

Incomplete Information and Small Cores in Matching Markets ^{*}

Lars Ehlers[†] Jordi Massó[‡]

September 2003

Abstract

The National Resident Matching Program (NRMP) is a centralized clearinghouse which annually matches approximately 20,000 graduating physicians to positions at hospitals in the United States. It has been observed that the cores of the profiles submitted to the clearinghouse are small *and* a significant number of physicians and hospitals truthfully reveal their preferences (Roth and Peranson, 1999). We study Bayesian Nash equilibria of stable mechanisms (such as the NRMP) in matching markets under incomplete information. We will explain the observed facts by showing that truth-telling is an equilibrium of the Bayesian 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

^{*}L. Ehlers acknowledges financial support from the SSHRC (Canada) and the FCAR (Québec). The work of J. Massó is partially supported by the Spanish Ministry of Science and Technology, through grant BEC2002-02130, and by the Generalitat de Catalunya, through grant 2001SGR-00162 and the Barcelona Economics Program (CREA). This research started when J. Massó was visiting the Université de Montréal; financial support from CIREQ is gratefully acknowledged.

[†]Département de Sciences Économiques and CIREQ, Université de Montréal, Montréal, Québec H3C 3J7, Canada; e-mail: lars.ehlers@umontreal.ca

[‡]Departament d'Economia i d'Història Econòmica and CODE, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès (Barcelona), Spain; e-mail: jordi.massó@uab.es

singleton cores. We also provide further reasons for the survival of stable mechanisms in entry-level professional labor markets.

JEL Classification: C78, D81, J44.

Keywords: Matching Market, Incomplete Information, Small 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 were reorganized by centralizing them from the 1950s. Each year a clearinghouse announces the open positions at each hospital and the finishing medical students which will be available. 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

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

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 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. On the other hand, it has been observed in centralized matching markets which use stable mechanisms that the opportunities for strategic manipulation are surprisingly small (Roth and Peranson, 1999). It seems that a significant number of participants truthfully reveal their preferences.

Another unexplained but observed fact is that the number of stable matchings for the submitted profile is usually small (Roth and Peranson, 1999). In the National Residents Matching Program (NRMP) the number of stable matchings has been 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). In addition, two recent papers have identified strong but meaningful sufficient conditions on the preference profiles under which the core is a singleton. Eeckhout (2000) proposes a condition, which is also necessary for markets with a

²Niederle 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.

³It 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.

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. Under the more realistic context of incomplete information, our paper provides an equilibrium explanation of why participants truthfully reveal their preferences *and* the cores of the submitted lists are small.

In centralized matching markets with a large number of participants the common knowledge assumption of the submitted lists is strong and we need to be able to relax it for real-life situations (for instance, the NRMP yearly fills approximately 20,000 jobs for new physicians). 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. 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 first 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

⁴This notion was introduced by d’Aspremont and Peleg (1988) who call it “ordinal Bayesian incentive-compatibility”.

if and only if the support of the common belief is contained in the set of profiles with singleton core. 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.

Second, we show in Theorem 2 that truncating the true preference lists (i.e., keeping the true rankings while possibly reducing the sets of acceptable partners) is an ordinal Bayesian Nash equilibrium in the Bayesian revelation game induced by a common belief and a stable mechanism if and only if at each preference profile in the support of the common belief the declared preference lists are a complete information Nash equilibrium of the direct preference revelation game induced by the stable mechanism. Observe that this condition is satisfied only if the support of the declared preference lists is contained in the set of profiles with singleton core. Since in the NRMP applicants can only list a position after being interviewed by the program offering it, Theorem 2 may explain the small core property of the reported profiles observed by Roth and Peranson (1999) during the last years in the NRMP.

Finally, in Theorem 3 we show 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 revelation games induced by stable mechanisms when participants have incomplete information about the ordinal preferences of all other agents. Roth and Rothblum (1999) and Ehlers (2002a,b) 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 hypotheses on the information structure of the beliefs held by the worker and for different mechanisms.

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-telling. Section 5 focuses on truncation strategies and Section 6 on general ordinal Bayesian Nash equilibria. Section 7 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 $v', \hat{v} \in V$, $v' R_v \hat{v}$ means that either $v' = \hat{v}$ or $v' \succ P_v \hat{v}$. 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., $A(P_w) = \{f \in F \mid f \succ P_w w\}$. 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 $\mu : V \rightarrow V$ 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.⁵

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).

