Giving Advice and Perfect Equilibria in Matching Markets

Joana Pais*

UECE - Research Unit on Complexity in Economics

ISEG/UTL

Very Preliminary

May 2006

Abstract

Our aim is to characterize perfect equilibria in matching markets. Ordinal preferences require the use of an ordinal perfect equilibrium concept. We show that, in the game induced by a random stable mechanism, an ordinal perfect equilibrium strategy lists all the acceptable partners. Moreover, when either the firm-optimal or the worker-optimal mechanisms are considered, truth telling is a very prudent form of behavior and is the unique ordinal perfect equilibrium that may emerge. Finally, in the game induced by these mechanisms, truth telling is an ordinal perfect equilibrium if and only if it is a Nash equilibrium in dominant strategies.

*I have received useful comments from Lars Ehlers, Flip Klijn, Antonio Romero-Medina, and especially from Jordi Massó. I acknowledge the financial support from the Fundação para a Ciência e a Tecnologia under grant n^o SFRH/BD/5308/2001, from the Spanish Ministry of Science and Technology through grant BEC2002-002130 and FEDER. Address: Rua Miguel Lupi, 20, 1249-078 Lisboa, Portugal; email: jpais@iseg.utl.pt

JEL Classification: C78

Keywords: Matching Markets; Perfect Equilibrium; Random Mechanisms

1 Introduction

There is a vast literature on two-sided matching markets. Theoretical investigations in matching exhaust issues on the existence of stable matchings, the structure of the set of such outcomes, and computational algorithms designed to reach them. The strategic decisions that confront individuals under matching mechanisms have also been broadly inspected, focusing particularly on incentives in stable matching mechanisms. That every individually rational matching can be reached as the outcome of an equilibrium play in the game induced by a stable mechanism is a well-known fact (Alcalde, 1996). Nevertheless, agents are, in general, poorly informed and this casts some doubts on the significance of the statement. Indeed, a great deal of information about the preferences of the other agents may be needed to compute an equilibrium; furthermore, the multiplicity of equilibria entails a lot of coordination among agents. Attention is then devoted to a more reasonable class of equilibria, narrowing the set of probable outcomes. In the mechanism that yields the optimal stable matching for one side of the market, Roth (1984) showed that, although agents may have an incentive to misrepresent their preferences, every equilibrium in undominated strategies produces a matching that is stable with respect to the true preferences.

The purpose of this paper is to take this analysis further aiming at a characterization of perfect equilibria in markets organized to produce stable outcomes. Ordinal preferences entail the use of a perfect equilibrium concept with an ordinal flavor. In fact, in an ordinal perfect equilibrium agents play best replies to particular profiles of completely mixed strategies. A best reply, in this context, first-order stochastically dominates every alternative strategy against the mixed strategy profile being considered. Surprisingly, in the mechanism that induces the optimal stable matching for one side of the market, truth telling emerges as the unique ordinal perfect equilibrium. Hence, if acting straightforwardly is, in fact, an ordinal perfect equilibrium, we may postulate that the unique stable matching for the true preferences is the outcome of the game.

Nevertheless, only seldom is truth a Nash equilibrium in the game induced by the

optimal stable mechanism for one side of the market. We can thus anticipate that the existence of ordinal perfect equilibria is exceptional. In reality, a necessary requirement for honesty to be an ordinal perfect equilibrium is being dominant for every agent. Hence, the set of ordinal perfect equilibria and Nash equilibria in dominant strategies coincide in these markets.

Still, the described results may be seen from a brighter perspective. Provided agents are poorly informed, truth telling may be prescribed as a very prudent form of behavior. In the complete information framework, Gale and Sotomayor (1985) have proved that, when any stable mechanism is in use, at least one agent can profitably misrepresent its preferences, except when there is a unique stable outcome. Yet, in order for participants to identify some strategies that perform better than truth telling, a lot of information about others' revealed preferences is needed. In the game induced by the mechanism that yields the firm-optimal stable matching, when each agent has certain beliefs about others' strategies, it is still a dominant strategy for each firm to act straightforwardly (Roth, 1989). On the other hand, Roth and Rothblum (1999) have shown that if workers do not have detailed information about the preferences revealed by other agents in the course of play, the scope of potentially profitable strategic behavior is significantly reduced, if we compare it with the complete information case. If such information exhibits a certain kind of symmetry, reversing the true order of two acceptable firms is to be considered imprudent behavior, but submitting a truncation of the true preferences may be beneficial. Informally, a truncation is a preference ordering that is order-consistent with the true preferences, but under which the worker restricts the number of firms he applies to. Ehlers (2004) takes a further step in the search of advice for workers in a matching market, providing a weaker condition on a worker's beliefs to obtain the conclusions of Roth and Rothblum (1999). Loosely speaking, a worker should not reverse the true ranking of two acceptable firms whenever he is not able to anticipate which new proposals he is going to receive after having rejected others. Moreover, Ehlers (2004) gives advice to workers who can distinguish between three sets of firms: the firms that will certainly propose to him, the firms that may propose, and those from which he does

not expect a proposal.

Hence, there seems to be a clear consensus about how harmful altering the true order of firms may be in a low information environment. The results in this paper suggest that, when a worker contemplates obtaining a proposal from any acceptable firm, he should reveal his whole true preference ordering if he wants to minimize the probability of being unmatched. In fact, truncations may lead to more favorable outcomes, but at the expense of increasing the chances of being alone. Regardless of the incentives to act strategically, honesty thus remains a fundamental form of behavior.

These conclusions already stem from Barberà and Dutta (1995). Barberà and Dutta (1995) show that acting straightforwardly is the unique *protective strategy* for every agent. Loosely speaking, this means that when an agent compares truth telling with any misrepresentation of its preferences, there exists a potential partner with whom, by manipulating, it ends up matched for a larger set of actions of the other players, while less preferred potential partners are obtained, by either acting straightforwardly or strategically, against the same profiles for the rest of society. The concept of protective behavior is based on a refinement of a maxmin criterion and is particularly appropriate for games where agents are poorly informed and sufficiently risk averse.

We proceed as follows. In Section 2, we formally present the marriage model and introduce notation. We define the concept of ordinal perfect equilibrium in Section 3. In Section 4 we develop the main results. We conclude in Section 5 on further research.

