Preprint de “Incomplete information and singleton cores in matching markets,” de Lars Ehlers i Jordi Massó. *Journal of Economic Theory* 136, 587-600 (2007). Lliurat a Elsevier el setembre de de 2006.

Incomplete Information and Singleton Cores in Matching Markets*

Lars Ehlers† Jordi Massó‡

September 2003 (revised September 2006)

*We are especially grateful to an anonymous referee and an anonymous associate editor for their helpful comments and suggestions. L. Ehlers acknowledges financial support form the SSHRC (Canada) and the FQRSC (Québec). The work of J. Massó is partially supported by the Spanish Ministry of Education and Science through grant SEJ2005-04081, and by the Generalitat de Catalunya, through grant SGR2005-00454 and the Barcelona Economics Program (XREA). This research started when J. Massó was visiting the Université de Montréal (financial support from CIREQ is gratefully acknowledged) and continued when L. Ehlers was visiting the Universitat Autònoma de Barcelona (financial support from XREA is gratefully acknowledged).

†Département de Sciences Économiques and CIREQ, Université de Montréal, C.P. 6128 Succursale Centre Ville, Montréal, Québec H3C 3J7, Canada; e-mail: lars.ehlers@umontreal.ca (Corresponding author)

‡Departament d'Economia i d'Història Econòmica and CODE, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain; e-mail: jordi.masso@uab.es

Abstract

We study ordinal Bayesian Nash equilibria of stable mechanisms in centralized matching markets under incomplete information. We show that truth-telling is an ordinal Bayesian Nash equilibrium of the revelation game induced by a common belief and a stable mechanism if and only if all the profiles in the support of the common belief have singleton cores. Our result matches the observations of Roth and Peranson (1999) in the National Resident Matching Program (NRMP) in the United States: (i) the cores of the profiles submitted to the clearinghouse are small and (ii) while truth-telling is not a dominant strategy most participants of the NRMP truthfully reveal their preferences.
JEL Classification: C78, D81, J44.

Keywords: Matching Market, Incomplete Information, Singleton Core.

1 Introduction

In entry-level professional labor markets new workers search for their first positions at firms. Such markets differ in how they match workers and firms. In a decentralized market, workers and firms are themselves responsible in looking for partners. For example, in the first half of the 20th century the entry-level medical markets in the United States and the United Kingdom were decentralized. This had the effect that hospitals (the firms) were offering promising medical students (or workers) earlier and earlier contracts.[1] By the 1950s students often signed a contract two years before finishing. This caused inefficiencies and subsequent regret among the participants of the entry-level medical market: either the student did not develop as expected and the hospital could have later hired a better physician or the student developed much better than expected and could have gotten a job at a better hospital. Thus, the realized matching was often unstable: some students and hospitals were committed to now unacceptable partners or unmatched pairs were preferring each other to their committed partners. Due to these problems entry-level medical markets in the U.S. were reorganized from the 1950s by centralizing them through the National Resident Matching Program (NRMP). Each year a clearinghouse announces the open positions at each hospital and the finishing medical students which will be available (around 20,000 per year). Salaries are not negotiated and included in the job description. Therefore, each participant's preference is a ranking over his potential partners. Then all participants submit their preference lists to the clearinghouse and a mechanism determines a matching for the submitted lists. In other words, a centralized matching market together with a mechanism induces a preference revelation game. The success of the reorganizations depended on which mechanism was used in determining the matching between students and hospitals. A mechanism is stable if it always selects a stable matching of the declared profile. It has been observed that stable mechanisms

[1] Roth and Xing (1994) and Niederle and Roth (2003) describe other entry-level professional labor markets experiencing unravelling of appointment dates.

perform better than unstable ones.[2]

There is a considerable amount of literature analyzing strategic incentives in centralized matching markets when the submitted lists are common knowledge among the participants. A central result is that no stable mechanism exists for which stating the true preferences is a dominant strategy for every agent under complete information (Roth, 1982). Thus, for any stable mechanism there are situations at which some agents gain by manipulation. Sönmez (1999) showed for general allocation problems with indivisibilities that a mechanism is incentive-compatible (truth-telling is a dominant strategy), Pareto-optimal and individually rational only if for each profile the core is a singleton and the mechanism chooses this allocation. Since a matching market may not have a singleton core and stability implies both individual rationality and Pareto-optimality in our model, Sönmez's result implies in our model Roth's (1982) result.

Roth and Peranson (1999) have examined submitted preference lists by hospitals and students in the National Residents Matching Program for the years 1987, 1993, 1994, 1995, and 1996 and found that the number of stable matchings for the submitted profiles were surprisingly small. To explain this unexpected fact, Roth and Peranson (1999) suggest the following conjecture (they call it a new kind of "core convergence" result):[3] As the size of the market increases, the number of stable matchings becomes smaller provided that each participant only ranks (in his/her reported preference ordering) at most a fixed number of positions (which remains small when the number of participants increase). Moreover, the small size of the core suggests limited ability to benefit from manipulating submitted preferences. Thus, Roth and Peranson (1999) infer that a significant number of participants truthfully reveal their preferences. Under the more realistic context of incomplete information, our paper will show in a

[2]Niederle and Roth (2003) report the existence of about 100 markets and submarkets organized via stable mechanisms and that only 3 of them were abandoned after being used for several years.

[3]It is well-known that in the two-sided, one-to-one matching markets the effective coalitions are only individuals or pairs, and hence, the core coincides with the set of stable matchings.

simplified matching market why participants truthfully reveal their preferences *and* the cores of the submitted lists are small.

In centralized matching markets the common knowledge assumption of the submitted lists is extremely strong. Thus, we will consider preference revelation games induced by a stable mechanism under incomplete information. Agents have a common belief and their beliefs of the others' submitted lists are calculated through Bayes' rule for every realization of an individual preference relation. Any stable mechanism is ordinal, i.e., it determines the stable matching through the submitted ordinal rankings. Thus, truth-telling is a Bayesian Nash equilibrium if for every von Neumann-Morgenstern utility function submitting the induced ordinal ranking maximizes the agent's expected utility in the Bayesian revelation game induced by the common belief and the stable mechanism.[4] This requirement is equivalent to the concept of ordinal Bayesian Nash equilibrium which is based on first-order stochastic dominance in the sense that each agent plays a best response to the others' strategies for every von Neumann-Morgenstern representation. We show in Theorem 1 that truth-telling is an ordinal Bayesian Nash equilibrium in the Bayesian revelation game induced by a common belief and a stable mechanism if and only if the support of the common belief is contained in the set of profiles with singleton core. Our result matches the following of Roth and Peranson (1999) in the NRMP: (i) they observed that the cores of the submitted lists are small and (ii) they conjecture that a significant number of participants truthfully reveal their preferences.