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 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

⁵We 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 $\mu'(f) \succ \mu(f)$ or else $(f) P_f \mu(f)$.

associates a matching with the reported profile. Formally, a *mechanism* is a function $\varphi : \mathcal{P} \rightarrow \mathcal{M}$ 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 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 $DA_F : \mathcal{P} \rightarrow \mathcal{M}$), 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 at which all offers are accepted; 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 $DA_W : \mathcal{P} \rightarrow \mathcal{M}$) is the stable matching μ_W ; i.e., $DA_W[P] = \mu_W$ for all $P \in \mathcal{P}$.

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) No stable matching mechanism exists for which it is always an equilibrium for every agent to state his true preferences (Roth and Sotomayor, 1990). (3) In the deferred-acceptance algorithm in which firms make offers it is a dominant strategy for each firm to state its true preferences; similarly, for workers (Dubins and Freedman (1981) and Roth (1982)). (4) In the deferred-acceptance algorithm in which firms make offers and each firm chooses its dominant strategy and states its true preferences, any equilibrium outcome is a stable matching of the true preferences (Roth, 1984b). (5) Any individually rational matching is an equilibrium outcome of the deferred-acceptance algorithm (Roth and Sotomayor, 1990).

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 .

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 $v = 3D w$ and $W \cup \{f\}$ if $v = 3D f$).

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 $s_v : \mathcal{P}_v \rightarrow \mathcal{P}_v$ specifying for each type of v a list that v submits to the mechanism. A *strategy profile* is a list $s = 3D_v \{s_{\in V}$ associating with each agent a strategy. Given a mechanism $\varphi : \mathcal{P} \rightarrow \mathcal{M}$ and a common belief \tilde{P} over \mathcal{P} , a strategy profile $s : \mathcal{P} \rightarrow \mathcal{P}$ induces a random matching in the following way: for all $\mu \in \mathcal{M}$, $\Pr\{\tilde{P} = 3D P \mid \varphi[s(P)] = 3D \mu\}$ 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 $\varphi[s_v(P_v), s_{-v}(\tilde{P}_{-v}|_{P_v})]$. 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 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 stochastically dominates any distribution

over his partners when submitting another ranking (given the others' strategies and the common belief).

A random matching $\tilde{\mu}$ *stochastically P_f -dominates* a random matching $\tilde{\mu}'$, denoted by $\tilde{\mu}(f) \succ_{P_f} \tilde{\mu}'(f)$, if for all $v \in W \cup \{f\}$, $\Pr\{\tilde{\mu}(f) = 3D | w'R_f v\} \geq \Pr\{\tilde{\mu}'(f) = 3D | v'R_f v\}$. Similarly, $\tilde{\mu}(w) \succ_{P_w} \tilde{\mu}'(w)$ means that random matching $\tilde{\mu}$ *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 OBNE in the mechanism φ iff for all $v \in V$ and all $P_v \in \mathcal{P}_v$ such that $\Pr\{\tilde{P}_v = 3D_v\} > 0$ we have*

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

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 $\Pr\{\tilde{P}_v = 3D_v\} > 0$ we have*

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

for all $P'_v \in \mathcal{P}_v$.⁶

⁶In the definition of an OBNE optimal behavior of agent v is only required for the preferences

4 Truth-Telling and Singleton Cores

In the U.S. entry-level medical market it has been observed that the cores of the submitted lists are small and a significant number of participants reveal their true preferences (Roth and Peranson, 1999). We will show that these observations have a simple 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 $\Pr\{\tilde{P} = 3D P\} > 0$.

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 of v which arise with positive probability under \tilde{P} . If $P_v \in \mathcal{P}_v$ is such that $\Pr\{\tilde{P}_v = 3D P_v\} = 0$, then the conditional belief $\tilde{P}_{-v}|P_v$ cannot be derived from \tilde{P}_v . 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 .

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 there is no Nash Equilibrium in pure strategies.

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 $\{P^1, \dots, P^K\}$ be an arbitrary set of profiles with the property that for all $1 \leq k \leq K$ and all $v \in V$, $|C(P^k)| = 1$ and

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

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 $\{P^1, \dots, P^K\}$. Consider any strategy profile $s = (s_v)_{v \in V}$, where $s_v : \mathcal{P}_v \rightarrow \mathcal{P}_v$ has the property that $s_v(P_v^k) = \bar{P}_v^k$ 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 .

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 (2003), and Klaus, Klijn and Massó (2003)). 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

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 $\Pr\{\tilde{P} = P\} > 0$,

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

Therefore, truth-telling is a joint best response for the couple w and w' .