2 The Marriage Model

Consider two finite and disjoint sets $F = \{f_1, ..., f_n\}$ and $W = \{w_1, ..., w_p\}$, where F is the set of firms and W is the set of workers. We let $V = W \cup F$ and sometimes refer to a generic agent by v, while w and f represent a generic worker and firm, respectively. Each agent has a strict, complete, and transitive preference relation over the agents on the other side of the market and the perspective of being unmatched. The preferences

of a firm f, for example, can be represented by $P_f = w_3, w_1, f, w_2, ..., w_4$, indicating that f's first choice is to be matched to w_3, its second choice is w_1 and it prefers remaining unmatched to being assigned to any other worker. Equivalently, we may say that w_3 is the lowest ranked worker in P_f, with rank 1 ($r_{P_f}(w_3) = 1$), w_1 is ranked second ($r_{P_f}(w_1) = 2$), being unmatched is ranked third ($r_{P_f}(f) = 3$), and every other worker has a higher ranking in P_w. A worker is *acceptable* if the firm ranks him lower than having its position unfilled; in the above example, the set of acceptable workers is $A(P_f) = \{w_1, w_3\}$. Similarly, given P_w we may define an acceptable firm and $A(P_w)$. It is sufficient to describe only the ordering of acceptable partners, so that the in the above example preferences can be abbreviated as $P_f = w_3,\ w_1$. Let $P = (P_{f_1}, ...,\ P_{f_n}, P_{w_1}, ..., P_{w_p})$ denote the profile of all agents' preferences; we sometimes write it as $P = (P_v, P_{-v})$ where P_{-v} is the set of preferences of all agents other than v. Further, we may use P_U, where $U \subseteq V$, to denote the profile of preferences $(P_v)_{v \in U}$. We write $v' P_v v''$ when v' is preferred to v'' under preferences P_v and we say that v *prefers* v' to v''. We write $v' R_v v''$, when v likes v' *at least as well as* v'' (it may be the case that v' and v'' are the same agent).

Formally, a *marriage market* is a triple (F, W, P). An outcome for a marriage market, a *matching*, is a function $\mu : V \longrightarrow V$ satisfying the following: *(i)* for each f in F and for each w in W, $\mu(f) = w$ if and only if $\mu(w) = f$; *(ii)* if $\mu(f) \neq f$ then $\mu(f) \in W$; *(iii)* if $\mu(w) \neq w$ then $\mu(w) \in F$. If $\mu(v) = v$, then v is *unmatched* under μ, while if $\mu(w) = f$, we say that f and w are *matched* to one another. A description of a matching is given by $\mu = \{(f_1, w_2), (f_2, w_3)\}$, indicating that f_1 is matched to w_2, f_2 is matched to w_3 and the remaining agents in the market are unmatched. A matching μ is *individually rational* if each agent is acceptable to its partner, *i.e.*, $\mu(v) R_v v$, for all $v \in V$. We denote the set of all individually rational matchings by $IR(P)$. Two agents f and w form a *blocking pair* for μ if they prefer each other to the agents they are actually assigned to under μ, *i.e.*, $f P_w \mu(w)$ and $w P_f \mu(f)$. A matching μ is *stable* if it is individually rational and it is not blocked by any pair of agents. We denote the set of all stable matchings by $S(P)$. A firm f and a worker w are *achievable* for each other if f and w are matched under some

stable matching.

The proof of existence of stable matchings in Gale and Shapley (1962) is constructed by means of the deferred-acceptance algorithm. At each step of the algorithm, proposals are issued by one side of the market according to its preferences, while the other side merely reacts to such offers by rejecting all but the best. Hence, in the case that firms make job offers, the algorithm starts with each firm proposing to the first worker on its list and each worker rejecting all proposals but the best. This yields the first tentative matching. Next, every rejected firm makes an offer to its second favorite worker and again workers only hold the one they prefer among those just received and the one held from the previous step. The algorithm proceeds by creating, at each step, a tentative matching and terminates when each firm is either held by a worker or has been rejected by every worker on its list of preferences. This algorithm arrives at the firm-optimal stable matching, with the property that all firms are in agreement that it is the best stable matching. The deferred-acceptance algorithm with workers proposing produces the worker-optimal stable matching with corresponding properties. Further, the optimal stable matching for one side of the market is the worst stable matching for every agent on the other side of the market, a result presented in Knuth (1976) but attributed to John Conway. Still, there is a set of agents who are indifferent between any stable matching. The first statement of this result appears in McVitie and Wilson (1970); later, it was proved in Roth (1984) and Gale and Sotomayor (1985). We state it formally in the next Proposition for further reference.

Proposition 1 *In a matching market (F, W, P), the set of unmatched agents is the same for all stable matchings.*

Finally, a *matching mechanism* $\tilde{\varphi}$ maps preference profiles into lotteries over matchings. In what follows, in a matching market (F, W, P), we consider the revelation game induced by $\tilde{\varphi}$ in which agents are each faced with the decision of what strategies to act on. The strategy space of a player in the game is the set of all possible preference lists: given the true preference ordering P_v, each player v may eventually reveal a different

order Q_v over the players on the other side of the market. A matching mechanism $\tilde{\varphi}$ and a preference profile Q induce a random matching $\tilde{\varphi}[Q]$. Throughout the paper, we only consider *stable* matching mechanisms. Hence, $\tilde{\varphi}[Q]$ denotes the probability distribution induced over the set of stable matchings $S(Q)$ and $\tilde{\varphi}[Q](v)$ is the probability distribution induced over agent v's achievable matches. We use, for example, $\Pr\{\tilde{\varphi}[Q](v)R_v\hat{v}\}$ to denote the probability that v obtains a partner at least as good as $\hat{v}$ according to v's true preferences R_v when the profile Q is used in the mechanism $\tilde{\varphi}$. In the particular case that the mechanism is deterministic, we let $\tilde{\varphi}[Q]$ denote the unique outcome matching. The mechanism that yields the firm-optimal stable matching with certainty is an example of a deterministic stable matching mechanism and will be denoted by φ^F. We let φ^W represent the mechanism that leads to the worker-optimal stable matching.

3 Ordinal Perfect Equilibria

In this section we define ordinal perfect equilibria. We present all definitions for stable mechanisms in general, even though many results refer to the particular case of deterministic mechanisms, namely the mechanisms that yield the optimal stable matching for one side of the market.

Consider $w \in W$ (what follows also holds for a representative firm, with obvious modifications) with true preferences P_w and let Q_{-w} be a strategy profile for all the agents other than w. Given a stable mechanism $\tilde{\varphi}$ and given Q_{-w}, we say that the strategy Q_w *stochastically P_w-dominates* Q'_w if, for all $v \in F \cup \{w\}$, $\Pr\{\tilde{\varphi}[Q_w, Q_{-w}](w)R_wv\} \geq \Pr\{\tilde{\varphi}[Q'_w, Q_{-w}(w)R_wv\}$. Thus, for all $v \in F \cup \{w\}$, the probability of w being assigned to v or to a strictly preferred agent is higher under $\tilde{\varphi}[Q_w, Q_{-w}](w)$ than under $\tilde{\varphi}[Q'_w, Q_{-w}](w)$. Hence, if we consider the problem that player w faces given the strategy choices Q_{-w} of the other players, a particular strategy choice Q_w may be preferred if it stochastically dominates every other alternative strategy. In this case we say that Q_w is a *best reply* to Q_{-w}.

