ECOLE POLYTECHNIQUE

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Ordinally Bayesian Incentive Compatible Stable Matching

Dipjyoti Majumdar

Décembre 2003

Cahier n° 2003-028

LABORATOIRE D'ECONOMETRIE

1rue Descartes F-75005 Paris

(33) 1 55558215

http://ceco.polytechnique.fr/

mailto:labecox@poly.polytechnique.fr

Ordinally Bayesian Incentive Compatible Stable Matching

Dipjyoti Majumdar[1]

Décembre 2003

Cahier n° 2003-028

Résumé: Nous étudions des questions d'incitation dans le cadre des problèmes de marriage, en remplassant la notion de non-manipulabilité par celle de Compatibilité Incitative Bayésienne Ordinale (OBIC). Cette condition exige que dire la vérité maximise l'utilité espérée calculée par rapport à la loi a priori de chaque agent et sous l'hypothèse que les autres agents disent la vérité. Nous montrons que, sans restriction sur les préférences, il n'existe aucune procédure stable et OBIC. On suppose ensuite que les préférences sont telles que rester célibataire est la pire option pour chaque agent. Dans ce cas, si les probabilités a priori sont uniformes, les mariages générés par les algorithmes d'acceptation différée sont OBIC. Cependant, pour des lois a priori génériques, il n'existe pas de procédures stables et OBIC, même pour des préférences restreintes.

Abstract: We study incentive issues related to the two-sided one-to-one stable matching problem after weakening the notion of strategy-proofness to Ordinal Bayesian Incentive Compatibility (OBIC). Under OBIC, truthtelling is required to maximize expected utility of every agent, expected utility being computed with respect to the agent's prior and under the assumption that everybody else is also telling the truth. We show that when preferences are unrestricted there exists no matching procedure that is both stable and OBIC. Next preferences are restricted to the case where remaining single is the worst alternative for every agent. We show that in this case, if agents have uniform priors the stable matching generated by the "deferred acceptance algorithms" are OBIC. However, for generic priors there are no procedures that are both stable and OBIC even with restricted preferences.

Mots clés : Marriage stable, incitation, manipulabilité.

Key Words : Stable matching, incentives, strategy-proofness.

Classification JEL: C72, D72

[2]

[1] CORE-UCL, 34 Voie du Roman Pays, B-1348 Louvain La Neuve, Belgique - majumdar@core.ucl.ac.be
Laboratoire d'Econométrie, Ecole polytechnique

Ordinally Bayesian Incentive Compatible Stable Matchings

Dipjyoti Majumdar*

August, 2003

Abstract

We study incentive issues related to two-sided one-to-one stable matching problem after weakening the notion of strategy-proofness to *Ordinal Bayesian Incentive Compatibility (OBIC)*. Under OBIC, truthtelling is required to maximize the expected utility of every agent, expected utility being computed with respect to the agent's prior beliefs and under the assumption that everybody else is also telling the truth. We show that when preferences are unrestricted there exists no matching procedure that is both stable and OBIC. Next preferences are restricted to the case where remaining single is the worst alternative for every agent. We show that in this case, if agents have uniform priors then the stable matchings generated by "deferred acceptance algorithms" are OBIC. However, for generic priors there are no matching procedures that are both stable and OBIC even with restricted preferences.

*CORE-UCL, 34 Voie du Roman Pays B-1348 Louvain La Neuve, Belgium. e-mail: majumdar@core.ucl.ac.be

1 Introduction

The objective of this paper is to explore issues in incentives related to matching problems and design of matching procedures.

Matching problems refer to problems which involve matching members of one set of agents to members of a second, disjoint set of agents all of whom have preferences over the possible resulting matches. We focus attention on two-sided, one-to-one matching where each agent is matched to at most one mate. A fundamental notion in this context is a *stable matching* which can be defined as a matching such that there does not exist a pair of agents who would prefer to be matched to each other than to their current partners. Such a matching is in the core of the corresponding cooperative game which would result if individual agents were able to freely negotiate their own matches. Gale and Shapley (1962) show that the set of stable matchings is non-empty.

In the strategic version of the model the preferences of the agents are private information. Therefore any stable matching is computed on the basis of the reported preferences. The agents know that by reporting different preferences they can alter the stable matching that is selected and hence change their mate. A natural question which arises is whether *matching procedures* can be designed which give the agents incentive to truthfully reveal their preferences in equilibrium, and which produce stable matchings. The truth-telling concept preponderant in the literature is *strategy-proofness*. Under strategy-proofness it is a dominant strategy for all the agents to truthfully reveal their preferences. The question is does there exist a stable matching procedure that is strategy-proof. Roth (1982) demonstrates that there does not exist any matching procedure which is strategy-proof and which also generates stable matching at every profile of preferences. This result is similar in spirit to a number of impossibility results present in the social choice literature, in the context of designing non-dictatorial social choice procedures which operate

in fairly unrestricted domains (Gibbard (1973), Satterthwaite(1975)).

In this paper we weaken the truth-telling requirement from strategy-proofness to *ordinal Bayesian incentive compatibility (OBIC)*. This notion was introduced in d'Aspremont and Peleg (1988) in the context of a different problem, that of representation of committees. It has also been analysed in standard voting environments in Majumdar and Sen (2003). Truth-telling is required to maximize the expected utility of each individual where expected utility is computed with reference to the individual's *prior beliefs* about the (possible) preferences of other individuals and based on the assumption that other individuals follow the truth-telling strategy. However, this truth-telling notion has one important difference with the standard notion of Bayesian incentive compatibility used widely in incentive theory (for example auction theory). Under OBIC truth-telling is required to maximize expected utility for *every* representation of an individual's true preference ordering. Roth (1989) applies the notion of Bayesian incentive compatibility to the stable matching problem. He generalizes the Roth (1982) result to the case where truth-telling is a Bayes-Nash equilibrium of the revelation game. However, he assumes particular cardinalization of utilities and makes specific assumptions about priors. Since stable matchings only considers preferences and since individual preferences are ordinal, a more appropriate equilibrium notion would be ordinal Bayesian incentive compatibility.

Even though ordinal Bayesian Incentive Compatibility is a significant weakening of the truth-telling requirement, our first result is that there does not exist any prior such that there exists a stable matching procedure that is ordinally Bayesian incentive compatible with respect to it.