5 Truncation Strategies and Complete Information Equilibria

A number of recent papers (Roth and Rothblum (1999), Ehlers (2002a,b)) search for advice for workers in matching markets which use the DA-algorithm with firms proposing. They show that when a worker's uncertainty about the others' submitted lists is "symmetric", any non-truncation of the true preference is stochastically dominated by a truncation. A truncation keeps the true ranking of the firms while possibly reducing the acceptable set. Thus, given a common belief and a worker's preference, it is optimal for the worker to submit a truncation if her information derived from the common belief given her realization is symmetric in the sense of these papers. This advice is independent of the vNM-utility representation of the worker's true preference. Note that these papers determine the first order stochastically dominated strategies for a single worker in the DA-algorithm with firms proposing. They neither provide any common advice for the whole class of stable mechanisms nor determine equilibria.

Observe that submitting a truncation is risky whenever being unmatched is perceived as a very bad outcome. A participant who plays a truncation strategy may end up unmatched even though the participant and a (non-listed) acceptable partner may form a blocking pair. So, truncation strategies might produce stronger regrets than

just switching the ordering of two acceptable partners. However, it has been observed that many participants in matching markets play indeed truncation strategies (Roth and Peranson, 1999).⁷ Our next result identifies a necessary and sufficient condition for a profile of truncation strategies to be an OBNE in the preference revelation game induced by a stable mechanism. Of course, truth-telling is a profile of truncation strategies.

Formally, given $w \in W$, a *truncation* of a list P_w containing k acceptable firms is a list P'_w containing $k' \leq k$ acceptable firms such that the k' acceptable firms of P'_w are the first k' elements of P_w , in the same order. In the same way we define truncations for a firm. Let $v \in V$ and $s_v : \mathcal{P}_v \rightarrow \mathcal{P}_v$ be a strategy. Then we call s_v a *truncation strategy* if for all $P_v \in \mathcal{P}_v$, $s_v(P_v)$ is a truncation of P_v .

Theorem 2 *Let \tilde{P} be a common belief, s be a profile of truncation strategies and φ be a stable mechanism. Then s is an OBNE in the stable mechanism φ if and only if for any profile P in the support of \tilde{P} , $s(P)$ is a Nash equilibrium under complete information P in the direct preference revelation game induced by φ .*

Theorem 2 tells us that if the participants in a matching market play an OBNE in truncation strategies, then for any realization of the common belief they play a Nash equilibrium with common knowledge of the realized profile, and thus, the submitted lists. Thus, the participants in the U.S. entry-level medical market truncate their preferences in the “right” way. In the Appendix we also show that this condition is *independent* of the stable mechanism. For any true preference profile the set of complete information Nash equilibria in truncation strategies is identical across all stable mechanisms.

⁷Gale and Sotomayor (1985, p. 266) are the first who study truncations under complete information.

6 Ex-post Unanimity and Small Cores

Theorems 1 and 2 characterized the common beliefs for which a specific strategy profile is an OBNE. In both results the 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 conditions of Theorems 1 and 2. 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 Theorems 1 and 2), then this condition is trivially satisfied.

Theorem 3 *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 $\Pr\{\tilde{P} \succeq P\} > 0$, we have $\varphi[s(P)](v) R_v \mu(v)$ for all $v \in V$ and all $\mu \in C(s(P))$.*

Generally, the condition in Theorem 3 is not sufficient for a profile of strategies to be an OBNE.⁸ 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

⁸For 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 3 is satisfied but $s(P)$ is obviously not an OBNE. Any firm gains by revealing its true preference.

of the submitted profile to be a singleton since the ex-post unanimity is in terms of the true profile.

7 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 (and given the observed singleton cores), truth-telling remains a plausible behavior; 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, some features of the equilibrium behavior in these matching markets are independent of the chosen stable mechanism. This is significant since the two sides of the 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.

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.

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 (2002a).
- [8] L. Ehlers. “In Search of Advice for Participants in Matching Markets which Use the Deferred-Acceptance Algorithm”. Mimeo, Université de Montréal (2002b).
- [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”, Mimeo, Universitat Autònoma de Barcelona (2003).
- [12] B. Klaus, F. Klijn and J. Massó. “Some Things Couples Always Wanted to Know about Stable Matchings (But Were Afraid to Ask)”, Mimeo, Universitat Autònoma de Barcelona (2003).