Definition 1 *Given the profile of preferences P, the profile of strategies Q is an* ordinal Nash equilibrium *(*ON equilibrium*) in the game induced by $\tilde{\varphi}$ if, for each agent v in V, Q_v is a best reply to Q_{-v}.*

The concept of ordinal Nash equilibrium deserves a couple of remarks. First, it was introduced in d'Aspremont and Peleg (1988) and its use is required given the very nature of random matching.[1] In fact, agents' preferences are ordinal in nature. Since no natural utility representation of these preferences exists (and no expected utilities can be computed), this ordinal criterion provides a means for comparing probability distributions over potential partners. Second, it is quite a strong equilibrium concept. Under an ordinal Nash equilibrium, each agent plays its best response to the others' strategies for every utility representation of the preferences. However, in the particular case that $\tilde{\varphi}$ is a deterministic mechanism, the concept boils down to plain Nash equilibrium.

For our purposes, some of the above definitions have to be extended to mixed strategies. We let σ denote a mixed strategy and we let $\sigma(Q) = \prod_{v \in V} \sigma_v(Q_v)$ be the probability of profile Q under the mixed strategy σ. Given a stable mechanism $\tilde{\varphi}$ and a mixed strategy profile σ, we let $\tilde{\varphi}[\sigma]$ denote the probability distribution induced over the whole set of matchings that satisfies the following: $\Pr\{\tilde{\varphi}[\sigma] = \mu\} = \sum_{Q \in supp\sigma} \sigma(Q) \cdot \Pr\{\tilde{\varphi}[Q] = \mu\}$. As before, given a mixed strategy profile σ_{-w}, the pure strategy Q_w *stochastically P_w-dominates* Q'_w if, for all $v \in F \cup \{w\}$, $\Pr\{\tilde{\varphi}[Q_w, \sigma_{-w}](w) R_w v\} \geq \Pr\{\tilde{\varphi}[Q'_w, \sigma_{-w}](w) R_w v\}$. The strategy Q_w is a *best reply* to σ_{-w} if it stochastically P_w-dominates every alternative pure strategy. We are now in condition to define ordinal perfect equilibria.

Definition 2 *Given the profile of preferences P, the profile of strategies Q is an* ordinal perfect equilibrium *in pure strategies (*OP equilibrium*) in the game induced by $\tilde{\varphi}$ if there exists a sequence of completely mixed strategies σ^k, $\{\sigma^k\}_{k \to \infty} \to Q$, with the property that, for every $k \geq 1$, Q_v is a best reply to σ^k_{-v}, for every agent v in V.*

Hence, we require that the profile Q be a limit of a sequence of totally mixed profiles

[1] This concept was also used in the context of voting theory in Majumdar and Sen (2004) and in matching markets in Ehlers and Massó (2003), Majumdar (2003), Pais (2004a), and Pais (2004b).

σ^k and that Q_v stochastically P_v-dominates every alternative pure strategy when the opponents use the perturbed strategies σ^k_{-v}.

4 Ordinal Perfect Equilibria

The first couple of results apply to any stable matching mechanism. In Theorem 1, we take a prescriptive point of view and establish that no unacceptable partners should be included in one's list if not matching unacceptable partners is the major concern. Moreover, if an agent wishes to minimize the probability of being unmatched, it should submit a comprehensive preference ordering. In fact, the existence of even the slightest chance of being matched to an acceptable partner should not be neglected.

Theorem 1 *Let $\tilde{\varphi}$ be a stable mechanism. If Q_v is agent v's best reply to a completely mixed strategy profile σ_{-v}, then Q_v lists all the partners that are acceptable according to v's true preferences P_v (i.e., $A(Q_v) = A(P_v)$).*

Proof. Let v be an arbitrary worker. Since the model is symmetric between firms and workers, what follows also holds for an arbitrary firm.

First, we will show that ranking an unacceptable firm f' as acceptable in Q_v is not a best reply to a completely mixed strategy profile for agents other than v, σ_{-v}.

(i) Take any Q_{-v} and note that under any matching $\mu \in S(P_v, Q_{-v})$, v is either always unmatched or always matched to an acceptable firm. Hence, $\Pr\{\tilde{\varphi}[P_v, Q_{-v}](v) R_v v\} = 1$, for all Q_{-v}. It follows that $\Pr\{\tilde{\varphi}[P_v, \sigma_{-v}](v) R_v v\} = \sum_{Q_{-v} \in supp\sigma_{-v}} \sigma(Q_{-v}) \cdot \Pr\{\tilde{\varphi}[P_v, Q_{-v}](v) R_v v\} = 1$.

(ii) Now consider $\hat{Q}_{-v}$ such that $\hat{Q}_{f'} = v$ and no other firm ranks v as acceptable. Then, v will be matched to f' in every matching that is stable for $(Q_v, \hat{Q}_{-v})$. Consequently, $\Pr\{\tilde{\varphi}[Q_v, \hat{Q}_{-v}](v) R_v v\} = 0$.

(iii) Given that $\hat{Q}_{-v}$ has positive probability under σ_{-v}, (ii) implies $\Pr\{\tilde{\varphi}[Q_v, \sigma_{-v}](v) R_v v\} = \sum_{Q_{-v} \in supp\sigma_{-v}} \sigma(Q_{-v}) \cdot \Pr\{\tilde{\varphi}[Q_v, Q_{-v}](v) R_v v\} < 1$. Hence, $\Pr\{\tilde{\varphi}[P_v, \sigma_{-v}](v) R_v v\} >$

$\Pr\{\tilde{\varphi}[Q_v, \sigma_{-v}](v) R_v v\}$ and Q_v is not a best reply to σ_{-v}.

So, let Q_v only rank acceptable firms. We will now prove that deleting an acceptable firm from P_v cannot be a best reply to the completely mixed strategy profile σ_{-v}. So let $f \in A(P_v)$, but $f \notin A(Q_v)$. Let Q'_v be such that the restriction of Q_v and of Q'_v to $A(Q_v)$ coincide, but $f \in A(Q'_v)$ and $f' Q'_v f$, for all $f' \in A(Q_v)$. Note that $A(Q'_v) \subseteq A(P_v)$. We will show that Q_v does not stochastically P_v-dominate Q'_v when the other players choose σ_{-v}.

(i) If, for every Q_{-v}, v is unmatched under $\mu \in S(Q_v, Q_{-v})$, we have $\Pr\{\tilde{\varphi}[Q_v, Q_{-v}](v) P_v v\} = 0$, for all Q_{-v}. Since $A(Q'_v) \subseteq A(P_v)$, we also have $\Pr\{\tilde{\varphi}[Q'_v, Q_{-v}](v) P_v v\} \geq 0$. Hence, $\Pr\{\tilde{\varphi}[Q'_v, Q_{-v}](v) P_v v\} \geq \Pr\{\tilde{\varphi}[Q_v, Q_{-v}](v) P_v v\}$, for every Q_{-v}.