Our next step is to look for possibility results by putting restrictions on the set of allowable preferences of the agents. Alcalde and Barberá (1994) look at possibility results by restricting the set of allowable preferences but maintaining strategy-proofness as the notion of truth-telling. We restrict attention to the class

of preferences where each agent prefers to be matched than to remain single and show that when each individual's belief about the preferences of others is uniformly and independently distributed then there *exist* stable matching procedures that are ordinally Bayesian incentive compatible. In a recent paper, Roth and Rothblum (1999) consider stable matching in an incomplete information environment where agents have what they call "symmetric beliefs". If beliefs are uniform then they are symmetric. Roth and Rothblum discuss stochastic dominance of one strategy over others in such an environment. They show that if the stable matching procedure is the *man proposing deferred acceptance algorithm* then for any woman with symmetric beliefs any strategy that changes her true preference ordering of men is stochastically dominated by a strategy that states the same number of acceptable men in their correct order. This latter strategy is called a *Truncation strategy*. Basically, a truncation strategy for a man or a woman is a preference ordering which is order-consistent with his or her true preference but has fewer acceptable men or women. Detailed analysis of truncation strategies can be found in Roth and Vande Vate (1991), Roth and Rothblum (1991), Ma (2002) among others. Roth and Peranson (1991) contains an empirical study using truncation strategies. Ehlers (2001) gives an alternative condition to the symmetry condition on beliefs that leads to the same result. However, neither Roth and Rothblum paper nor the Ehlers paper analyse equilibrium behaviour of agents. Our possibility result with uniform priors follows immediately from and can be seen to be an equilibrium interpretation of the Roth and Rothblum (1999) and the Ehlers (2001) results. One remark needs to be made here. Our result with uniform priors is an existence result. There are a number of distributions that satisfy the criterion of symmetric beliefs. Uniform beliefs is just one of them.

Our main result in this paper is to show that this possibility result is non-generic. We assume common independently distributed prior for all individuals and show that for each individual i there exists a set of conditional beliefs $\mathcal{C}_i$ which

is open and dense in the set of all conditional beliefs and whose complement set is of Lebesgue measure zero, such that *no* stable matching procedure exists that is ordinally Bayesian incentive compatible with respect to a prior belief μ such that the conditionals generated by μ lie in $\mathcal{C}_i$.

The paper is organized as follows. In section 2 we set out the basic notation and definitions. Section 3 deals with the case of unrestricted preferences. In section 4 we consider restricted preferences. In subsection 4.1 we deal with uniform priors while subsection 4.2 considers generic priors. Section 5 concludes. Appendix A contains the deferred acceptance algorithm while Appendix B briefly discusses symmetric beliefs.

2 Preliminaries

We assume that there are two disjoint sets of individuals which we refer to as the set of men and women. These sets are denoted by M and W respectively. Elements in M are denoted by m, m' etc and elements is W are denoted by w, w' etc. Let $I \equiv M \cup W$ denote the entire set of agents. Each man $m \in M$ has a preference ordering P_m over the set $W \cup \{m\}$. Let $\mathcal{P}_m$ be the set of all possible preference orderings for man m. Each woman w has a preference ordering P_w over the set $M \cup \{w\}$. Let $\mathcal{P}_w$ denote the set of all possible preference orderings for woman w. We denote by $P = ((P_m)_{m \in M}, (P_w)_{w \in W})$, a preference profile for all the agents. Let $\mathcal{P} = \times_{i \in I} \mathcal{P}_i$ denote the set of all such preference profiles. We assume that these orderings are strict. We denote by P_{-i} the collection of preferences for all agents other than i. The set of all such P_{-i}'s is denoted by $\mathcal{P}_{-i} = \times_{j \neq i} \mathcal{P}_j$.

We will usually describe an agent's preferences by writing only the ordered set of people that the agent *weakly* prefers to remaining single. Thus the preference P_m described below,

$$P_m := w_1 P_m w_2 P_m m P_m, \cdots, P_m w_k$$

will be abbreviated to,

$$P_m := w_1 P_m w_2 P_m m$$

For reasons that will be obvious shortly, it will suffice only to consider these abbreviated preferences.

DEFINITION 2.1 *A matching is a function $\nu : I \to I$ satisfying the following properties:*

- $\nu(m) \in W \cup \{m\}$
- $\nu(w) \in M \cup \{w\}$
- $\nu(\nu(i)) = i \qquad \forall i \in I$

We now define a *stable matching*. Let $A(P_i) = \{j \in I | j P_i i\}$ denote the set of acceptable mates for agent i. Obviously for a man m with preference ordering P_m, $A(P_m) \subseteq W$ and similarly for a woman w with preference P_w, $A(P_w) \subseteq M$.

DEFINITION 2.2 *A matching ν is stable if the following two conditions are satisfied*

- *for all $i \in I$, $\nu(i) \in A(P_i) \cup \{i\}$*
- *there does not exist $(m, w) \in M \times W$ such that $w P_m \nu(m)$ and $m P_w \nu(w)$*

Let $\mathcal{S}(P)$ denote the set of stable matches under P. Gale and Shapley (1962) shows that $\mathcal{S}(P)$ is always non-empty for all $P \in \mathcal{P}$.

Let $\mathcal{M}$ denote the set of all possible matchings. A *matching procedure* is a mapping that associates a matching with every preference profile P.

DEFINITION 2.3 *A matching procedure is a function $f : \mathcal{P} \to \mathcal{M}$*

If f is a matching procedure and P is a profile, then $f_i(P)$ denotes the match for i selected by f under P.

A *stable* matching procedure f selects an element from the set $\mathcal{S}(P)$ for every $P \in \mathcal{P}$. The rest of the essay is concerned only with stable matching procedures.

We now look at strategic issues in the model. In the strategic version of this problem each agent's preference over his/her possible mates is private information. A question of fundamental interest is the following: does there exist a stable, strategy-proof matching procedure? The answer is negative.

DEFINITION 2.4 *A matching procedure f is strategy-proof if there does not exist $i \in I$, $P_i, P_i' \in \mathcal{P}_i$, and $P_{-i} \in \times_{j \neq i} \mathcal{P}_j$ such that*

$$f_i(P_i', P_{-i}) P_i f_i(P_i, P_{-i})$$

THEOREM **2.1** *Roth (1982)*

A stable and strategy-proof matching procedure does not exist.