Theorem 1 is the first result for matching problems which relates singleton cores to incentive-compatibility in an incomplete information setup. For matching markets it extends Sönmez (1999) by allowing information to be incomplete and it confirms the importance of singleton cores for incentive-compatibility of stable mechanisms. Because dominant-strategy incentive-compatibility is equivalent to requir-

[4]This notion was introduced by d'Aspremont and Peleg (1988) who call it "ordinal Bayesian incentive-compatibility". Majumdar and Sen (2004) use it to relax strategy-proofness in the Gibbard-Satterthwaite Theorem.

ing Bayesian incentive-compatibility for all common beliefs, for stable mechanisms the need for singleton cores is very robust and persists even if dominant-strategy incentive-compatibility is given up and instead Bayesian incentive-compatibility for one common belief is adopted.

Two recent papers have identified strong but meaningful sufficient conditions on preference profiles under which the core of a matching market is a singleton. Eeckhout (2000) proposes a condition, which is also necessary for markets with a small number of participants, that includes the following two special cases. (1) Vertical heterogeneity, where all firms have identical preferences over workers (for instance, according to the student's grades) and all workers have identical preferences over firms (for instance, according to a public and objective ranking of hospitals). (2) Horizontal heterogeneity, where all agents have different preferences over the other side of the market, but each agent has a different most preferred partner and in addition is the most preferred by this partner. Clark (2003) proposes a (stronger) sufficient condition (called the No Crossing Condition), which is closely related to the well-known Single Crossing Condition.

We also argue why, even under the assumption of a common belief, (1) there are other ordinal Bayesian Nash equilibria in which agents misreport their preferences, and (2) members of couples jointly looking for jobs do not have incentives to misrepresent coordinately their preferences at a truth-telling ordinal Bayesian Nash equilibrium of the game induced by a stable mechanism.

We complement Theorem 1 in Theorem 2 by showing that a list of strategies is an ordinal Bayesian Nash equilibrium in the Bayesian revelation game induced by a belief and a stable mechanism only if for each preference profile in the support of the common belief all agents unanimously agree that the matching selected by the stable mechanism for the declared preference lists is most preferred among all matchings in the core. This suggests a new and additional reason, based on the incomplete information nature of real matching markets, of why stable mechanisms last and why

cores are small.

Our paper is the first complete analysis of equilibria of preference revelation games induced by stable mechanisms when participants have incomplete information about the *ordinal* preferences of all other agents.[5] Roth and Rothblum (1999), Ehlers (2002), and Ehlers (2004) provide advice on the list that a particular worker should submit to the clearinghouse, given her uncertainty about the rankings submitted by the other participants. These papers give advice under different hypothesis on the information structure of the beliefs held by the worker and for different mechanisms. Following the mechanism design literature on direct revelation games under incomplete information we assume that agents have a common belief on the set of all profiles which may limit (as in all games with incomplete information) the applicability of our results. However, note that a priori we do not impose any condition on the common belief (such as symmetry of agents' beliefs or independence). Furthermore we consider a simplified one-to-one version of matching markets. Nevertheless one-to-one matching markets are a reasonable approximation of many-to-one matching markets. Think of each firm representing a department of a hospital and suppose that each department has at most one position for its medical specialty. Each department possesses its own ranking over students. Then Theorem 1 remains unchanged if several departments together are allowed to misrepresent their true preferences, i.e. hospitals cannot misrepresent their preferences such that each department strictly gains. Moreover, one-to-one matching markets may arise for instance in regional medical markets for a certain specialty where each institution has (at most) one position for this speciality.

The paper is organized as follows. Section 2 defines the matching market and preference revelation games. Section 3 introduces incomplete information in these games and ordinal Bayesian Nash equilibrium. Section 4 contains the result for truth-

[5]Roth (1989) contains the first strategic analysis of games with incomplete information (on the profile of *utility functions*) induced by stable mechanisms using expected utility. He shows that the most important results concerning dominant and dominated strategies carry over from complete information to (cardinal) incomplete information whereas results concerning Nash equilibria do not.

telling. Section 5 focuses on general ordinal Bayesian Nash equilibria. Section 6 concludes. The Appendix collects all the proofs.

2 The Matching Market

The *agents* in our market consist of two disjoint sets, the set of *firms* F and the set of *workers* W. Generic agents are denoted by $v \in V \equiv F \cup W$ while generic firms and workers are denoted by f and w, respectively. Each worker $w \in W$ has a strict, transitive, and complete preference relation P_w over $F \cup \{w\}$, and each firm $f \in F$ has a strict, transitive, and complete preference relation P_f over $W \cup \{f\}$. Let $\mathcal{P}_v$ denote the set of all preference relations of agent v. In order to compare (potentially) identical partners of v according to the preference relation P_v we denote by R_v the binary relation where for all $v', \widehat{v} \in V$, $v' R_v \widehat{v}$ means that either $v' = \widehat{v}$ or $v' P_v \widehat{v}$. Given $P_w \in \mathcal{P}_w$ and $v \in F \cup \{w\}$, let $B(v, P_w)$ denote the weak upper contour set of P_w at v; i.e., $B(v, P_w) = \{v' \in F \cup \{w\} \mid v' R_w v\}$. Let $A(P_w)$ denote the set of firms which are *acceptable* to w under P_w; i.e., $A(P_w) = \{f \in F \mid f P_w w\}$. Given P_w and a subset of firms $S \subseteq F$, let $P_w|S$ denote the strict ordering on S consistent with P_w. Similarly, given $P_f \in \mathcal{P}_f$, $v \in W \cup \{f\}$ and $S \subseteq W$, we define $B(v, P_f)$, $A(P_f)$ and $P_f|S$. Let $\mathcal{P} \equiv \times_{v \in V} \mathcal{P}_v$. Elements of $\mathcal{P}$ are called (preference) *profiles*. To emphasize the role of agent v's preference in the profile $P \in \mathcal{P}$ we will write it as (P_v, P_{-v}).