(ii) Otherwise, take any Q_{-v} such that $\mu(v) \in F$, with $\mu \in S(Q_v, Q_{-v})$. Let $Q' = (Q'_v, Q_{-v})$. We will prove that $\mu \in S(Q')$. Clearly, $\mu \in IR(Q')$, by definition of Q'. Now assume, by contradiction, that (f', w) block μ, *i.e.*, $f' Q'_w \mu(w)$ and $w Q'_{f'} \mu(f')$. Since $Q'_{f'} = Q_{f'}$, for every $f' \in F$, we have $w Q_{f'} \mu(f')$. Also, given that $Q'_w = Q_w$, for every $w \neq v$, the stability of μ for Q implies that $w = v$. Hence, $f' Q'_v \mu(v)$ and $v Q_{f'} \mu(f')$. It follows from the definition of Q'_v and the stability of μ for Q that $f' = f$ and that $\mu(v) = v$. This contradicts the initial assumption $\mu(v) \in F$.

We proved that, if $\mu(v) \in F$, for some $\mu \in S(Q_v, Q_{-v})$, we have $\mu \in S(Q')$. Since the set of unmatched agents is the same for all stable matchings (the first statement of this result appears in McVitie and Wilson, 1970; it also appears in Gale and Sotomayor, 1985, and Roth, 1984), v is matched under every matching that is stable for Q'. It follows that, for all Q_{-v}, $\Pr\{\tilde{\varphi}[Q'_v, Q_{-v}](v) P_v v\} \geq \Pr\{\tilde{\varphi}[Q_v, Q_{-v}](v) P_v v\}$.

(iii) To see that there exists some $\hat{Q}_{-v}$ for which v ends up alone when stating Q_v, but matched when using Q'_v, suppose $\hat{Q}_f = v$ and no other firm ranks v as acceptable. Then, v is matched to f with certainty at $(Q'_v, \hat{Q}_{-v})$, whereas he stays alone if using Q_v. As a consequence, there exists a $\hat{Q}_{-v}$ such that $1 = \Pr\{\tilde{\varphi}[Q'_v, \hat{Q}_{-v}](v) P_v v\} > \Pr\{\tilde{\varphi}[Q_v, \hat{Q}_{-v}](v) P_v v\} = 0$.

(iv) Since all profiles of preferences Q_{-v} have positive probability in the completely mixed strategy profile σ_{-v}, v will be unmatched with higher probability when using Q_v than when using Q'_v, we have $\Pr\{\tilde{\varphi}[Q'_v, \sigma_{-v}](v) P_v v\} > \Pr\{\tilde{\varphi}[Q_v, \sigma_{-v}](v) P_v v\}$ and Q_v is not a best reply to σ_{-v}. ∎

Since ordinal perfect equilibrium strategies are best replies to completely mixed strategy profiles it immediately follows from the above result that ordinal perfect equilibrium strategies have to be exhaustive, listing all the acceptable partners, but leaving out those considered unacceptable. We state this formally in the following corollary.

Corollary 1 *Let $\tilde{\varphi}$ be a stable mechanism. If Q is an ordinal perfect equilibrium in the game induced by $\tilde{\varphi}$, every agent v ranks in Q_v all the partners that are acceptable according to its true preferences P_v (i.e., $A(Q_v) = A(P_v)$, for all $v \in V$).*

Two further implications of the above theorem are worth noticing. The first and most immediate goes against the celebrated properties of strict truncations. Formally, a strict truncation of an agent v's true preferences P_v containing p acceptable partners is a strategy that lists the first p', $p' < p$, elements of P_v as acceptable, preserving their order in P_v. Revealing a strict truncation of the true preferences may not be wise when one highly esteems being matched; furthermore, strict truncation strategies cannot be part of an ordinal perfect equilibrium in the game induced by a stable mechanism.

Second, it allows us to restrict the set of potential outcomes. As it will be readily understood, not every individually rational matching is sustainable as the outcome of an equilibrium play where agents fully reveal whom they are willing to match. In what follows, we describe those matchings that are beyond reach and state the result.

Definition 3 *Let $U(P)$ be the set of all individually rational matchings μ such that at least one of its blocking pairs either includes one agent that is unmatched under μ or consists of a pair of unmatched agents under μ.*

Proposition 2 *Let $\tilde{\varphi}$ be a stable mechanism. Let μ be a matching in $U(P)$. In the game induced by $\tilde{\varphi}$, μ is not sustainable in an ordinal perfect equilibrium.*

Proof. By Theorem 1, listing all the acceptable partners is a necessary requirement for an ordinal perfect equilibrium strategy. So, let Q be an ordinal Nash equilibrium such that $A(Q_v) = A(P_v)$ for all v. We will show that Q cannot support $\mu \in U(P)$.

By contradiction, assume that it does. In the game induced by a stable mechanism, every Nash equilibrium yields a single matching with probability one (Pais, 2004). Hence, Q leads to μ with probability one. By definition of μ, there exists at least one blocking pair for μ consisting of a firm f and a worker w such that either f or w or both are unmatched under μ. If both are unmatched under μ, since $f \in A(Q_w)$ and $w \in A(Q_f)$, we have $fQ_w\mu(w)$ and $wQ_f\mu(f)$. Hence, μ is not stable for Q and it cannot be reached as the outcome of a random stable mechanism where agents use Q.

So, let μ be blocked by (f, w) such that one of its members is unmatched, while the other one is matched under μ. Since the model is symmetric between firms and workers, it is sufficient to prove the proposition for, say, f unmatched and w matched under μ. Now let Q'_w be identical to Q_w, but such that f is listed first in Q'_w even if it occupies a worse position in Q_w. Define $Q' = (Q'_w, Q_{-w})$. We will show that Q'_w is different from Q_w (i.e., Q_w does not list f first); in addition, Q'_w is a profitable deviation to Q_w, since f and w are matched with certainty under any matching in $S(Q')$.

To prove that f is matched to w under the firm-optimal stable matching at Q', we will use the deferred-acceptance algorithm with firms proposing. Since $Q'_v = Q_v$, for every agent $v \neq w$, all proposals, acceptances, and rejections take place exactly the same way as when Q was being used, up to the point where w holds f's proposal. This moment comes since $w \in A(Q_f)$ and f is the first firm in Q'_w. Also, w will not reject f until the final matching is reached, so that f and w are together under the firm-optimal stable matching.

It follows from the definition of worker-optimal stable matching that w holds f or a firm ranked higher than f at Q'_w under any stable matching for Q'. Since f is the first firm in Q'_w, w is matched to f under all stable matchings. This implies that $Q'_w \neq Q_w$; otherwise, (f, w) could not block μ.

To conclude, we have $1 = \Pr\{\tilde{\varphi}[Q'_w, Q_{-w}](w) = f\}$. Since Q leads to μ with certainty, $\Pr\{\tilde{\varphi}[Q](w) = f\} = 0$. Hence, $\Pr\{\tilde{\varphi}[Q'_w, Q_{-w}](w) R_w f\} > \Pr\{\tilde{\varphi}[Q_w, Q_{-w}](w) R_w f\}$ and Q_w is not a best reply to Q_{-w}, contradicting that Q is a Nash equilibrium. Since no Nash equilibrium where agents list all the acceptable partners can sustain a matching in $U(P)$, no ordinal perfect equilibrium will. ∎

In the other direction, it may be shown that every matching in $IR(P)\backslash U(P)$ may be sustained as the unique outcome of an equilibrium play of the game where all agents reveal the full set of acceptable partners.