- [13] D. Knuth. *Marriages Stables*. Les Presses de l'Université de Montréal, Montréal (1976).
- [14] M. Niederle and A. Roth. "Unraveling Reduces Mobility in a Labor Market: Gastroenterology with and without a Centralized Match", forthcoming in the *Journal of Political Economy* (2003).
- [15] A. Roth. "The Economics of Matching: Stability and Incentives", *Mathematics of Operations Research* **7**, 617-628 (1982).
- [16] A. 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).
- [17] A. Roth. "Misrepresentation and Stability in the Marriage Problem", *Journal of Economic Theory* **34**, 383-387 (1984b).
- [18] A. Roth. "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics", *Econometrica* **70**, 1341-1378 (2002).
- [19] A. 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).
- [20] A. Roth and U. Rothblum. "Truncation Strategies in Matching Markets- In Search of Advice for Participants", *Econometrica* **67**, 21-43 (1999).
- [21] A. Roth and M. Sotomayor. *Two-sided Matching: A Study in Game-Theoretic Modelling and Analysis*. Cambridge University Press, Cambridge, England. [Econometric Society Monograph] (1990).

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

APPENDIX

In this Appendix we prove Theorems 1, 2 and 3.

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 $v \in W$ and $P'_v \in \mathcal{P}_v$. By $|C(P)| = 1$,

$$DA_W[P] = \varphi[P]. \quad (2)$$

Let $P''_v \in \mathcal{P}_v$ be such that $A(P''_v) = \{v\}$ if $\varphi[P'_v, P_{-v}](v) \in F$ and $A(P''_v) = \emptyset$ if $\varphi[P'_v, P_{-v}](v) = v$. By stability of φ and by construction of P''_v

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

Because for DA_W a worker cannot gain by misrepresentation we have

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

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

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 $\Pr\{\tilde{P}_{-v} = P_{-v}\} > 0$ 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 $\Pr\{\tilde{P} = 3D P\} > 0$ 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 $DA_W[P](w) = 3D .f$ Let $P'_w \in \mathcal{P}_w$ be such that $P'_w|F = 3D_w \mathbb{P}$ and $A(P'_w) = 3D B(\mathbb{P}_w)$.

Let $P'_{-w} \in \mathcal{P}_{-w}$ be such that $\Pr\{\tilde{P} = 3D_w(\mathbb{P}'_{-w})\} > 0$. 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 $\varphi[P_w, P'_{-w}](w)R_w f'$. Then $\varphi[P_w, P'_{-w}] \in C(P_w, P'_{-w})$ implies $\varphi[P_w, P'_{-w}] \in C(P'_w, P'_{-w})$ since individual rationality of $\varphi[P_w, P'_{-w}]$ at profile (P_w, P'_{-w}) implies individual rationality of $\varphi[P_w, P'_{-w}]$ at profile (P'_w, P'_{-w}) and (\hat{w}, \hat{f}) blocks $\varphi[P_w, P'_{-w}]$ at profile (P'_w, P'_{-w}) implies (\hat{w}, \hat{f}) blocks $\varphi[P_w, P'_{-w}]$ at profile (P_w, P'_{-w}) as well. Thus, by $A(P'_w) = 3D B(f', P_w)$ and the fact that under any two stable matchings the set of unmatched agents is identical, $\varphi[P'_w, P'_{-w}](w)R_w f'$. We next show that $\varphi[P'_w, P_{-w}](w) = 3D .f$. Suppose $\varphi[P'_w, P_{-w}](w) = 3D w$. Then $DA[P'_w, P_{-w}](w) = 3D w$. Therefore

$$DA_W[P_w, P_{-w}](w) = 3D P'_w w = 3D DA[P'_w, P_{-w}](w),$$

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 $\varphi[P'_w, P_{-w}](w) = 3D .f$. Furthermore, $\Pr\{\tilde{P}_{-w}|P_w = 3D_w \mathbb{P}\} > 0$, $f'P_w\varphi[P](w)$, and $\varphi[P'_w, P_{-w}](w) = 3D .f$. 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 Truncation Strategies

Theorem 2 *Let \tilde{P} be a common belief, s be a profile of truncation strategies and φ be stable mechanism. Then s is an OBNE in the stable mechanism φ if and only if for any profile P in the support of \tilde{P} , $s(P)$ is a Nash equilibrium under complete information P in the direct preference revelation game induced by φ .*

Proof. Let φ be a stable mechanism and s be a profile of truncation strategies.

