanging the position $\varphi_m^2(Q)$ with m. Let $Q' = (Q_{-m}, P'_m)$ and observe that $\varphi^1(Q) \in S(Q')$; thus $m \notin S_{\varphi^1(Q)} = S_\mu$ for any $\mu \in S(Q')$ by Theorem 2.22 in Roth and Sotomayor [14] again. But because $\varphi^2 \in S^*$, we have $\varphi^2(Q') \in S(Q')$. Thus $\varphi_m^2(Q') \neq m$. By individual rationality, we actually have $\varphi_m^2(Q')\, P'_m\, m$. But by the construction of P'_m, we also have $\varphi_m^2(Q')\, P_m\, \varphi_m^2(Q)$, which contradicts that $\varphi^2 \in \Phi(T, P)$. ∎

[9] The highlights in the proof were the same as in the proof of Theorem 4.6 in Roth and Sotomayor [14] (Proposition 2). The proof given here applies to any subset T of N while the proof of Theorem 4.6 pays attention to N only. Our proof seems to generalize slightly the proof of Theorem 4.6. It should be noted that the impossibility results of Propositions 3 and 5 are derived from the positive results of Proposition 17 that ought to satisfy on Ω. Thus one cannot assume that all other players tell the truth as Theorem 4.6 does. To assume all other players are telling the truth will not harm Theorem 4.6 because Theorem 4.6 derives the impossibility theorem directly. It is enough to show Theorem 4.6 by taking an example.

PROPOSITION 3. *$\bigcap_{P \in \Omega} \Phi(M, P) = \{\varphi^M\}$ and $\bigcap_{P \in \Omega} \Phi(W, P) = \{\varphi^W\}$.*

Proof. It is well known that $\varphi^M \in \bigcap_{P \in \Omega} \Phi(M, P)$. If $T = M$, by Proposition 17, for any $\varphi \in \bigcap_{P \in \Omega} \Phi(T, P)$, $\varphi^M_m(Q) = \varphi_m(Q)$ for all $m \in M$ at all $Q = (P_M, Q_W)$. Hence, $\varphi = \varphi^M$. ∎

PROPOSITION 5. *For all T in $\mathcal{T}$ with both $T \cap M$ and $T \cap W$ nonempty, $\bigcap_{P \in \Omega} \Phi(T, P) = \varnothing$.*

Proof. First, observe that $\varphi_m(Q) = \varphi^M_m(Q)$ for all $Q = (P_T, Q_{-T})$, all $T \subset M$, all $m \in T$, and all $\varphi \in \Phi(T, P)$ by Proposition 17. Similarly, observe that $\varphi_w(Q) = \varphi^W_w(Q)$ for all $Q = (P_T, Q_{-T})$, all $T \subset W$, all $w \in T$, and all $\varphi \in \Phi(T, P)$. Now let $T = \{m, w\}$ and assume that $\varphi \in \bigcap_{P \in \Omega} \Phi(T, P)$, where $m \in M$ and $w \in W$. By the observation, we have (★): $\varphi_m(Q) = \varphi^M_m(Q)$ and $\varphi_w(Q) = \varphi^W_w(Q)$ for all $Q = (Q_{-\{m, w\}}, P_m, P_w)$. But by Theorem 3.19 in Roth and Sotomayor [14], there exist a couple $(i, j) \in M \times W$ and a profile $(Q_{-\{i, j\}}, P_i, P_j)$ such that $\varphi^M_i(Q) \neq \varphi^W_i(Q)$ or $\varphi^M_j(Q) \neq \varphi^W_j(Q)$. If $m = i$ and $w = j$ with $\varphi^M_i(Q) = j$ but $\varphi^W_j(Q) \neq i$, then $\varphi_i(Q) = \varphi^M_i(Q) = j$ implies that $\varphi_j(Q) = i \neq \varphi^W_j(Q)$, a contradiction to (★). If $m = i$ and $w = j$ with $\varphi^M_i(Q) \neq j$ but $\varphi^W_j(Q) = i$, then $\varphi_j(Q) = \varphi^W_j(Q) = i$ implies that $\varphi_i(Q) = j \neq \varphi^M_i(Q)$, a contradiction to (★) too.

The proof is the same for any other $T \in \mathcal{T}$ with both $T \cap M$ and $T \cap W$ nonempty. ∎

REFERENCES

1. K. J. ARROW, "Social Choice and Individual Values," second ed., Wiley, New York, 1963.
2. R. J. AUMANN, Acceptable points in general cooperation *n*-persons games, *in* "Contributions to the Theory of Games IV," Princeton Univ. Press, Princeton, 1959.
3. L. E. DUBINS AND D. A. FREEDMAN, Machiavelli and the Gale–Shapley Algorithm, *Amer. Math. Monthly* **88** (1981), 485–494.
4. D. FUDENBERG AND J. TIROLE, "Game Theory," MIT Press, Cambridge, MA, 1992.
5. D. GALE AND L. S. SHAPLEY, College admissions and the stability of marriage, *Amer. Math. Monthly* **69** (1962), 9–15.
6. D. GALE AND M. SOTOMAYOR, Ms Machiavelli and the stable matching problem, *Amer. Math. Monthly* **92** (1985), 223–232.
7. A. GIBBARD, Manipulation of voting schemes: A general result, *Econometrica* **41** (1973), 587–601.
8. S. MONGELL AND A. E. ROTH, Sorority rush as a two-sided matching mechanism, *Amer. Econ. Rev.* **81** (1991), 441–464.
9. A. E. ROTH, The evolution of the labor market for medical interns and residents: A case study in game theory, *J. Polit. Econ.* **92** (1984), 991–1016.
10. A. E. ROTH, The economics of matching: Stability and incentives, *Math. Oper. Res.* **7** (1982), 617–628.
11. A. E. ROTH, Misrepresentation and stability in the marriage problem, *J. Econ. Theory* **34** (1984), 383–387.

12. A. E. ROTH, A natural experiment in the organization of entry-level labor markets: Regional markets for new physicians and surgeons in the United Kingdom, *Amer. Econ. Rev.* **81** (1991), 415–440.
13. A. E. ROTH, Two-sided matching markets: An overview of some theory and empirical evidence, *in* "Game Theory and Applications" (T. Ichiishi, A. Neyman and Y. Tauman, Eds.), Academic Press, New York, 1990.
14. A. E. ROTH AND M. SOTOMAYOR, "Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis," Cambridge Univ. Press, Cambridge, UK, 1990.
15. A. E. ROTH AND J. VANDE VATE, Random paths to stability in two-sided matching, *Econometrica* **58** (1991), 1475–1480.
16. A. E. ROTH AND J. VANDE VATE, Incentives in two-sided matching with random stable mechanisms, *Econ. Theory* **1** (1991), 31–44.
17. M. SATTERTHWAITE, Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions, *J. Econ. Theory* **10** (1975), 187–217.
18. L. ZHOU, Stable matchings and equilibrium outcomes of the Gale–Shapley's algorithm for the marriage problem, *Econ. Lett.* **36** (1991), 25–29.