Proposition 3 *Let $\tilde{\varphi}$ be a stable mechanism. Let μ be a matching in $IR(P)\backslash U(P)$. Then, there exists an ordinal Nash equilibrium Q in the game induced by $\tilde{\varphi}$ with the following properties:*

1. *every agent v ranks as acceptable in Q_v all of its acceptable partners under P_v*
2. *Q sustains μ with probability one.*

Proof. Let μ be a matching in $IR(P)\backslash U(P)$ and consider agent v. Let Q_v be such that (i) $A(Q_v) = A(P_v)$ and (ii) *if* v is matched under μ, $\mu(v) Q_v v'$, for all $v' \in A(P_v)$, *i.e.*:

$$Q_v = \begin{cases} \overbrace{\mu(v), \ldots\ldots\ldots\ldots\ldots\ldots\ldots}^{\text{All elements of } A(P_v) \text{ in any order}}, v & \text{if } \mu(v) \neq v \\ \\ \underbrace{\ldots\ldots\ldots\ldots\ldots\ldots\ldots}_{\text{All elements of } A(P_v) \text{ in any order}}, v & \text{if } \mu(v) = v \end{cases}.$$

We will show that Q has all the described properties.

First, we will show, by contradiction, that $\mu \in S(Q)$. By definition of Q, it is clear that $\mu \in IR(Q)$. So assume (f, w) blocks μ when the profile Q is considered. By (ii), this implies that f and w are unmatched under μ. We thus have $w Q_f f$ and $f Q_w w$. Since $A(Q_v) = A(P_v)$ for every agent v, it follows that $w P_f f$ and $f P_w w$. Hence, (f, w) are unmatched under μ and block μ for preferences P. This contradicts that $\mu \in IR(P)\backslash U(P)$.

Now we will prove that μ is the unique matching in $S(Q)$. Assume not and take any matching $\hat{\mu}$, $\hat{\mu} \neq \mu$, in $S(Q)$. Let v be such that $\mu(v) = v$. Then, $\hat{\mu}(v) = v$ since, by Proposition 1, the same set of agents is unmatched under every matching that belongs to $S(Q)$. On the other hand, for $\hat{v}$ such that $\mu(\hat{v}) \neq \hat{v}$, we must have $\hat{\mu}(\hat{v}) = \mu(\hat{v})$ by (ii). Otherwise, $(\hat{v}, \mu(\hat{v}))$ blocks $\hat{\mu}$. Hence, $\hat{\mu} = \mu$ and μ is the only matching in $S(Q)$. As a consequence, every random stable mechanism leads to μ with probability one.

To complete the proof, we must show that Q is an ordinal Nash equilibrium. By contradiction, suppose firm f can profitably deviate by matching worker w (the same argument holds for an arbitrary worker). This implies that there exists a worker w willing to accept f, *i.e.*, such that $fQ_w\mu(w)$. By (ii), we must have $\mu(w) = w$ and, by (i), $fP_w\mu(w)$. Since $\mu \in IR(P)\backslash U(P)$, it follows that $\mu(f)P_f w$ and we contradict the initial assumption: matching w is not a profitable deviation for f. ∎

We now state an important result for deterministic stable mechanisms that follows from Barberà and Dutta (1995). In this paper, revealing the true preferences is most convenient for agents who are extremely risk averse. In fact, when an agent compares straightforward behavior with any misrepresentation of its preferences, there exists a potential partner with whom, by manipulating, it ends up matched for a larger set of actions of the other players; further, less preferred potential partners are obtained, by either acting straightforwardly or strategically, against the same profiles for the rest of society. It thus follows that when an agent's beliefs are such that all preference profiles for the other agents may be revealed with positive probability, behaving strategically is never a best reply. We state this formally in the next Theorem. Even though the result applies to the mechanism producing the firm-optimal stable matching, it is straightforward to extend it to a market using the worker-optimal stable mechanism.

Theorem 2 *[Barberà and Dutta (1995)] In the game induced by φ^F, if Q_v is a best reply to a completely mixed strategy profile σ_{-v}, then Q_v are agent v's true preferences (i.e., $Q_v = P_v$).*

It clearly follows that only truth telling may be an ordinal perfect equilibrium in the game induced by the mechanism that yields an optimal stable matching. We state this as a corollary to Theorem 2.

Corollary 2 *In an ordinal perfect equilibrium of the game induced by φ^F every agent states its true preferences.*

This result anticipates that ordinal perfect equilibria only seldom exist in matching markets. The following example supports this observation.

Example 1 *A market where there are no ordinal perfect equilibria in pure strategies.*

Let (F, W, P) be a marriage market with P such that

$$P_{w_1} = f_2,\ f_1 \quad P_{f_1} = w_1,\ w_2$$
$$P_{w_2} = f_1,\ f_2 \quad P_{f_2} = w_2,\ w_1.$$

Consider the game induced by the mechanism that produces the firm-optimal stable matching. Corollary 2 establishes that, under an ordinal perfect equilibrium in this game, every agent chooses the honest announcement of preferences. In this case, the mechanism would yield the matching $\mu = \{(f_1, w_1), (f_2, w_2)\}$. Nevertheless, truth telling fails to meet the basic requirement of being a Nash equilibrium of the game, since both workers can profitably deviate. For example, submitting $Q_{w_1} = f_2$ is a deviation for worker w_1, conveying the position in f_2. ◇

We can extend the above result to random stable mechanisms that only assign positive probability to the firm-optimal and to the worker-optimal stable matchings.

Proposition 4 *Let $\tilde{\varphi}$ be a stable mechanism that yields the firm-optimal stable matching with probability α and the worker-optimal stable matching with probability $1 - \alpha$, $0 < \alpha < 1$. In an ordinal perfect equilibrium in the game induced by $\tilde{\varphi}$ every agent states its true preferences.*

Proof. Let w be an arbitrary worker. We will show that P_w is the only strategy that can be part of an OP equilibrium in the game induced by the random stable mechanism $\tilde{\varphi}$ as defined above. We omit the proof for an arbitrary firm f, since the same arguments, with obvious modifications, can be applied.