In this paper, we explore the consequences of weakening the incentive requirement for stable matching procedures from strategy-proofness to *ordinal Bayesian incentive compatibility*. This concept originally appeared in d'Aspremont and Peleg (1988) and we describe it formally below.

DEFINITION 2.5 *A belief for an individual i is a probability distribution on the set $\mathcal{P}$, i.e. it is a map $\mu_i : \mathcal{P} \to [0,1]$ such that $\sum_{P \in \mathcal{P}} \mu_i(P) = 1$.*

We assume that all individuals have a common prior belief μ. For all μ, for all P_{-i} and P_i, we shall let $\mu(P_{-i}|P_i)$ denote the conditional probability of P_{-i} given P_i.

Consider a man m. The utility function $u_m : W \cup \{m\} \to \Re$ represents $P_m \in \mathcal{P}_m$, if and only if for all $i, j \in W \cup \{m\}$,

$$iP_m j \Leftrightarrow u_m(i) > u(j)$$

The utility function u_w for a woman w is similarly defined.

For any agent $i \in I$ we will denote the set of utility functions representing P_i by $\mathcal{U}_i(P_i)$.

We can now define the notion of incentive compatibility that we use in the essay.

DEFINITION 2.6 *A matching procedure f is ordinally Bayesian Incentive Compatible (OBIC) with respect to the belief μ if for all $i \in I$, for all $P_i, P_i' \in \mathcal{P}_i$, for all $u_i \in \mathcal{U}(P_i)$, we have*

$$\sum_{P_{-i} \in \mathcal{P}_{-i}} u_i\left(f_i(P_i, P_{-i})\right) \mu(P_{-i}|P_i) \geq \sum_{P_{-i} \in \mathcal{P}_{-i}} u_i\left(f_i(P_i', P_{-i})\right) \mu(P_{-i}|P_i) \qquad (1)$$

Let f be a matching procedure and consider the following game of incomplete information as formulated in Harsanyi (1967). The set of types for a player i is $\mathcal{P}_i$ which is also the set from which i chooses an action. If player i's type is P_i, and if the action tuple chosen by the players is P', then player i's payoff is $u(f(P'))$ where u is a utility function which represents P_i. Player i's beliefs are given by the probability distribution μ. The matching procedure is OBIC if truth-telling is a Bayes-Nash equilibrium of this game. Since matching procedures under consideration are ordinal by assumption there is no "natural" utility function for expected utility calculations. Under these circumstances OBIC requires that a player cannot gain in expected utility (conditional on type) by unilaterally misrepresenting his preferences no matter what utility function is used to represent his true preferences.

It is possible to give an alternative definition of OBIC in terms of stochastic dominance. Let f be a matching procedure and pick an individual i and a preference ordering P_i. Suppose that j is the first-ranked mate for i under P_i. Let α

denote the probability conditional on P_i that i is matched with j when i announces P_i assuming that other agents are truthful as well. Thus α is the sum of $\mu(P_{-i}|P_i)$ over all P_{-i} such that $f_i(P_i, P_{-i}) = j$ (that is, i is matched to j). Similarly, let β be the probability that i is matched to j when i announces P_i', i.e., β is the sum of $\mu(P_{-i}|P_i)$ over all P_{-i} such that $f_i(P_i', P_{-i}) = j$. If f is OBIC with respect to μ we must have $\alpha \geq \beta$. Suppose this is false. Consider now a utility function that gives a utility of one to j (i's top-ranked mate under P_i) and virtually zero to all other possible mates for i. This utility function will represent P_i and the expected utility from announcing the truth for agent i with preferences P_i is strictly lower than from announcing P_i'. Using a similar argument, it follows that the probability of obtaining the first k ranked mates according to P_i under truth-telling must be at least as great as under misreporting via P_i'. We make these ideas precise below.

For any agent $i \in I$, let I_i be the set of possible mates for i. Thus if $i \equiv m \in M$, then $I_i = W \cup \{m\}$ and if $i \equiv w$ then $I_i = M \cup \{w\}$. For all $P_i \in \mathcal{P}_i$ and $k = 1, \cdots, |I_i|$, let $r_k(P_i)$ denote the k th ranked mate in P_i, i.e., $r_k(P_i) = j$ implies that $|\{l \neq j | l P_i j\}| = k - 1$. For all $i \in I$, for any $P_i \in \mathcal{P}_i$ and for any $j \in I_i$, let $B(j, P_i) = \{l \in I_i | l P_i j\} \cup \{j\}$. Thus $B(j, P_i)$ is the set of mates that are weakly preferred to j under P_i.

The stable matching procedure f is OBIC with respect to the belief μ if for all $i \in I$, for all integers $k = 1, \cdots, |I_i|$ and for all P_i and P_i',

$$\mu(\{P_{-i} | f_i(P_i, P_{-i}) \in B(r_k(P_i), P_i)\} | P_i) \geq \mu(\{P_{-i} | f_i(P_i', P_{-i}) \in B(r_k(P_i), P_i)\} | P_i) \tag{2}$$

A similar definition of OBIC appears in Majumdar and Sen (2003). We omit the proof of the equivalence of the two definitions of OBIC. The proof is easy and we refer the interested reader to Theorem 3.11 in d'Aspremont and Peleg (1988).

3 The Case of Unrestricted Preferences

The main result of this section is to show that there does not exist any stable marriage procedure that is OBIC with respect to any prior belief μ[1]. In an earlier paper, Roth (1989) extends the analysis of Roth (1982) by weakening the truth-telling requirement to Bayesian incentive compatibility. However, he assumes particular cardinalization of utilities. The paper shows that there exists specific utility values and probability distributions for which no stable matching procedure is Bayesian incentive compatible. The paper therefore, does not rule out the possibility that there may exist utility profiles and probability distributions for which there exist Bayesian incentive compatible stable procedures. However, since stable matchings are based only on ordinal preferences, it is possible to argue that OBIC is a more appropriate equilibrium notion. We have the following strong negative result.