A *matching market* is a triple (F, W, P), where F is a set of firms, W is a set of workers, and P is a preference profile. Because F and W will remain fixed, a matching market is simply a profile $P \in \mathcal{P}$. The assignment problem consists of matching workers with firms, keeping the bilateral nature of their relationship and allowing for the possibility that both, firms and workers, may remain unmatched. Namely, a *matching* is a function $\mu : V \to V$ satisfying the following properties: (m1) for all $w \in W$, $\mu(w) \in F \cup \{w\}$; (m2) for all $f \in F$, $\mu(f) \in W \cup \{f\}$; and (m3) for all $v \in V$, $\mu(\mu(v)) = v$. We say that agent v is *unmatched under* μ if $\mu(v) = v$. Let

$\mathcal{M}$ denote the set of all matchings.[6]

A matching is stable if no worker or firm prefers to be unmatched (*individual rationality*) and no unmatched pair mutually prefer each other to their assigned partners (*pair-wise stability*). Namely, given a profile $P \in \mathcal{P}$ a matching $\mu \in \mathcal{M}$ is *stable under* P if (s1) for all $v \in V$, $\mu(v) R_v v$; and (s2) there exists no pair $(w, f) \in W \times F$ such that $f P_w \mu(w)$ and $w P_f \mu(f)$. Gale and Shapley (1962) show that the set of stable matchings under P is non-empty and coincides with the *core* of the matching market P; that is, there is no loss of generality if we assume that all blocking power is carried out only by individual agents and by worker-firm pairs. We denote by $C(P)$ the set of stable matchings under P (or the core induced by P). Given a profile $P \in \mathcal{P}$ a matching $\mu \in \mathcal{M}$ is *Pareto-optimal* if there exists no matching $\mu' \in \mathcal{M}$ such that $\mu'(v) R_v \mu(v)$ for all $v \in V$ with strict preference holding for at least one agent.

The core of a matching market has a lattice structure (Knuth (1976) attributes this result to John Conway; see Roth and Sotomayor (1990) for the formal statement and proof of this result). Therefore, the core of a matching market contains two stable matchings, μ_F and μ_W, the two extremes of the complete lattice (called the firms-optimal stable matching and the workers-optimal stable matching, respectively) which have the property that firms (workers) unanimously agree that μ_F (μ_W) is the best stable matching; moreover, the optimal stable matching for one side of the market is the worst stable matching for the other side.

Whether or not a matching is stable depends on the preferences of agents, and since they constitute private information, agents have to be asked about them. A mechanism requires each agent v to report some preference relation $P_v \in \mathcal{P}_v$ and associates a matching with the reported profile. Formally, a *mechanism* is a function $\varphi : \mathcal{P} \to \mathcal{M}$ mapping each preference profile $P \in \mathcal{P}$ to a matching $\varphi[P] \in \mathcal{M}$. Therefore, $\varphi[P](v)$ is the agent matched to v at preference profile P by mechanism

[6]We are following the convention of extending the preference relation P_v from the original set of potential partners to the set of all matchings $\mathcal{M}$ by identifying a matching μ with $\mu(v)$. For instance, to say that firm f prefers μ' to μ means that either $\mu'(f) = \mu(f)$ or $\mu'(f) P_f \mu(f)$.

φ. A mechanism φ is stable if for all $P \in \mathcal{P}$, $\varphi[P] \in C(P)$.

The *deferred-acceptance algorithm* defined by Gale and Shapley (1962) is a stable mechanism that produces either μ_F or μ_W depending on the side of the market that makes the offers. At any step of the algorithm in which firms make offers (denoted by $DA_F : \mathcal{P} \to \mathcal{M}$), each firm f proposes to the most-preferred worker among the set of workers that have not already rejected f during previous steps, while a worker w accepts the most-preferred firm among the set of current offers plus the firm provisionally matched to w in the previous step (if any). The algorithm stops at the step when either all offers are accepted or firms have no more acceptable workers to whom they want to make an offer; the provisional matching becomes then definite and is the stable matching μ_F; i.e., $DA_F[P] = \mu_F$ for all $P \in \mathcal{P}$. Symmetrically if workers make offers, and the outcome of the algorithm (denoted by $DA_W : \mathcal{P} \to \mathcal{M}$) is the stable matching μ_W; i.e., $DA_W[P] = \mu_W$ for all $P \in \mathcal{P}$.[7]

When each agent has complete information about the preference relations of all other agents then: (1) No stable mechanism exists for which stating the true preferences is a dominant strategy for every agent (Roth, 1982). (2) Truth-telling is a dominant strategy for one side of the market if the deferred-acceptance algorithm selects that side's optimal stable matching (Dubins and Freedman (1981) and Roth (1982)). Therefore, if the core is singleton for a matching market, then the deferred-acceptance algorithm chooses the same matching independently of the side which makes offers. However, in general this fact does not allow us to conclude that for *any* stable mechanism truth-telling is a Nash equilibrium whenever the true profile has a singleton core.

[7]Strictly speaking, the DA-algorithm is an algorithm that finds the matching chosen by the "DA-mechanism". However, most of the matching literature uses the term DA-algorithm when referring to both the algorithm and the mechanism. We follow this convention.

3 Incomplete Information

We give up the usual assumption that the submitted lists are common knowledge and consider Bayesian preference revelation games induced by a stable mechanism and a common belief which is shared among all agents. A *common belief over* $\mathcal{P}$ is a probability distribution $\tilde{P}$ over $\mathcal{P}$. Given $v \in V$, let $\tilde{P}_v$ denote the marginal distribution of $\tilde{P}$ over $\mathcal{P}_v$. Given a common belief $\tilde{P}$ and a preference relation P_v, let $\tilde{P}_{-v}|_{P_v}$ denote the probability distribution over $\mathcal{P}_{-v}$ conditional on P_v.[8] Given a common belief $\tilde{P}$, (i) information is complete if $\tilde{P}$ puts probability one on a unique profile $P \in \mathcal{P}$ and (ii) information is incomplete if information is not complete.