Let Q_w be a strategy for w such that $Q_w \neq P_w$. Assume that Q_w is a best reply to a completely mixed strategy profile σ_{-w}. This has two implications. First, by Corollary 1, $A(Q_w) = A(P_w)$. Second, $\Pr\{\tilde{\varphi}[Q_w, \sigma_{-w}](w) R_w v\} \geq \Pr\{\tilde{\varphi}[P_w, \sigma_{-w}](w) R_w v\}$, for every v, a potential partner of w. By definition of $\tilde{\varphi}$, we have $\alpha \Pr\{\varphi^F[Q_w, \sigma_{-w}](w) R_w f\} + (1-\alpha) \Pr\{\varphi^W[Q_w, \sigma_{-w}](w) R_w f\} \geq \alpha \Pr\{\varphi^F[P_w, \sigma_{-w}](w) R_w f\} + (1-\alpha) \Pr\{\varphi^W[P_w, \sigma_{-w}](w) R_w f\}$, for every v. Nevertheless, since truth telling is a dominant strategy for workers in the game induced by the worker-optimal stable mechanism, it is a best reply to any mixed strategy profile and $\Pr\{\varphi^W[P_w, \sigma_{-w}](w) R_w f\} \geq \Pr\{\varphi^W[Q_w, \sigma_{-w}](w) R_w f\}$. Moreover, Theorem 2 states that honestly revealing the true preferences is a best reply in the game induced by the firm-optimal stable mechanism, so that $\Pr\{\varphi^F[P_w, \sigma_{-w}](w) R_w f\} \geq \Pr\{\varphi^F[Q_w, \sigma_{-w}](w) R_w f\}$. It follows that $\varphi^W[P_w, \sigma_{-w}](w) = \varphi^W[Q_w, \sigma_{-w}](w)$ and $\varphi^F[P_w, \sigma_{-w}](w) = \varphi^F[Q_w, \sigma_{-w}](w)$. Since σ_{-w} is a completely mixed strategy profile, these distributions have full support, *i.e.*, every firm in $A(P_w)$ is obtained as a partner with positive probability. This implies that $Q_w = P_w$, contradicting the initial assumption.

As a consequence, only P_w can be a best reply to a mixed strategy; thus only P_w can be part of an OP equilibrium. ∎

Finally, confirming our conjecture on the existence of ordinal perfect equilibria, in the next result we show that truth telling can only be an ordinal perfect equilibrium if it is a dominant strategy for every agent. Hence, the concept of ordinal perfect equilibrium and the apparently stronger concept of Nash equilibria in dominant strategies coincide.

Theorem 3 *In the game induced by φ^F, the sets of ordinal perfect equilibria and of Nash equilibria in dominant strategies coincide.*

Proof. It is clear that every Nash equilibrium in dominant strategies is an OP equilibrium. In fact, a dominant strategy is a best reply to all profiles of preferences

stated by the other players; hence, it is also a best reply to any completely mixed strategy profile. The converse statement will be shown in what follows.

Theorem 2 imposes as a necessary requirement for an OP equilibrium that every agent states its true preferences. Hence, let P be an OP equilibrium in φ^F, but assume that stating the true preferences is not a dominant strategy for some worker w. Then, there exists at least one instance, *i.e.*, a strategy profile for the other players, under which playing strategically pays for worker w. Denote by Q_{-w} such a strategy profile and let Q_w be the best reply to Q_{-w}. Formally,

$$\varphi^F[Q_w, Q_{-w}](w) P_w \varphi^F[P_w, Q_{-w}](w) \text{ and} \tag{1}$$

$$\varphi^F[Q_w, Q_{-w}](w) R_w \varphi^F[\bar{Q}_w, Q_{-w}](w), \text{ for every } \bar{Q}_w. \tag{2}$$

Let, without loss of generality, $P_w = f_1, f_2, ..., f_m$ and $f_j = \varphi^F[Q_w, Q_{-w}](w)$, with $1 \leq j \leq m$.

Now define $Q'_w = f_j, f_{j-1}, ..., f_1$. Observe that $\varphi^F[Q_w, Q_{-w}] \in S(Q'_w, Q_{-w})$, since it remains individually rational once w uses Q'_w and there are potentially fewer blocking pairs for $\varphi^F[Q_w, Q_{-w}]$. Hence, Proposition 1 implies that w is matched under every matching in $S(Q'_w, Q_{-w})$; in addition, the definition of Q'_w implies that he is matched to a firm at least as good as f_j according to P_w. By (2), $\varphi^F[Q_w, Q_{-w}](w) R_w \varphi^F[Q'_w, Q_{-w}](w)$, so that we must have $\varphi^F[Q'_w, Q_{-w}](w) = f_j$. Then, if Q_w gives w matched to f_j against Q_{-w}, Q'_w also matches w to f_j. Hence, for the profile Q_{-w}, condition (1) yields $f_j = \varphi^F[Q'_w, Q_{-w}](w) P_w \varphi^F[P_w, Q_{-w}](w)$.

Now let us prove that there is no instance $\hat{Q}_{-w}$ under which P_w matches w to a firm at least as good as f_j, while Q'_w leaves w unmatched. By contradiction, assume that, by playing truthfully, w is matched to f_i, $i \leq j$, but unmatched when using Q'_w against $\hat{Q}_{-w}$ in the game induced by φ^F. If this is so, by Proposition 1, w is unmatched under every matching that is stable for $(Q'_w, \hat{Q}_{-w})$ and, in particular, we have $\varphi^W[Q'_w, \hat{Q}_{-w}](w) = w$. On the other hand, by definition of worker-optimal stable matching, $\varphi^W[P_w, \hat{Q}_{-w}](w) R_w \varphi^F[P_w, \hat{Q}_{-w}](w) = f_i$. Since $f_i R_w f_j$, we have $\varphi^W[P_w, \hat{Q}_{-w}](w) \in A(Q'_w)$. Now imagine Q'_w are w's true preferences. By acting

strategically and using P_w, w is better off than by straightforwardly revealing Q'_w. This contradicts the fact that truth is a dominant strategy for workers in the game induced by φ^W. Hence, w is matched to a firm at least as good as f_j with P_w, by manipulating and using Q'_w, w will also be matched to a firm at least as good as f_j. We thus have, for every $\bar{Q}_{-w}$ that yields $\varphi^F[P_w, \bar{Q}_{-w}](w)R_w f_j$, that $\varphi^F[Q'_w, \bar{Q}_{-w}](w)R_w f_j$.

Consider a completely mixed strategy profile σ_{-w}. In the game induced by φ^F, when playing against σ_{-w}, Q'_w yields a higher probability of being matched to a firm at least as good as f_j than P_w. Clearly, P_w cannot be part of an OP equilibrium, contradicting the initial assumption. ∎

The whole picture changes when we depart from the ordinal framework. As shown in the following example, if agents are able to go beyond an ordering of the possible matches and provide a measure of their preferences, strategic behavior may be held in a perfect equilibrium.

Example 2 (Example 1 (revisited)) *Acting strategically may be a perfect equilibrium when agents can give a cardinal meaning to their preferences.*

Consider the game induced by the mechanism that yields the firm-optimal stable matching in the matching market described above. Consider the profile of strategies $Q = (P_F, Q_W)$, such that each worker only finds his first choice acceptable in Q (*i.e.*, $Q_{w_1} = f_2$ and $Q_{w_2} = f_1$). We will show that, depending on the utility representation of the workers' preferences, Q may be a perfect equilibrium of the game.