THEOREM **3.1** *Let $|M|$, $|W| \geq 2$ and assume that there are no restrictions on the preferences of individuals. Then for any prior belief μ, there does not exist a stable matching procedure f such that f is OBIC with respect to μ.*

Let f be a stable matching procedure. We first establish a lemma which says the following: consider an agent $i \in I$ and two preference orderings P_i and P_i' such that $r_1(P_i) = r_1(P_i') = j$. However under preference ordering P_i' agent i prefers to remain single than to be matched to any agent other than j. Lemma 3.1 shows that if for some combination of others preferences P_{-i}, f picks j to be i's mate when i reports P_i, then f should pick j as i's mate when i reports P_i'. Formally we show the following:

Lemma 3.1 *Consider an agent $i \in I$ and two preferences P_i and P_i' such that $r_1(P_i) = r_1(P_i') = j$ and $r_2(P_i') = i$. Then for any $P_{-i} \in \mathcal{P}_{-i}$,*

[1]The result holds even if we do away with the assumption of common priors

$$[f_i(P_i, P_{-i}) = j] \Rightarrow [f_i(P_i', P_{-i}) = j]$$

PROOF: It follows from the definition of stable matching that $f_i(P_i', P_{-i}) \in \{j, i\}$. Suppose that $f_i(P_i', P_{-i}) = i$. Observe that for agent j, $i \in A(P_j) \cup \{j\}$. Also since the preferences for all the agents other than i have not changed, we claim that any k such that kP_ji will not be matched to j under the preference profile (P_i', P_{-i}). Suppose that the claim is not true and suppose that there exists a k with kP_ji such that, $k = f_j(P_i', P_{-i})$. Since f is a stable matching procedure it follows that $f_k(P)P_kj$, otherwise (k, j) would have blocked the matching selected by f under the profile P. Let $l = f_k(P)$. Replicating the arguments above one can show that $f_l(P_i', P_{-i})P_lf_l(P) = k$. Otherwise k and l would block the matching $f(P_i', P_{-i})$. Let $f_l(P_i', P_{-i}) = k' \neq k$. Observe that $k' \neq i$ for in the matching $f(P_i', P_{-i})$, i is remaining single. Again by analogous arguments it follows that $f_{k'}(P)P_{k'}f_{k'}(P_i', P_{-i}) = l$. Thus there exists a sequence of pairs $\{(k_n, l_n)|n = 1, 2, 3, \cdots\}$ where any two pairs are distinct (i.e., for any n_1 and n_2, $k_{n_1} \neq k_{n_2}$ and $l_{n_1} \neq l_{n_2}$) such that $k_1 = i$, $l_1 = j$ and

$$l_n = f_{k_n}(P)P_{k_n}f_{k_n}(P_i', P_{-i}) = l_{n+1} \text{ and}$$
$$k_{n+1} = f_{l_n}(P_i', P_{-i})P_{l_n}f_{l_n}(P) = k_n$$

Since I is finite there exists a $n^\star$ such that,

$$l_{n^\star} = f_{k_{n^\star}}(P)P_{k_{n^\star}}f_{k_{n^\star}}(P_i', P_{-i}) \text{ and,}$$

there does not exist a $\hat{k} \in I \setminus \{k_n\}_{n=1}^{n^\star}$ such that $\hat{k}P_{l_{n^\star}}f_{l_{n^\star}}(P)$. Then $(k_{n^\star}, l_{n^\star})$ will block the matching $f(P_i', P_{-i})$. Therefore it follows that any k such that kP_ji will not be matched to j under (P_i', P_{-i}). This proves the claim. Therefore if $f_i(P_i', P_{-i}) = i$ it implies that for agents i and j,

$$jP_i'f_i(P_i', P_{-i}) = i \text{ and}$$
$$iP_jf_j(P_i', P_{-i})$$

Then f is not a stable matching procedure. We thus have a contradiction. Therefore $f_i(P'_i, P_{-i}) = j$. ∎

PROOF OF THEOREM 3.3.1 Let f be OBIC with respect to μ. Pick $i \in I$ and preferences P_i and P'_i. From (3.2) we get,

$$\mu(\{P_{-i}|f_i(P_i, P_{-i}) = r_1(P_i)\}|P_i) \geq \mu(\{P_{-i}|f_i(P'_i, P_{-i}) = r_1(P_i)\}|P_i) \qquad (3)$$

Consider a preference profile P such that, $P_{m_1} := w_1 P_{m_1} w_2 P_{m_1} m_1$; $P_{m_2} := w_2 P_{m_2} w_1 P_{m_2} m_2$; $P_{w_1} := m_2 P_{w_1} m_1 P_{w_1} w_1$; $P_{w_2} := m_1 P_{w_2} m_2 P_{w_2} w_2$; also let $P_j := j$ for all $j \in I \setminus \{m_1, m_2, w_1, w_2\}$. It is easy to check that $\mathcal{S}(P)$ consists of two matchings ν_1 and ν_2 where $\nu_1(m_1) = w_1$, $\nu_1(m_2) = w_2$, $\nu_1(j) = j$ for all $j \in I \setminus \{m_1, m_2, w_1, w_2\}$, $\nu_2(m_1) = w_2$, $\nu_2(m_2) = w_1$ and $\nu_2(j) = j$ for all $j \in I \setminus \{m_1, m_2, w_1, w_2\}$. Suppose $f(P) = \nu_1$. Now consider $P'_{w_2} := m_1 P'_{w_2} w_2$. Then we claim that the only stable matching in the profile (P'_{w_2}, P_{-w_2}) is ν_2. Suppose $f(P'_{w_2}, P_{-w_2}) = \nu$. Note that $\nu(w_2)$ is either m_1 or w_2. Suppose $\nu(w_2) = w_2$. Then either $\nu(m_1) = m_1$ or $\nu(m_2) = m_2$. If $\nu(m_1) = m_1$, then (m_1, w_2) blocks ν. Therefore $\nu(m_1) = w_1$ and $\nu(m_2) = m_2$. Then (m_2, w_1) blocks ν. Therefore $\nu(w_2) = m_1$ and $\nu(w_1) = m_2$. But then $\nu = \nu_2$. Since $f_{w_2}(P) = m_2$, $f_{w_2}(P'_{w_2}, P_{-w_2}) = m_1$ and $r_1(P_{w_2}) = m_1$, it must be the case in order for (3.3) to hold that there exists $\tilde{P}_{-w_2}$ such that $f_{w_2}(P_{w_2}, \tilde{P}_{-w_2}) = m_1$ and $f_{w_2}(P'_{w_2}, \tilde{P}_{-w_2}) \neq m_1$. But from Lemma 3.3.1 this will never be the case. Thus $f(P) \neq \nu_1$. Therefore $f(P) = \nu_2$. Now consider $P'_{m_1} := w_1 P'_{m_1} m_1$. The only stable matching under the profile (P'_{m_1}, P_{-m_1}) is ν_1. By replicating the earlier argument it follows that if $f(P'_{m_1}, P_{-m_1}) = \nu_1$ then $f(P)$ can never be ν_2. But this is a contradiction. This completes the proof of the theorem. ∎