A random matching $\tilde{\mu}$ is a probability distribution over the set of matchings $\mathcal{M}$. Let $\tilde{\mu}(v)$ denote the probability distribution which $\tilde{\mu}$ induces over v's set of potential partners ($F \cup \{w\}$ if $v = w$ and $W \cup \{f\}$ if $v = f$).[9]

A mechanism φ and a common belief $\tilde{P}$ define an (ordinal) game of incomplete information as follows. A strategy of v is a function $s_v : \mathcal{P}_v \to \mathcal{P}_v$ specifying for each type of v a list that v submits to the mechanism. A *strategy profile* is a list $s = (s_v)_{v \in V}$ associating with each agent a strategy. Given a mechanism $\varphi : \mathcal{P} \to \mathcal{M}$ and a common belief $\tilde{P}$ over $\mathcal{P}$, a strategy profile $s : \mathcal{P} \to \mathcal{P}$ induces a random matching in the following way: for all $\mu \in \mathcal{M}$, $\Pr\{\tilde{P} = P \mid \varphi[s(P)] = \mu\}$ is the probability of matching μ. However, the relevant random matching for agent v, given his type P_v and a strategy profile s, is $\varphi[s_v(P_v), s_{-v}(\tilde{P}_{-v}|_{P_v})]$ (where $s_{-v}(\tilde{P}_{-v}|_{P_v})$ is the probability distribution over $\mathcal{P}_{-v}$ which s_{-v} induces conditional on P_v). Note that $\varphi[s_v(P_v), s_{-v}(\tilde{P}_{-v}|_{P_v})](v)$ is the distribution which the random matching $\varphi[s_v(P_v), s_{-v}(\tilde{P}_{-v}|_{P_v})]$ induces over v's set of potential partners.

All mechanisms used in centralized matching markets are ordinal. In other words the only relevant information for a mechanism are the agents' rankings over their sets

[8] Note that we do not impose any condition on the common belief such as symmetry of agents' beliefs or independence.

[9] Formally, if $v = w$, then $\Pr\{\tilde{\mu}(v) = f\} = \sum_{\mu \in \mathcal{M}: \mu(v) = f} \Pr\{\tilde{\mu} = \mu\}$ for all $f \in F$ and $\Pr\{\tilde{\mu}(v) = v\} = \sum_{\mu \in \mathcal{M}: \mu(v) = v} \Pr\{\tilde{\mu} = \mu\}$.

of potential partners. In this environment truth-telling is a Bayesian Nash equilibrium whenever for every von Neumann-Morgenstern (vNM)-utility submitting the induced ordinal ranking maximizes an agent's expected utility in the Bayesian preference revelation game induced by the common belief and the mechanism. Equivalently, truth-telling is an ordinal Bayesian Nash equilibrium (OBNE) if the distribution over his partners when reporting the true ranking first-order stochastically dominates any distribution over his partners when submitting another ranking (given the others' strategies and the common belief).

A random matching $\tilde{\mu}$ *first-order stochastically* P_f-*dominates* a random matching $\tilde{\mu}'$, denoted by $\tilde{\mu}(f) \succcurlyeq_{P_f} \tilde{\mu}'(f)$, if for all $v \in W \cup \{f\}$, $\Pr\{\tilde{\mu}(f) \in B(v, P_f)\} \geq \Pr\{\tilde{\mu}'(f) \in B(v, P_f)\}$. Similarly, $\tilde{\mu}(w) \succcurlyeq_{P_w} \tilde{\mu}'(w)$ means that random matching $\tilde{\mu}$ *first-order stochastically* P_w-*dominates* random matching $\tilde{\mu}'$.

Definition 1 *Let* $\tilde{P}$ *be a common belief over* $\mathcal{P}$. *Then* truth-telling *is an* Ordinal Bayesian Nash Equilibrium (OBNE) *in the mechanism* φ *iff for all* $v \in V$ *and all* $P_v \in \mathcal{P}_v$ *such that* $\Pr\{\tilde{P}_v = P_v\} > 0$ *we have*

$$\varphi[P_v, \tilde{P}_{-v}|_{P_v}](v) \succcurlyeq_{P_v} \varphi[P'_v, \tilde{P}_{-v}|_{P_v}](v)$$

for all $P'_v \in \mathcal{P}_v$.

More generally, a strategy profile is an ordinal Bayesian Nash equilibrium whenever for any agent's true ordinal preference submitting the ranking specified by his strategy maximizes his expected utility for every vNM-utility representation of his true preference. This requires that an agent's strategy only depends on the ordinal ranking induced by his vNM-utility function. Of course, this is true for truth-telling. Furthermore, ordinal strategies are meaningful if an agent only observes his ordinal ranking and may have (still) little information about his utilities of his potential partners.

Definition 2 *Let* $\tilde{P}$ *be a common belief over* $\mathcal{P}$. *Then a strategy profile* s *is an* Ordinal Bayesian Nash Equilibrium (OBNE) *in the mechanism* φ *iff for all* $v \in V$

and all $P_v \in \mathcal{P}_v$ *such that* $\Pr\{\tilde{P}_v = P_v\} > 0$ *we have*

$$\varphi[s_v(P_v), s_{-v}(\tilde{P}_{-v}|_{P_v})](v) \succcurlyeq_{P_v} \varphi[P'_v, s_{-v}(\tilde{P}_{-v}|_{P_v})](v)$$

for all $P'_v \in \mathcal{P}_v$.[10]

Observe that for any common belief the set of OBNE in a mechanism φ is non-empty. For instance, the constant strategy in which all agents declare that no agent in the other side of the market is acceptable is an OBNE in φ since the mechanism selects, at all profiles P and (P_{-v}, P'_v), the empty matching. Furthermore, for any stable mechanism φ, any matching μ can be connected to an OBNE s^μ in φ in the following way: for any $v \in V$ and any $P_v \in \mathcal{P}_v$, let $A(s^\mu_v(P_v)) = \{\mu(v)\}$ if $\mu(v) \in A(P_v)$ and let $A(s^\mu_v(P_v)) = \emptyset$ otherwise. Then s^μ is an OBNE in φ under any common belief because for any preference relation P_v, agent v ranks either as the unique acceptable match the agent specified by μ (if this agent is acceptable under P_v) or no partner acceptable. If information is complete, then any s^μ is a Nash equilibrium in φ and the outcomes of the strategies s^μ is the set of all individually rational and Pareto-optimal matchings. Both under complete and incomplete information there is a multiplicity of OBNE. Remark 2 in Section 4 further illustrates that the multiplicity of equilibria is a robust property of the direct revelation game (under incomplete information) induced by a stable mechanism even under very strong properties of the common belief, included those that "transform" the game into a game of complete information.

4 Truth-Telling and Singleton Cores

We will show that the observation that the cores of the submitted lists are small (Roth and Peranson, 1999) and that participants reveal their true preferences are

[10] In the definition of an OBNE optimal behavior of agent v is only required for the preferences of v which arise with positive probability under $\tilde{P}$. If $P_v \in \mathcal{P}_v$ is such that $\Pr\{\tilde{P}_v = P_v\} = 0$, then the conditional belief $\tilde{P}_{-v}|_{P_v}$ cannot be derived from $\tilde{P}$. However, we could complete the belief of v in the following way: let $\tilde{P}_{-v}|_{P_v}$ put probability one on a profile where all other agents submit lists which do not contain v.

intimately related in our simplified matching market since both have a simple and simultaneous equilibrium explanation.