Each agent has five different strategies at its disposal (two of them stating two acceptable matches, other two naming only one, and the strategy where all potential partners are unacceptable). Let σ^k be a sequence of completely mixed strategy profiles such that, for $k \geq 1$ and for every agent v, $\sigma^k(\hat{Q}_v) = \frac{1}{k+4}$, for all $\hat{Q}_v \neq Q_v$, and $\sigma^k(Q_v) = 1 - \frac{4}{k+4}$. Note that $\{\sigma^k\} \xrightarrow[k\to\infty]{} Q$. Revealing the true preferences is a dominant strategy for each firm f in this game (Dubins and Freedman, 1981, and Roth, 1982), outperforming every alternative strategy for every profile chosen by the other agents, namely σ^k_{-f}, for every

$k \geq 1$. So, consider worker w_1 (by symmetry, what follows also holds for w_2); we will prove that Q_{w_1} is a best reply to $\sigma^k_{-w_1}$.

(i) Consider $Q'_{w_1} = w_1$; note that w_1 is always unmatched when using this strategy against any profile of strategies of the other players. Hence, w_1 is unmatched with certainty when playing Q'_{w_1} against $\sigma^k_{-w_1}$. It follows that Q'_{w_1} is stochastically P_{w_1}-dominated by every other strategy that w_1 may use, in particular by $P_{w_1}$when playing against $\sigma^k_{-w_1}$.

(ii) The strategy $Q''_{w_1} = f_1$ is also stochastically P_{w_1}-dominated by $P_{w_1} = f_2, f_1$ against $\sigma^k_{-w_1}$. In fact, if w_1 is unmatched when using P_{w_1}against Q_{-w_1}, he will certainly be unmatched with Q''_{w_1}. So, when playing against $\sigma^k_{-w_1}$, w_1 is matched with higher probability if he uses P_{w_1}. Moreover, there exist profiles of strategies for the other players such that w_1 is matched to f_2 under the outcome of the deferred-acceptance algorithm when revealing Q'''_{w_1}, but not with Q''_{w_1}, where f_2 is considered unacceptable. The conclusion follows.

(iii) Now consider $Q'''_{w_1} = f_1, f_2$. Note that w_1 is unmatched when using this strategy if and only if he is unmatched with $P_{w_1} = f_2, f_1$. Furthermore, for every profile of the other players such that w_1 is assigned to f_2 with Q'''_{w_1}, he is also assigned to f_2 when using P_{w_1}; and there are profiles of strategies for the other players such that w_1 is matched to f_2 when revealing P_{w_1}, but not with Q'''_{w_1}. It follows that P_{w_1} stochastically P_{w_1}-dominates Q'''_{w_1} against any completely mixed strategy profile $\sigma^k_{-w_1}$.

(iv) Since P_{w_1} outperforms Q'_{w_1}, Q''_{w_1}, and Q'''_{w_1}, it is sufficient to find under which conditions Q_{w_1} may be preferred to P_{w_1}. There is an instance under which submitting Q_{w_1} provides w_1 with a better partner. In fact, f_2 is w_1's partner under the firm-optimal stable matching with $(Q_{w_1}, \hat{Q}_{w_2}, P_{f_1}, P_{f_2})$, where $\hat{Q}_{w_2} = f_1, f_2$; by using P_{w_1} against the same profile for the others, w_1 ends up matched to f_1. Nevertheless, w_1 is unmatched when revealing Q_{w_1} against a larger set of profiles of the other players, than when using P_{w_1}. It turns out that Q_{w_1} is a best reply to $\sigma^k_{-w_1}$, if the following condition holds:

$k^2[u(f_2)-u(f_1)] \geq (4(k+4)^2-17k-51)[u(f_1)-u(w_1)]$.[2] In particular, when $u(f_2)-u(f_1)$ is much larger than $u(f_1)-u(w_1)$, w_1 benefits from listing only his first choice. $\diamond$

5 Weak Ordinal Perfect Equilibria

Definition 4 *Given the profile of preferences P, the profile of strategies Q is a* weak ordinal perfect equilibrium *in pure strategies (*OP equilibrium*) in the game induced by $\tilde{\varphi}$ if there exists a sequence of completely mixed strategies σ^k, $\{\sigma^k\}_{k\to\infty} \to Q$, with the property that, for every $k \geq 1$, Q_v is not (strictly) stochastically P_v-dominated by any alternative pure strategy when played against σ^k_{-v}, for every agent v in V.*

Proposition 5 *Let $\tilde{\varphi}$ be a stable mechanism that yields the firm-optimal stable matching with probability α and the worker-optimal stable matching with probability $1-\alpha$, $0 \leq \alpha \leq 1$. In a weak ordinal perfect equilibrium in the game induced by $\tilde{\varphi}$ every agent ranks its best partner first.*

Proof. Let Q_w be a strategy for worker w that doesn't list w's best partner, f_1, first. We will show that Q_w is (strictly) stochastically P_w-dominated when played against any perfectly mixed strategy profile in the game induced by $\tilde{\varphi}$. By symmetry, the same arguments apply to a firm's strategy that doesn't list its preferred worker first.

Let Q'_w be a strategy identical to Q_w, except for the fact that f_1 is listed first in Q'_w (formally, $vQ'_wv' \iff vQ_wv'$, for every v, $v' \neq f_1$ and $f_1Q'_wf$, for every firm f). In what follows, we let f be a firm different from f_1. Also, Q_{-w} is any profile of strategies for agents other than w and we let $Q = (Q_w, Q_{-w})$, $Q' = (Q'_w, Q_{-w})$.

Claim 1: Q'_w dominates Q_w in the game induced by φ^W.

Proof: We start by showing that $\varphi^W[Q](w) = f$ implies that $\varphi^W[Q'](w) \in \{f_1, f\}$. By contradiction, let $\varphi^W[Q](w) = f$ and assume that $\varphi^W[Q'](w) \notin \{f_1, f\}$. If Q'_w were

[2] This expression results from considering the outcomes of the deferred-acceptance algorithm when w_1 uses P_{w_1} and Q_{w_1} for all possible combinations of the other agents' preferences. This calls for performing a total of 5^3 tedious comparisons that we leave out for obvious reasons.

w's true preferences, w could not obtain a better partner than $\varphi^W[Q'](w)$ by manipulating and using Q_w, since truth is a dominant strategy in the game induced by φ^W. Hence, $\varphi^W[Q'](w)Q'_w f$. By definition of worker-optimal stable matching, f is the best achievable partner for w when Q_w is used against Q_{-w}, implying that $\varphi^W[Q'] \notin S(Q)$. Since $Q'_v = Q_v$ for any $v \neq w$ and $\varphi^W[Q'](w) \in A(Q_w)$ by definition of Q'_w, we have $\varphi^W[Q'](w) \in IR(Q)$. It thus follows that a pair of agents $(\hat{f}, \hat{w})$ blocks $\varphi^W[Q']$ in Q. It must be the case that $\hat{w} = w$, otherwise $(\hat{f}, \hat{w})$ would block $\varphi^W[Q']$ in Q'. As $(\hat{f}, w)$ block $\varphi^W[Q']$, $\hat{f}Q_w\varphi^W[Q'](w)$ and $wQ_{\hat{f}}\varphi^W[Q'](\hat{f})$. By definition of Q'_w and since $\varphi^W[Q'](w) \neq f_1$, $\hat{f}Q'_w\varphi^W[Q'](w)$; since $Q'_{\hat{f}} = Q_{\hat{f}}$, $wQ'_{\hat{f}}\varphi^W[Q'](\hat{f})$. Therefore, we obtain a contradiction: $(\hat{f}, w)$ block $\varphi^W[Q']$ in Q'.