The result in this section assumes unrestricted preferences, i.e., each man m is allowed to have any ordering over the set $W \cup \{m\}$ and similarly each woman w is

allowed to have any ordering over the set $M \cup \{w\}$. Alcalde and Barberà (1994) put restrictions on preferences to obtain strategy-proof stable matchings. In the next section we put weaker restrictions on preferences to see whether possibility results with OBIC can be obtained.

4 Restricted Preferences

In this section we examine the stable matching problem for a special class of preferences. We restrict our attention to the class of preferences where remaining single is the worst alternative for every agent. That is, each agent prefers to be matched to some other agent than to remain single.

Formally, the domain $\mathcal{D}$ consists of *all* preferences (P_m, P_w) satisfying the following conditions:

- for all $w_i \in W$, $w_i P_m m$
- for all $m_i \in M$, $m_i P_w w$

In this environment a stable matching procedure is a function $f : \mathcal{D} \to \mathcal{M}$ with the restriction that $f(P) \in \mathcal{S}(P)$ for all $P \in \mathcal{D}$. We denote by $\mathcal{D}_{-i}$ the set of all P_{-i}'s, where P_{-i} is the collection of preferences of all agents other than i.

The man proposing and the woman proposing *deferred acceptance* algorithms are ways to obtain a stable matching given the preference reports of men and women. Both algorithms are discussed in Appendix A.

Let $f^{DA(m)}$ denote the stable matching procedure that uses the man proposing deferred acceptance algorithm and let $f^{DA(w)}$ denote the woman proposing deferred acceptance algorithm. Roth (1982) demonstrates that with the man proposing deferred acceptance algorithm it is a dominant strategy for men to truthfully reveal their preferences i.e., it is strategy-proof for men. Since men and women are

symmetric in this model, the woman proposing deferred acceptance algorithm is strategy-proof for women.

THEOREM **4.1** *Roth (1982)*

The stable matching procedure $f^{DA(m)}$, is strategy-proof for men. Similarly, $f^{DA(w)}$ is strategy-proof for women.

4.1 Uniformly and Independently Distributed Priors

In this section, we assume that the beliefs are independently and uniformly distributed.

DEFINITION 4.1 *Individual i's beliefs are independent if for all $k = 1, \cdots, |I|$ there exist probability distributions $\mu_k : \mathcal{P}_k \to [0, 1]$ such that, for all P_{-i} and P_i,*

$$\mu(P_{-i}|P_i) = \times_{k \neq i} \mu_k(P_k)$$

An individual's belief is independent if his conditional belief about the types of the other individuals is a product measure of the marginals over the types of the other individuals. We also assume that the beliefs are uniform.

DEFINITION 4.2 *For all profiles $P, P' \in \mathcal{P}$, we have*

$$\mu(P) = \mu(P')$$

We denote these independent, uniform priors by $\bar{\mu}$. Restating Definition 3.6 in the present context, we have

PROPOSITION 4.1 *The matching procedure f is OBIC with respect to the belief $\bar{\mu}$ if, for all i, for all integers $k = 1, \cdots, |I_i|$, for all P_i and P_i', we have*

$$|\{P_{-i}|f_i(P_i, P_{-i}) \in B(r_k(P_i), P_i)\}| \geq |\{P_{-i}|f_i(P_i', P_{-i}) \in B(r_k(P_i), P_i)\}| \qquad (4)$$

We omit the trivial proof of this Proposition.

Roth and Rothblum (1999) define a particular type of belief for agents which they call "symmetric" beliefs. Symmetric beliefs are discussed in Appendix B. We note that independent, uniform beliefs are symmetric. They show that if the stable matching procedure is $f^{DA(m)}$ then for a woman with symmetric beliefs, a strategy that changes her true preference ordering of men is stochastically dominated by a strategy that states the same number of acceptable men in their correct order, i.e., in the order of the true preference ordering. The same is true for men when the matching procedure is $f^{DA(w)}$. The following theorem can be treated as an equilibrium interpretation of the Roth and Rothblum results.

THEOREM **4.2** *The stable marriage procedures* $f^{DA(m)} : \mathcal{D} \to \mathcal{M}$ *and* $f^{DA(w)} : \mathcal{D} \to \mathcal{M}$ *are OBIC with respect to the uniform prior.*

PROOF: We give the proof for $f^{DA(m)}$. The proof for $f^{DA(w)}$ is analogous. From Theorem 4.1 we know that $f^{DA(m)}$ is strategy-proof for men. So we only need to check whether $f^{DA(m)}$ is OBIC with respect to the uniform prior for women. Observe that if any $w \in W$ has uniformly and independently distributed prior belief then her conditional belief is $\{M\}$-symmetric (the concept of $\{M\}$-symmetry is discussed in Appendix B). So Proposition 7.1 (again we refer the reader to Appendix B) applies and hence any strategy that changes her true preference ordering of men is stochastically dominated by a strategy that states the same number of acceptable men in their correct order. However, when preference profiles are in $\mathcal{D}$, for any $w \in W$ with preference order P_w, the only strategy that states w's set of acceptable men in their correct order is P_w. Since OBIC is equivalent to the stochastic domination of truth-telling this proves the theorem. ■

4.2 Generic Priors

The main result in this section is to show that the possibility result of the previous section vanish if the beliefs are slightly perturbed. We continue to assume first that the beliefs are independent.