We will be interested in the profiles with a singleton core. The *support of* $\tilde{P}$ is the set of profiles on which $\tilde{P}$ puts a positive weight. Formally, for all $P \in \mathcal{P}$, P belongs to the support of $\tilde{P}$ if and only if $\Pr\{\tilde{P} = P\} > 0$.

Theorem 1 *Let $\tilde{P}$ be a common belief. Then truth-telling is an OBNE in a stable mechanism if and only if the support of $\tilde{P}$ is contained in the set of all profiles with a singleton core.*

By Theorem 1, participants truthfully reveal their true preference because the submitted lists have a singleton core. Profiles with a singleton core can arise very easily. For instance, let each hospital offer a position for a certain medical speciality and suppose that each hospital ranks as acceptable only the students who studied its position specific speciality. Furthermore, suppose that all hospitals who have a position for specialty A rank the students who studied speciality A in the same way, say according to some objective criterion like their grades. Then, independently of the students' preferences, the core is always a singleton. Now if the common belief is such that any profile in its support has the properties as described above, then Theorem 1 applies and each participant cannot do better than truthfully reveal his preferences.

Remark 1 Theorem 1 can be read as truth-telling is an OBNE if and only if the support of $\tilde{P}$ is contained in the profiles for which under complete information truth-telling is a best response to the other's strategies. Obviously Theorem 1 is not necessarily true in general Bayesian games. For instance, consider the game of matching pennies. Interpret each of the two player's strategies (heads or tails) as his possible types. If each player's type arises with the same probability, then truth-telling is an OBNE but under complete information there is no Nash Equilibrium in pure strategies.[11]

[11] We conjecture that the existence of OBNE hinges crucially on the "strong indivisibilities" prop-

Remark 2 Of course, truth-telling is not the unique OBNE in a stable mechanism even when the support of $\tilde{P}$ is contained in the set of all profiles with a singleton core. To see this, let $\{P^1, ..., P^K\}$ be an arbitrary set of profiles with the property that for all $1 \leq k \leq K$ and all $v \in V$, $\left|C\left(P^k\right)\right| = 1$ and

$$P_v^{k'} \neq P_v^k \quad \text{for all } k' \neq k. \tag{1}$$

For each k, let μ^k be an individually rational matching relative to the profile P^k and let φ be a stable mechanism. We know, by Roth and Sotomayor (1990), that there exists $\bar{P}^k \in \mathcal{P}$ such that $\varphi[\bar{P}^k] = \mu^k$ and $\bar{P}^k$ is a NE of the direct preference revelation game induced by φ. Observe that, in general, $\bar{P}^k$ is not equal to P^k. Let $\tilde{P}$ be a common belief over $\mathcal{P}$ with support on $\{P^1, ..., P^K\}$. Consider any strategy profile $s = (s_v)_{v \in V}$, where $s_v : \mathcal{P}_v \to \mathcal{P}_v$ has the property that $s_v\left(P_v^k\right) = \bar{P}_v^k$ for all k and all $v \in V$. It is immediate to see that, since condition (1) holds and $\bar{P}^k$ is a NE of the complete information game induced by the mechanism φ (with preferences P^k), s is an OBNE in the stable mechanism φ. However, this equilibrium is arbitrary and without much predictive power since it requires extremely large amounts of coordination among all agents. In contrast, truth-telling arises as a natural and simple behavior in large markets where this coordination is literally unfeasible.

Remark 3 Much attention has been paid to the incentives that members of a couple who want to live together face when looking coordinately for two jobs in entry-level professional markets (see Roth (1984a), Roth and Sotomayor (1990), Dutta and Massó (1997), Roth and Peranson (1999), Cantala (2002), Roth (2002), Klaus and Klijn (2005), and Klaus, Klijn and Massó (2006)). A straightforward extension of the proof of Theorem 1 shows that, under its assumptions, no couple can misrepresent coordinately their preferences in a stable mechanism φ such that both members of the couple strictly benefit. To see this, let $\tilde{P}$ be a common belief with support contained

erty (which induces a "natural ordinality") in a matching market and that this does not necessarily remain true for general NTU or TU games (where no natural ordinality is induced; these games are cardinal).

in the set of all profiles with a singleton core. Let w and w' be a couple and assume that all remaining agents are truth-telling. Because in the stable mechanism DA_W no subset of workers can gain by jointly misrepresenting their preferences we have that, similarly as in the proof of Theorem 1, for all P such that $\Pr\{\tilde{P} = P\} > 0$,

$$\varphi\left[P\right](v)\, R_v \varphi[P'_w, P'_{w'}, P_{-\{w,w'\}}]\,(v) \quad \text{for } v = w \text{ or } v = w'.$$

Therefore, truth-telling is a joint best response for the couple w and w'. Note that the same is true for any set of firms, i.e. truth-telling is a best response for any set of firms. Thus, if each firm represents the department of a hospital and each department has at most one open position (in its medical speciality), then hospitals cannot misrepresent the preferences of their departments such that all departments strictly gain.

Remark 4 In matching markets Sönmez's (1999) result says the following:

Let $\varphi : \mathcal{P} \to \mathcal{M}$ be a mechanism choosing for each profile $P \in \mathcal{P}$ an individually rational and Pareto-optimal matching. If for all $v \in V$ and all $P_v \in \mathcal{P}_v$, P_v is a (weakly) dominant strategy in the game induced by φ at P_v[12]*, then $|C(P)| = 1$ for all $P \in \mathcal{P}$.*

The following are the important differences between Sönmez's result and Theorem 1. First, Sönmez (1999) requires full incentive-compatibility (for all $v \in V$ and all $P_v \in \mathcal{P}_v$) whereas we only require Bayesian incentive-compatibility (for all $v \in V$ and all $P_v \in \mathcal{P}_v$ such that $\Pr\{\tilde{P}_v = P_v\} > 0$. Under complete information (say under P) Bayesian incentive-compatibility reduces to the requirement that P is one of the Nash equlibria of the preference revelation game induced by the mechanism φ and P. Second, since in matching markets stability implies Pareto-optimality and individual rationality, our requirement on the mechanism is stronger than in Sönmez (1999). Third, in Theorem 1 "singleton cores" is both a necessary and sufficient condition.[13]

[12]For all $P'_v \in \mathcal{P}_v$ and all $P_{-v} \in \mathcal{P}_{-v}$, $\varphi[P_v, P_{-v}](v) R_v \varphi[P'_v, P_{-v}](v)$

[13]Takamiya (2003) showed that the converse of Sönmez's general result is not necessarily true:

5 Ex-post Unanimity and Small Cores

Theorem 1 characterized the common beliefs for which a specific strategy profile is an OBNE. In this result the singleton core condition on the common belief was independent of which stable mechanism is used. The key feature of the mechanism was its stability and not whether workers or firms make proposals like in the DA-algorithm.