Now it is left to prove that there exists at least one instance under which w gets f_1 when using Q'_w but not by means of Q_w. Let f' be such that $f'Q_w f_1$ (such a firm exists since by assumption Q_w doesn't list f_1 first); let Q'_{-w} be such that $Q'_{f_1} = Q'_{f'} = w, w'$, $Q'_{w'} = f_1, f'$, and no other firm lists w. It is clear to see that $\varphi^W[Q_w, Q'_{-w}](w) = f'$, while $\varphi^W[Q'_w, Q'_{-w}](w) = f_1$.

Claim 2: Q'_w dominates Q_w in the game induced by φ^F.

Proof: We show that $\varphi^F[Q](w) = f$ implies that $\varphi^F[Q'](w) \in \{f_1, f\}$ by means of the deferred-acceptance algorithm with firms proposing. Since $Q'_v = Q_v$, for every agent $v \neq w$, all proposals, acceptances, and rejections take place exactly the same way as when Q was being used, unless w obtains f_1's proposal, in which case he accepts it. Hence, either $\varphi^F[Q'](w) = \varphi^F[Q](w) = f$ or, if f_1 indeed proposes to w, $\varphi^F[Q'](w) = f_1$.

Furthermore, there exists a profile of strategies for players other than w under which w is matched to f_1 when using Q'_w but not when using Q_w. Let f' be such that $f'Q_w f_1$.Let Q'_{-w} be such that $Q'_{f_1} = Q'_{f'} = w, w'$, $Q'_{w'} = f_1, f'$, and such that no other firm considers w acceptable. Then, $\varphi^F[Q_w, Q'_{-w}](w) = f'$, but $\varphi^F[Q'_w, Q'_{-w}](w) = f_1$.

Claim 3: Q'_w stochastically P_w-dominates Q_w against any completely mixed strategy profile in the game induced by $\tilde{\varphi}$.

Proof: It follows from *Claim 1*, *Claim 2*, and by definition of $\tilde{\varphi}$ that $\Pr\{\tilde{\varphi}[Q'_w, Q_{-w}](w)R_w v\} \geq \Pr\{\tilde{\varphi}[Q_w, Q_{-w}](w)R_w v\}$, for any agent v. Hence, $\Pr\{\tilde{\varphi}[Q'_w, \sigma_{-w}](w)R_w v\} = \sum_{Q_{-w}} \sigma_{-w}(Q_{-w}) \cdot \Pr\{\tilde{\varphi}[Q'_w, Q_{-w}](w)R_w v\} \geq \sum_{Q_{-w}} \sigma_{-w}(Q_{-w}) \cdot \Pr\{\tilde{\varphi}[Q_w, \sigma_{-w}](w)R_w v\} = \Pr\{\tilde{\varphi}[Q_w, \sigma_{-w}](w)R_w v\}$ and Q'_w stochastically P_w-dominates Q_w against a completely mixed strategy profile σ_{-w} in the game induced by $\tilde{\varphi}$. ∎

6 Further Research

As mentioned in the Introduction, the aim of this paper is to narrow the set of potential equilibrium outcomes by imposing stronger rationality constraints than those underlying the concept of Nash equilibrium. Nevertheless, the analysis performed here should be considered very preliminary. We have shown that only truth telling may be a best reply to a completely mixed strategy profile and, thus, part of an ordinal perfect equilibrium. Such negative result on the existence of ordinal perfect equilibria calls for a weaker concept. One possible course of action lies in considering that agents never submit strategies that are stochastically dominated against a completely mixed strategy profile, while those that are not stochastically dominated should be regarded as potential choices. Such concept is closer in spirit to the notion of perfect equilibrium in expected utilities. In fact, each of the latter strategies should be a best reply to a completely mixed strategy profile for some utility representation of the true preferences. We can then show that strategies that do not list the best partner first are stochastically dominated and conclude that a strict subset of the individually rational matchings can be sustained in weak ordinal perfect equilibria.

References

Barberà, S., and B. Dutta (1995): "Protective Behavior in Matching Models," *Games and Economic Behavior*, 8, 281–296.

d'Aspremont, C., and B. Peleg (1988): "Ordinal Bayesian Incentive Compatible Representations of Committees," *Social Choice and Welfare*, 5, 261–279.

Dubins, L. E., and D. A. Freedman (1981): "Machiavelli and the Gale-Shapley Algorithm," *American Mathematical Monthly*, 88, 485–494.

Ehlers, L. (2004): "In Search of Advice for Participans in Matching Markets which Use the Deferred-Acceptance Algorithm," *Games and Economic Behavior*, 48, 249–270.

Ehlers, L., and J. Massó (2003): "Incomplete Information and Small Cores in Matching Markets," Mimeo, Universitat Autònoma de Barcelona.

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

Gale, D., and M. A. O. Sotomayor (1985): "Ms Machiavelli and the Stable Matching Problem," *American Mathematical Monthly*, 92, 261–268.

Knuth, D. E. (1976): *Marriages Stables*. Les Presses de l'Université de Montréal, Montréal.

Majumdar, D. (2003): "Ordinally Bayesian Incentive Compatible Stable Matchings," Mimeo.

Majumdar, D., and A. Sen (2004): "Ordinally Bayesian Incentive Compatible Voting Rules," *Econometrica*, 72, 523–540.

McVitie, D. G., and L. B. Wilson (1970): "Stable Marriage Assignment for Unequal Sets," *BIT*, 10, 295–309.

Pais, J. (2004a): "Incentives in Decentralized Random Matching Markets," Mimeo, Universitat Autònoma de Barcelona.

——— (2004b): "On Random Matching Markets: Properties and Equilibria," Mimeo, Universitat Autònoma de Barcelona.

——— (2004c): "Random Matching in the College Admissions Problem," Mimeo, Universitat Autònoma de Barcelona.

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

——— (1984a): "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory," *Journal of Political Economy*, 92, 991–1016.

——— (1984b): "Misrepresentation and Stability in the Marriage Problem," *Journal of Economic Theory*, 34, 383–387.

——— (1989): "Two-Sided Matching with Incomplete Information about Others' Preferences," *Games and Economic Behavior*, 1, 191–209.

Roth, A. E., and U. G. Rothblum (1999): "Truncation Strategies in Matching Markets - In Search of Advice for Participants," *Econometrica*, 67, 21–43.