For each agent i, we let $\Delta(i)$ denote the set of all beliefs over the possible types of i. If i is a man, $\Delta(i)$ is a unit simplex of dimension $(|W|+1)!-1$. If i is a woman, $\Delta(i)$ is a unit simplex of dimension $(|M|+1)!-1$. The set of all independent priors $\Delta^I = \times_{i\in I}\Delta(i)$. For an agent i and belief $\mu \in \Delta^I$, we shall let μ_{-i} i's conditional belief over the types of agents other than i. For instance $\mu_{-i}(P_{-i})$ will denote the probability under μ that the preferences of agents other than i, is P_{-i}. The set of all such conditional beliefs will be denoted by Δ^{CI}. Clearly, $\Delta^{CI} = \times_{k\neq i}\Delta(k)$.

We now state the main result of this section.

THEOREM **4.3** *Let $|M| = |W| \geq 3$ and assume that all individuals have independent beliefs. Then for all $i \in I$, there exists a subset $\mathcal{C}_i$ of $\Delta^{CI}(i)$ such that*

- *$\mathcal{C}_i$ is open and dense in $\Delta^{CI}(i)$*
- *$\Delta^{CI}(i) - \mathcal{C}_i$ has Lebesgue measure zero*
- *there does not exist a stable marriage procedure $f : \mathcal{D} \to \mathcal{M}$ that is OBIC w.r.t the belief μ where $\mu_{-i} \in \mathcal{C}_i$ for all $i \in I$.*

PROOF: The proof proceeds in three steps. In Step 1 we define the sets $\mathcal{C}_i$ and show that they are open and dense subsets of $\Delta^{CI}(i)$ and the Lebesgue measure of their complement sets are zero. In Step 2 we show that if a matching procedure f is OBIC with respect to μ with $\mu_{-i} \in \mathcal{C}_i$ for all i, then f must satisfy a certain property which we call *Top Monotonicity(TM)*. In Step 3 we complete the proof by showing that stable matching procedure violates TM.

STEP1:

Pick an individual i. We define the set $\mathcal{C}_i$ below.

For any $Q \subseteq \mathcal{D}_{-i}$, let $\mu_{-i}(Q) = \sum_{P_{-i} \in Q} \mu_{-i}(P_{-i})$. The set $\mathcal{C}_i$ is defined as the set of conditional beliefs μ_{-i} satisfying the following property:For all $Q, T \subseteq \mathcal{D}_{-i}$

$$[\mu_{-i}(Q) = \mu_{-i}(T)] \Rightarrow [Q = T]$$

For any belief μ and agent i the conditional belief $\mu{-}i$ belongs to $\mathcal{C}_i$ if it assigns equal probabilities to two "events" Q and T only if $Q = T$. Obviously the events O and T are defined over preference orderings of individuals other than i. In this step we show that $\mathcal{C}_i$ is open and dense in $\Delta^{CI}(i)$ and that its complement set has Lebesgue measure zero. Observe that $\mathcal{C}_i$ is generic in the space of *conditional probabilities generated by an independent prior*. It is *not* generic in the space of all probability distributions.

We first show that $\mathcal{C}_i$ is open in $\Delta^{CI}(i)$. Consider any μ such that for all $i \in I$, $\mu_{-i} \in \mathcal{C}_i$. Let,

$$\phi(\mu) = \min_{S,T \subset \times_{k \neq i} \mathcal{P}_k,\, S \neq T} |\mu_{-i}(S) - \mu_{-i}(T)|$$

Observe that $\phi(\mu) > 0$. Since ϕ is a continuous function of μ, there exists $\epsilon > 0$ such that for all product measures $\hat{\mu} \in \delta^I$ with $d(\hat{\mu}, \mu)w < \epsilon$,[2] we have $\phi(\hat{\mu}) > 0$. But this implies $\hat{\mu}_{-i} \in \mathcal{C}_i$. Therefore $\mathcal{C}_i$ is open in $\Delta^{CI}(i)$.

We now show that, $\Delta^{CI}(i) - \mathcal{C}_i$ has Lebesgue measure zero. we begin with the observation that $\Delta^{CI}(i) = \times_{k \neq i} \Delta(k)$. That is, $\Delta^{CI}(i)$ is the cartesian product of unit simplices $\Delta(k)$s, and each $\Delta(k)$ is of dimension $(|M|+1)! - 1 = (|W|+1)! - 1$. On the other hand,

$$\Delta^{CI}(i) - \mathcal{C}_i = \bigcup_{Q,T \subset \times_{k \neq i} \mathcal{P}_k} \{\mu \in \Delta^{CI} | \mu_{-i}(Q) = \mu_{-i}(T)\}$$

[2] $d(.,.)$ here signifies Euclidean distance

Therefore the set $\Delta^{CI}(i) - \mathcal{C}_i$ is the union of a finite number of hyper-surfaces intersected with $\Delta^{CI}(i)$. It follows immediately that it is a set of lower dimension and hence has zero Lebesgue measure.

Pick a product measure μ such that for some i, $\mu_{-i} \in \Delta^{CI}(i) - \mathcal{C}_i$ and consider an open neighborhood of radius $\epsilon > 0$ with center μ_{-i}. Since this neighborhood has strictly positive measure and since $\Delta^{CI}(i) - \mathcal{C}_i$ has measure zero, it must be the case that the neighborhood has non-empty intersection with $\mathcal{C}_i$. This establishes that $\mathcal{C}_i$ is dense in $\Delta^{CI}(i)$.

This completes Step 1.

STEP 2:

Let f be a matching procedure that is OBIC with respect to the belief μ where $\mu_{-i} \in \mathcal{C}_i$ for all i. In this step we show that f must satisfy Property TM. Let P be a preference profile, let i be an individual and let P_i' be an ordering such that the top-ranked mate in P_i is the same as the top-ranked element in P_i'. Let us denote this top-ranked mate for i by j. Then property TM requires that if i is matched to j when the reported preference profile is P i.e., $f_i(P) = j$, then i must be matched to j when the reported preference profile is $(P_i', P{-}i)$ i.e., $f_i(P_i', P{-}i) = j$. We give the formal definition below.