Generally, however, whether a strategy profile is an OBNE may depend on the stable mechanism. We will generalize the necessary condition of Theorem 1. We will show that a necessary condition for a strategy profile to be an OBNE is that for any profile belonging to the support of the common belief, the stable mechanism chooses the matching which is unanimously most preferred in the core of the submitted profile. This is more likely when the core of the submitted profile is "small" in terms of the true profile. If the submitted profile is one with singleton core (like in Theorem 1), then this condition is trivially satisfied.

Theorem 2 *Let $\tilde{P}$ be a common belief, s be a strategy profile, and φ be a stable mechanism. If s is an OBNE in the stable mechanism φ, then any profile belonging to the support of $\tilde{P}$ has the following property: all agents unanimously agree that the matching chosen by φ for the submitted profile is most preferred among all matchings in the core of the submitted profile. Formally, for all $P \in \mathcal{P}$ such that $\Pr\{\tilde{P} = P\} > 0$, we have $\varphi[s(P)](v) R_v \mu(v)$ for all $v \in V$ and all $\mu \in C(s(P))$.*

If a common belief, a strategy profile and a mechanism satisfy the ex-post unanimity condition of Theorem 2, then by strictness of preferences, for any (true) profile P belonging to the support there is a unique matching μ in the core of the submitted profile which is most preferred under the true profile and which has to be chosen by the mechanism, i.e., $\mu = \varphi[s(P)]$. This implies that the belief cannot attribute positive probability to a profile where some agents' preferences are opposed for any two

there are general allocation problems with indivisibilities where the core is a singleton for each profile but the mechanism choosing the unique core allocation for each profile is not incentive-compatible.

matchings belonging to the core of the submitted profile. However, ex-post unanimity does not require that the core of the submitted profile is a singleton.

Generally, the condition in Theorem 2 is not sufficient for a profile of strategies to be an OBNE.[14] Whether or not it is satisfied depends on both the stable mechanism and the agents' strategies. Furthermore, this condition is not sufficient for the core of the submitted profile to be a singleton since ex-post unanimity is in terms of the true profile.

6 Conclusion

Our analysis of ordinal Bayesian Nash equilibria of stable mechanisms under incomplete information confirms some already known reasons of why stable mechanisms arose and lasted for many years in centralized two-sided matching markets, and suggests some additional ones. First, under incomplete information, truth-telling remains a plausible behavior if and only if the cores of the support of the common belief are singleton; hence, the stability of the realized matching is guaranteed. This is an important property and becomes critical if the market has to be redesigned. Second, this feature of equilibrium behavior is independent of the chosen stable mechanism. This is significant since the two sides of the matching market have opposite interests on the set of stable matchings (and thus, on possible alternative stable mechanisms). Third, equilibrium is reinforced because each participant is matched to the best possible partner, that is, the partner most preferred among those he is matched to by any stable matching relative to the declared profile.

At a more conceptual level, one may ask why a centralized market mechanism

[14]For instance, let $\tilde{P}$ be a belief putting probability one on a profile P under which all agents rank acceptable all potential partners. Further let $s(P)$ be such that each worker truthfully reveals her preference and each firm submits an empty list (ranking all workers unacceptable). Then the condition in Theorem 2 is satisfied but $s(P)$ is obviously not an OBNE. Any firm gains by revealing its true preference.

is needed when truth-telling is an equilibrium. The problem is that in decentralized markets there are frictions because it is difficult for agents to communicate with all possible partners to find out their preferences. Furthermore, during this search agents are unlikely to reveal their complete preferences and unravelling may occur.

Overall, we (unexpectedly) found that the more realistic and potentially richer strategic setting of incomplete information reinforces some of the reasons already given to explain why many of the entry-level professional labor markets have been operating in a relatively smooth way for so many years.

All our results apply to a simplified one-to-one version of matching markets. Although one-to-one matching markets are a reasonable approximation of many-to-one matching markets, the following example, based on Sönmez (1997), shows that Theorem 1 does not generalize from one-to-one matching markets to many-to-one matching markets.

Example 1 Consider a matching market with two firms $F = \{f_1, f_2\}$ and two workers $W = \{w_1, w_2\}$. Firm f_1 has one position but firm f_2 has two positions. Consider any common belief with support $\{P, \bar{P}\}$, where P and $\bar{P}$ differ in firm f_2's preference and are the following. The profile P is

$$\begin{array}{cccc} \underline{P_{f_1}} & \underline{P_{f_2}} & \underline{P_{w_1}} & \underline{P_{w_2}} \\ w_1 & \{w_1, w_2\} & f_2 & f_1 \\ w_2 & w_2 & f_1 & f_2 \\ & w_1 & & \end{array},$$

and the profile $\bar{P}$ is

$$\begin{array}{cccc} \underline{\bar{P}_{f_1}} & \underline{\bar{P}_{f_2}} & \underline{\bar{P}_{w_1}} & \underline{\bar{P}_{w_2}} \\ w_1 & w_2 & f_2 & f_1 \\ w_2 & & f_1 & f_2 \end{array}.$$

Both profiles have singleton cores: $C(P) = \{\mu\}$ where $\mu(w_1) = f_2$ and $\mu(w_2) = f_1$; and $C(\bar{P}) = \{\bar{\mu}\}$ where $\bar{\mu}(w_1) = f_1$ and $\bar{\mu}(w_2) = f_2$. Nevertheless it is in firm f_2's best interest to report preference $\bar{P}_{f_2}$, even when its true preference is P_{f_2}. Thus,

truth-telling is not an ordinal Bayesian Nash equilibrium in this example although each profile in the support of the common belief has singleton core.

References

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

[2] D. Cantala. "Matching Markets: The Particular Case of Couples". Mimeo, Universidad de Guanajuato (2002).

[3] S. Clark. "Uniqueness of Equilibrium in Two-Sided Matching". Mimeo, University of Edinburgh (2003).

[4] L. Dubins and D. Freedman. "Machiavelli and the Gale-Shapley Algorithm", *American Mathematical Monthly* **88**, 485-494 (1981).