Definition 4.3 *The marriage procedure f satisfies TM, if for all individuals i, for all P_{-i} and for all P_i, P_i' such that $r_1(P_i) = r_1(P_i')$, we have*

$$f_i(P_i, P_{-i}) = r_1(P_i) \Rightarrow f_i(P_i', P_{-i}) = r_1(P_i')$$

Let i be an individual and let P_i and P_i' be such that $r_1(P_i) = r_1(P_i')$. Suppose i's "true" preference is P_i. Since f is OBIC with respect to μ, we have, by using equation (3.2)

$$\mu(\{P_{-i} | f_i(P_i, P_{-i}) = r_1(P_i)\}) \geq \mu(\{P_{-i} | f_i(P_i', P_{-i}) = r_1(P_i')\}) \tag{5}$$

Suppose i's true preference is P_i'. Applying equation (3.2) we have

$$\mu(\{P_{-i}|f_i(P_i', P_{-i}) = r_1(P_i')\}) \geq \mu(\{P_{-i}|f_i(P_i, P_{-i}) = r_1(P_i)\}) \tag{6}$$

Combining (3.5) and (3.6) and using the fact that $r_1(P_i) = r_1(P_i')$ we get,

$$\mu(\{P_{-i}|f_i(P_i, P_{-i}) = r_1(P_i)\}) = \mu(\{P_{-i}|f_i(P_i', P_{-i}) = r_1(P_i')\}) \tag{7}$$

Since $\mu(P_{-i}) \in \mathcal{C}_i$ it follows from (3.7) that,

$$\{P_{-i}|f_i(P_i, P_{-i}) = r_1(P_i)\} = \{P_{-i}|f_i(P_i', P_{-i}) = r_1(P_i')\} \tag{8}$$

Thus, if for some P_i $f_i(P_i, P_{-i}) = r_1(P_i)$, then (3.8) implies that $f_i(P_i', P_{-i}) = r_1(P_i')$. Therefore f satisfies TM.

STEP 3: In this step we complete the proof of the theorem by showing that a stable matching procedure does not satisfy TM.

Let $|M| = |W| \geq 3$ and let $f : \mathcal{D} \to \mathcal{M}$ be a stable matching procedure, i.e., for all $P \in \mathcal{D}$, $f(P) \in \mathcal{S}(P)$. Consider a preference profile P defined as follows:

$$\begin{aligned}
P_{m_1} &:= w_2 P_{m_1} w_1 P_{m_1} w_3 P_{m_1}, \cdots, P_{m_1} m_1 \\
P_{m_2} &:= w_1 P_{m_2} w_2 P_{m_2} w_3 P_{m_2}, \cdots, P_{m_2} m_2 \\
P_{m_3} &:= w_1 P_{m_3} w_2 P_{m_3} w_3 P_{m_3}, \cdots, P_{m_3} m_3 \\
P_{w_1} &:= m_1 P_{w_1} m_3 P_{w_1} m_2 P_{w_1}, \cdots, P_{w_1} w_1 \\
P_{w_2} &:= m_3 P_{w_2} m_1 P_{w_2} m_2 P_{w_2}, \cdots, P_{w_2} w_2 \\
P_{w_3} &:= m_1 P_{w_3} m_2 P_{w_3} m_3 P_{w_3}, \cdots, P_{w_3} w_3
\end{aligned}$$

For all $k \neq 1, 2, 3$, $P_{m_k} := w_k P_{m_k}, \cdots, P_{m_k} m_k$ and $P_{w_k} := m_k P_{w_k}, \cdots, P_{w_k} w_k$. We claim that $\mathcal{S}(P) = \{\nu_1, \nu_2\}$ where,

$$\begin{aligned}
\nu_1 &= [(m_1, w_2), (m_2, w_3), (m_3, w_1), (m_k, w_k),\ k \neq 1, 2, 3] \\
\nu_2 &= [(m_1, w_1), (m_2, w_3), (m_3, w_2), (m_k, w_k),\ k \neq 1, 2, 3]
\end{aligned}$$

Observe that , in any stable matching m_2 must be matched with w_3; otherwise either (m_1, w_2) or (m_3, w_1) will block. Given that, there are only two other possible combinations: one where m_1 is matched with w_1 and the other where m_1 is matched to w_2. Both give rise to stable outcomes since there is no pair that will block the matching. Let $f(P) = \nu_1$. Then $f_{w_1}(P) = m_3$. Now consider the preference ordering $\hat{P}_{w_1}$ given by

$$\hat{P}_{w_1} := m_1 \hat{P}_{w_1} m_2 \hat{P}_{w_1} m_3 \hat{P}_{w_1}, \cdots, w_1$$

We claim that $\mathcal{S}(\hat{P}_{w_1}, P_{-w_1}) = \nu_2$. Observe that in any stable matching in the profile $(\hat{P}_{w_1}, P_{-w_1})$, m_3 must be matched to w_2; otherwise, either (m_1, w_1) or (m_3, w_2) will block. Also, m_2 has to be matched to w_3; otherwise, m_1 and w_1 would block the matching. Hence the only stable matching is ν_2. Then $f_{w_1}(\hat{P}_{w_1}, P_{-w_1}) = m_1$. But if $f_{w_1}(\hat{P}_{w_1}, P_{-w_1}) = m_1$ it follows from TM that, $f_{w_1}(P)$ should also be m_1. Hence $f(P) \neq \nu_1$. Therefore, $f(P) = \nu_2$. Now consider a preference ordering for m_1, $\hat{P}_{m_1}$ given by,

$$\hat{P}_{m_1} := w_2 \hat{P}_{m_1} w_3 \hat{P}_{m_1} w_1 \hat{P}_{m_1}, \cdots, \hat{P}_{m_1} m_1$$

Replicating the earlier arguments we conclude that $\mathcal{S}(\hat{P}_{m_1}, P_{-m_1}) = \nu_1$. Then $f_{m_1}(\hat{P}_{m_1}, P_{-m_1}) = w_2$. But then TM implies that $f_{m_1}(P)$ should also be w_2 i.e., $f(P) = \nu_1$. But this is a contradiction for we have shown above that $f(P) \neq \nu_1$. This completes the proof of the theorem. ■

REMARK 3.4.1: The result in Theorem 3.4.3 is valid even when $|M| \neq |W|$. Let $M = \{m_1, \cdots, m_n\}$ and $W = \{w_1, \cdots, w_m\}$. Without loss of generality assume that $m < n$. Consider the preference profile P defined in the following way: for all $k \leq m$, P_{i_k} is defined in the same way as above; for $k > m$, $P_{m_k} := w_3 P_{m_k}, \cdots, P_{m_k} m_k$. Observe that under the preference profile P any selection from the set of stable marriages divides the set of agents into three groups: men

m_1, m_2 and m_3 and women w_1, w_2 and w_3 form matchings among themselves; w_k is matched to m_k for all $3 < k \leq m$ and the remaining set of men are forced to remain single. Now replicating the arguments above we obtain the result in Theorem 3.4.3.