[5] B. Dutta and J. Massó. "Stability of Matchings When Individuals Have Preferences over Colleagues", *Journal of Economic Theory* **75**, 464-475 (1997).

[6] J. Eeckhout. "On the Uniqueness of Stable Marriage Matchings", *Economics Letters* **69**, 1-8 (2000).

[7] L. Ehlers. "In Search of Advice for Physicians in Entry-Level Medical Markets". Mimeo, Université de Montréal (2002).

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

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

[10] D. Gale and M. Sotomayor. "Ms. Machiavelli and the Stable Matching Problem", *American Mathematical Monthly* **92**, 261-268 (1985).

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

[12] B. Klaus, F. Klijn and J. Massó. "Some Things Couples Always Wanted to Know about Stable Matchings (But Were Afraid to Ask)", *Review of Economic Design*, forthcoming (2006).

[13] D. Knuth. *Marriages Stables*. Les Presses de l'Université de Montréal, Montréal (1976).

[14] D. Majumdar and A. Sen. "Ordinally Bayesian Incentive-Compatible Voting Rules", *Econometrica* **72**, 523-540 (2004).

[15] M. Niederle and A.E. Roth. "Unraveling Reduces Mobility in a Labor Market: Gastroenterology with and without a Centralized Match", *Journal of Political Economy* **111**, 1342-1352 (2003).

[16] K. Takamiya. "On Strategy-Proofness and Essentially Single-Valued Cores: A Converse Result," *Social Choice and Welfare* **20**, 77-83 (2003).

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

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

[19] A.E. Roth. "Misrepresentation and Stability in the Marriage Problem", *Journal of Economic Theory* **34**, 383-387 (1984b).

[20] A.E. Roth. "Two-Sided Matching with Incomplete Information about Others' Preferences", *Games and Economic Behavior* **1**, 191-209 (1989).

[21] A.E. Roth. "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics", *Econometrica* **70**, 1341-1378 (2002).

[22] A.E. Roth and E. Peranson. "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design", *American Economic Review* **89**, 748-780 (1999).

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

[24] A.E. Roth and M. Sotomayor. *Two-sided Matching: A Study in Game-Theoretic Modelling and Analysis*. Cambridge University Press, Cambridge, England. [Econometric Society Monograph] (1990).

[25] A.E. Roth and X. Xing. "Jumping the Gun: Imperfections and Institutions related to the Timing of Market Transactions", *American Economic Review* **84**, 992-1044 (1994).

[26] T. Sönmez. "Manipulation via Capacities in Two-sided Matching Markets", *Journal of Economic Theory* **77**, 197-204 (1997).

[27] T. Sönmez. "Strategy-Proofness and Essentially Single-Valued Cores", *Econometrica* **67**, 677-689 (1999).

[28] K. Takamiya. "On Strategy-proofness and Essentially Single-valued Cores: A Converse Result", *Social Choice and Welfare* **20**, 77-83 (2003).

APPENDIX

In the Appendix we prove Theorems 1 and 2.

A Truth-Telling

Theorem 1 *Let $\tilde{P}$ be a common belief. Then truth-telling is an OBNE in a stable mechanism if and only if the support of $\tilde{P}$ is contained in the set of all profiles with a singleton core.*

Proof. Let φ be a stable mechanism.
($\Leftarrow$) Let $\tilde{P}$ be such that for all $P \in \mathcal{P}$, $\Pr\{\tilde{P} = P\} > 0$ implies $|C(P)| = 1$. Let $P \in \mathcal{P}$ be such that $|C(P)| = 1$. We show that under complete information P is a Nash Equilibrium in the direct preference revelation game induced by φ. We show that P_v is a best response to P_{-v} for all $v \in V$.

Let $v \in W$. By $|C(P)| = 1$,

$$DA_W[P] = \varphi[P]. \tag{2}$$

Truth-telling is dominant strategy for v under DA_W (Dubins and Freedman, 1981; Roth, 1982). However, in general this fact does not allow us to conclude that P_v is a best response to P_{-v} for v under the stable mechanism φ. We will show that for any possible deviation of v at φ, there exists a deviation of v at DA_W such that v is matched to the same partner as under v's deviation at φ.

Let $P'_v \in \mathcal{P}_v$. Let $P''_v \in \mathcal{P}_v$ be such that $A(P''_v) = \{\varphi[P'_v, P_{-v}](v)\}$ if $\varphi[P'_v, P_{-v}](v) \in F$ and $A(P''_v) = \emptyset$ if $\varphi[P'_v, P_{-v}](v) = v$. By construction of P''_v and stability of φ, $\varphi[P'_v, P_{-v}] \in C(P'_v, P_{-v})$ implies $\varphi[P'_v, P_{-v}] \in C(P''_v, P_{-v})$. Since the set of unmatched agents is identical under any two stable matchings, the stability of DA_W and the construction of P''_v yield

$$DA_W[P''_v, P_{-v}](v) = \varphi[P'_v, P_{-v}](v). \tag{3}$$

Because for DA_W a worker cannot gain by misrepresentation we have

$$DA_W[P](v)R_vDA_W[P''_v, P_{-v}](v). \tag{4}$$

Hence, by (2), (3), and (4), $\varphi[P](v)R_v\varphi[P'_v, P_{-v}](v)$, the desired conclusion.

Using $|C(P)| = 1$ and $DA_F[P] = \varphi[P]$, similarly as above it follows that for all $v \in F$ and all $P'_v \in \mathcal{P}_v$, $\varphi[P](v)R_v\varphi[P'_v, P_{-v}](v)$.

Let $v \in V$ and $P_v \in \mathcal{P}_v$ be such that $\Pr\{\tilde{P}_v = P_v\} > 0$. Because for all $P_{-v} \in \mathcal{P}_{-v}$ such that $\Pr\{\tilde{P}_{-v}|_{P_v} = P_{-v}\} > 0$ we have $|C(P_v, P_{-v})| = 1$ and under complete information P_v is a best response to P_{-v} in the direct preference revelation game, it follows that submitting P_v is a best response for agent v. Hence, truth-telling is an OBNE in the stable mechanism φ.

($\Rightarrow$) Suppose that there exists $P \in \mathcal{P}$ such that $\Pr\{\tilde{P} = P\} > 0$ and $|C(P)| \geq 2$. Then (i) there exists $w \in W$ such that $DA_W[P](w)P_w\varphi[P](w)$ or (ii) there exists $f \in F$ such that $DA_F[P](f)P_f\varphi[P](f)$. Without loss of generality, suppose that (i) holds. Let $DA_W[P](w) = f'$. Let $P'_w \in \mathcal{P}_w$ be such that $P'_w|F = P_w|F$ and $A(P'_w) = B(f', P_w)$.

Let $P'_{-w} \in \mathcal{P}_{-w}$ be such that $\Pr\{\tilde{P} = (P_w, P'_{-w})\} > 0$. Since we will show that truth-telling is not an OBNE in the stable mechanism φ by looking at the probability $\Pr\{\varphi[P_w, \tilde{P}_{-w}|_{P_w}](w) \in B(f', P_w)\}$, assume P'_{-w} is such that $\varphi[P_w, P'_{-w}](w)R_wf'$. Then $\varphi[P_w, P'_{-w}] \in C\left(P_w, P'_{-w}\right)$ implies $\varphi[P_w, P'_{-w}] \in C(P'_w, P'_{-w})$ since individual rationality of $\varphi[P_w, P'_{-w}]$ at profile $\left(P_w, P'_{-w}\right)$ implies individual rationality of $\varphi[P_w, P'_{-w}]$ at profile $\left(P'_w, P'_{-w}\right)$ and $(\hat{w}, \hat{f})$ blocks $\varphi[P_w, P'_{-w}]$ at profile $\left(P'_w, P'_{-w}\right)$ implies $(\hat{w}, \hat{f})$ blocks $\varphi[P_w, P'_{-w}]$ at profile $\left(P_w, P'_{-w}\right)$ as well. Thus, by $A(P'_w) = B(f', P_w)$ and the fact that under any two stable matchings the set of unmatched agents is identical, $\varphi[P'_w, P'_{-w}](w)R_wf'$. We next show that $\varphi[P'_w, P_{-w}](w) = f'$. Suppose $\varphi[P'_w, P_{-w}](w) = w$. Then $DA_W[P'_w, P_{-w}](w) = w$. Therefore

$$DA_W[P_w, P_{-w}](w) = f'P'_ww = DA_W[P'_w, P_{-w}](w),$$

which contradicts the fact that for w truth-telling is a dominant strategy in the direct preference revelation mechanism induced by DA_W under complete information.

A similar argument shows that $f' R'_w \varphi[P'_w, P_{-w}](w)$. Thus $\varphi[P'_w, P_{-w}](w) = f'$. Furthermore, $\Pr\{\tilde{P}_{-w}|_{P_w} = P_{-w}\} > 0$, $f' P_w \varphi[P](w)$, and $\varphi[P'_w, P_{-w}](w) = f'$. Hence,

$$\Pr\{\varphi[P_w, \tilde{P}_{-w}|_{P_w}](w) \in B(f', P_w)\} < \Pr\{\varphi[P'_w, \tilde{P}_{-w}|_{P_w}](w) \in B(f', P_w)\},$$

which means that truth-telling is not an OBNE in the stable mechanism . ■

B Small Cores

Theorem 2 *Let $\tilde{P}$ be a common belief, s be a strategy profile, and φ be a stable mechanism. If s is an OBNE in the stable mechanism φ, then any profile belonging to the support of $\tilde{P}$ has the following property: for all $P \in \mathcal{P}$ such that $\Pr\{\tilde{P} = P\} > 0$, we have $\varphi[s(P)](v) R_v \mu(v)$ for all $v \in V$ and all $\mu \in C(s(P))$.*

Proof. Suppose not. Then there exist $P \in \mathcal{P}$ such that $\Pr\{\tilde{P} = P\} > 0$ and $\mu(v) P_v \varphi[s(P)](v)$ for some $v \in V$ and $\mu \in C(s(P))$. Because φ is stable and the set of unmatched agents is identical under any two stable matchings, we have $\mu(v) \neq v$ and $\varphi[s(P)](v) \neq v$. Without loss of generality, suppose that $v \in W$, $\mu(v) = f$, and f is P_v-most preferred in $C(s(P))$.

Let $s_v(P_v) = P'_v$. Let $P''_v \in \mathcal{P}_v$ be such that (i) $A(P''_v) = A(P'_v) \cap B(f, P_v)$ and (ii) $P''_v|A(P''_v) = P'_v|A(P''_v)$. We show that

$$\Pr\{\varphi[P''_v, s_{-v}(\tilde{P}_{-v}|_{P_v})](v) \in B(f, P_v)\} > \Pr\{\varphi[P'_v, s_{-v}(\tilde{P}_{-v}|_{P_v})](v) \in B(f, P_v)\}, \quad (5)$$

which contradicts the fact that s is an OBNE.

Let $P'_{-v} \in \mathcal{P}_{-v}$ be such that $\Pr\{\tilde{P}_{-v}|_{P_v} = P'_{-v}\} > 0$ and $\varphi[P'_v, s_{-v}(P'_{-v})](v) \in B(f, P_v)$. By stability of φ, $\varphi[P'_v, s_{-v}(P'_{-v})] \in C(P'_v, s_{-v}(P'_{-v}))$. By construction of P''_v, $\varphi[P'_v, s_{-v}(P'_{-v})] \in C(P''_v, s_{-v}(P'_{-v}))$. Since the set of unmatched agents is identical under any two stable matchings, $\varphi[P'_v, s_{-v}\left(P'_{-v}\right)](v) \in B(f, P_v)$ implies that $\varphi[P'_v, s_{-v}\left(P'_{-v}\right)](v) \neq v$, and hence $\varphi[P''_v, s_{-v}\left(P'_{-v}\right)](v) \neq v$. Thus, $\varphi[P''_v, s_{-v}(P'_{-v})](v) \in A(P''_v)$, and by $A(P''_v) \subseteq B(f, P_v)$, $\varphi[P''_v, s_{-v}(P'_{-v})](v) \in B(f, P_v)$.

By construction of P_v'' and since $\mu \in C(P_v', s_{-v}(P_{-v}))$, $\mu \in C(P_v'', s_{-v}(P_{-v}))$. Moreover, since $\mu(v) = f$ and the set of unmatched agents is identical under any two stable matchings, $\varphi[P_v'', s_{-v}(P_{-v})](v) \neq v$. By $A(P_v'') \subseteq B(f, P_v)$, $\varphi[P_v'', s_{-v}(P_{-v})](v) \in B(f, P_v)$. Furthermore, $\Pr\{\tilde{P}_{-v}|_{P_v} = P_{-v}\} > 0$ and $\varphi[P_v', s_{-v}(P_{-v})](v) \notin B(f, P_v)$. Hence, (5) is true. ■