REMARK 3.4.2: When there are only two agents on each side of the market and preferences are restricted to the set $\mathcal{D}$, Alcalde and Barberà (1994) show that the stable matching selections obtained by the man-proposing and woman-proposing deferred acceptance algorithms are both strategy-proof.

5 Conclusion

We have examined the implications of weakening the incentive requirement in the theory of two-sided one-to-one matching from dominant strategies to ordinal Bayesian incentive compatibility. Truth-telling is no longer assumed to be optimal for every conceivable strategy-tuple of the other players. It is only required to maximize expected utility given an agents' prior beliefs about the types of other players and the assumption that these players are following truth-telling strategies. The set of ordinal Bayesian incentive compatible stable matching procedures clearly depends on the beliefs of each agent. However, we show that when preferences are unrestricted, there is no stable matching procedure that is ordinally Bayesian incentive compatible with respect to *any* prior. When we put restrictions on the set of allowable preferences, by requiring that every agent prefers to be matched than to remain single, one obtains possibility results with independently and uniformly distributed priors. However the possibility result is non-generic. If we perturb beliefs we get back the impossibility result.

6 Appendix A: Deferred Acceptance Algorithm

Man Proposing Deferred Acceptance Algorithm

STEP 1: Each man makes an offer to the first woman on his preference list of acceptable women. Each woman rejects the offer of any firm that is unacceptable to her, and each woman who receives more than one acceptable offer rejects all but her most preferred of these which she "holds".

STEP K: Any man whose offer was rejected at the previous step makes an offer to his next choice (i.e., to his most preferred woman among those who have not yet rejected it), so long as there remains an acceptable woman to whom he has not yet made an offer. If a man has already made an offer to all the women he finds acceptable and has been rejected by all of them, then he makes no further offers. Each woman receiving offers rejects any from unacceptable men, and also rejects all but her most preferred among the set consisting of the new offers together with an offer she may have held from the previous step.

STOP: The algorithm stops after any step in which no man's offer has been rejected. At this point, every man is either being matched to some woman or his offer has been rejected by every woman in his list of acceptable women. The output of the algorithm is the matching at which each woman is matched to the man whose offer she is holding at the time the algorithm stops. Women who do not receive any acceptable offer or men who were rejected by all women acceptable to them remain unmatched.

7 Appendix B: Symmetric beliefs

In this section, we briefly discuss symmetric beliefs. For the ensuing analysis some definitions are in order. For a given preference profile, denote by P_S the preference orders of the agents in the subset $S \subseteq I$. Denote by $P_S^{m \leftrightarrow m'}$ the preference orders

of the agents in S obtained from P by switching m and m', i.e., each woman in S exchanges the places of m and m' in her preference list and if m is in S his preference is $P_{m'}$ and if m' is in S its preference is P_m. Note that if woman w's true preferences are given by P_w, then $P_w^{m\leftrightarrow m'}$ is the preference in which she reverses the order of m and m' (but otherwise states her true preferences). Similarly, P_{-w} and $P_{-w}^{m\leftrightarrow m'}$ are assessments by agent w of the preferences of all other agents that are identical except that the roles of m and m' are everywhere reversed.

We model agent w's uncertainty about the about the differences in the preferences of men m and m', and about the other women's preferences for the two men as follows:

DEFINITION 7.1 *Given distinct men m and m' we say woman w's conditional belief $\mu(.|P_w)$ is $\{m, m'\}$ -symmetric if $\mu(P_{-w}|P_w) = \mu(P_{-w}^{m\leftrightarrow m'}|P_w)$.*

Observe that w may know a great deal about m and m' (for example w may know that both men prefer w' to some w''). What w does not know about m and m', if her conditional beliefs are $\{m, m'\}$-symmetric are any *differences* in their preferences, or in other women's preferences between them.

DEFINITION 7.2 *For a woman $w \in W$ and a set of men $U \subseteq M$, we say that w's conditional belief $\mu(.|P_w)$ is $\{U\}$-symmetric if it is $\{m, m'\}$-symmetric for every pair (m, m') of distinct members of U.*

If $U = M$ then woman w's belief is $\{M\}$-symmetric. We can similarly define $\{W\}$-symmetric beliefs for a man $m \in M$.

PROPOSITION 7.1 *(Corollary 1 in Roth and Rothblum (1999))*

For a woman with $\{M\}$-symmetric conditional belief, any strategy that changes the true preference ordering of men is stochastically dominated by a strategy that states the same number of acceptable men in their correct order.

Observe that the uniform prior $\bar{\mu}$ is $\{M\}$-symmetric for the women and $\{W\}$-symmetric for men.

8 References

- Alcalde, J.,and S. Barberà (1994), "Top Dominance and the Possibility of Strategy-proof Stable Solutions to Matching Problems", *Economic Theory*, 4: 417-435.

- d'Aspremont, C., and B. Peleg (1988), "Ordinal Bayesian Incentive Compatible Representation of Committees", *Social Choice and Welfare*, 5:261-280.

- Ehlers, L. (2001), "In Search of Advice for Participants in the Matching Markets which use the Deferred-Acceptance Algorithm" *mimeograph.*

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

- Harsanyi, J. (1967), "Games with Incomplete Information Played by 'Bayesian' Players: I-III", *Management Science*, 14: 159-182, 320-334, 486-502.

- Majumdar, D., and A. Sen (2003), "Ordinally Bayesian Incentive Compatible Voting Schemes" *mimeograph*

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

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

- Roth, A., and J. Vande Vate (1991), "Incentives in Two-sided Matching with Random Stable Mechanisms",*Economic Theory*, 1: 31-44.

- Roth, A., and E. Peranson (1999), "The Redesign of Matching Markets for American Physicians: some engineering aspects of Economic Design",*American Economic Review*, 89: 748-780.

- Roth, A., and U. Rothblum (1999), "Truncation Strategies in Matching Markets – in Search of Advice for Participants", *Econometrica*, 67(1): 21-43.