(\Leftarrow) Let $v \in V$ and $P_v \in \mathcal{P}_v$ be such that $\Pr\{\tilde{P}_v = 3D_v\} > 0$. Because for all $P_{-v} \in \mathcal{P}_{-v}$ such that $\Pr\{\tilde{P}_{-v} = 3D_{-v}(P_{-v})\} > 0$, $s_v(P_v)$ is a best response to $s_{-v}(P_{-v})$ under P_v in the direct preference revelation game, it follows that submitting $s_v(P_v)$ is a best response for agent v . Hence, $s(P)$ is an OBNE in the stable mechanism φ .

(\Rightarrow) Suppose that s is an OBNE. First, we show that the support of $s(\tilde{P})$ is contained in the set of all profiles with a singleton core. Let $v \in V$ and $P_v \in \mathcal{P}_v$ be such that $\Pr\{\tilde{P}_v = 3D_v\} > 0$. Let $s_v(P_v) = 3D_v$. Because s is an OBNE we have

$$\varphi[P'_v, s_{-v}(\tilde{P}_{-v}|P_v)](v) \succ_{P_v} \varphi[P''_v, s_{-v}(\tilde{P}_{-v}|P_v)](v) \quad (5)$$

for all $P''_v \in \mathcal{P}_v$. Without loss of generality, suppose that $v \in W$. By stability of φ , we have $\Pr\{\varphi[P'_v, s_{-v}(\tilde{P}_{-v}|P_v)](v) \in A(P'_v) \cup \{v\}\} = 3D_v 1$. Because \mathcal{B} a truncation of P_v , $A(P'_v) \subseteq A(P_v)$ and $\Pr\{\varphi[P'_v, s_{-v}(\tilde{P}_{-v}|P_v)](v) \in A(P_v) \cup \{v\}\} = 3D_v 1$. Furthermore, for all $f \in A(P'_v)$, $B(f, P'_v) = 3D_v B(f_v)$. Hence, by (5),

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

for all $P''_v \in \mathcal{P}_v$. Thus, submitting P'_v is a best response under $s_{-v}(\tilde{P}_{-v}|P_v)$ when v 's true preference is P'_v . Therefore, truth-telling is an OBNE if $s(\tilde{P})$ is the common belief. Hence, by Theorem 1, the support of $s(\tilde{P})$ is contained in the set of all profiles with a singleton core.

Second, suppose that P is in the support of \tilde{P} and $s(P)$ is not a Nash equilibrium under complete information P in the direct preference revelation game induced by φ . Then there is some $v \in V$ and $P'_v \in \mathcal{P}_v$ such that $\varphi[P'_v, s_{-v}(P_{-v})](v) \succ_{P_v} \varphi[s(P)](v)$. If

$\varphi[s(P)](v) \neq 3D v$, then by the fact that (\mathcal{P}_v) is a truncation of P_v , $\varphi[P'_v, s_{-v}(P_{-v})](v)$ is strictly $s_v(P_v)$ -preferred to $\varphi[s(P)](v)$. This is a contradiction since $|C(s(P))| = 3D 1$ and similarly to the proof of Theorem 1, $s(P)$ is a Nash equilibrium under complete information $s(P)$ in the direct preference revelation game induced by φ .

Hence, $\varphi[s(P)](v) = 3D v$. Because (\mathcal{P}_v) is a truncation of P_v and $\varphi[P'_v, s_{-v}(P_{-v})](v) \neq 3D v$, we have $\varphi[P'_v, s_{-v}(P_{-v})](v) \in A(P_v) \setminus A(s_v(P_v))$. We show that

$$\varphi[P_v, s_{-v}(P_{-v})](v) \neq 3D v. \quad (6)$$

If $\varphi[P_v, s_{-v}(P_{-v})](v) = 3D v$, then by the fact that (\mathcal{P}_v) is a truncation of P_v , $\varphi[P_v, s_{-v}(P_{-v})]$ is stable under $s(P)$. Because $|C(s(P))| = 3D 1$ and φ is stable, we have

$$\varphi[P_v, s_{-v}(P_{-v})] = 3D \varphi[s(P)]. \quad (7)$$

Let $P''_v \in \mathcal{P}_v$ be such that $A(P''_v) = 3D \{\varphi[s_v(P_v)](v)\}$. Then $\varphi[P'_v, s_{-v}(P_{-v})]$ is stable under $(P''_v, s_{-v}(P_{-v}))$. On the other hand, by (7), $\varphi[s(P)]$ is stable under $(P_v, s_{-v}(P_{-v}))$. Thus, by $A(P''_v) \subseteq A(P_v)$, $\varphi[s(P)]$ is stable under $(P''_v, s_{-v}(P_{-v}))$. Hence, $\varphi[s(P)]$ and $\varphi[P'_v, s_{-v}(P_{-v})]$ are stable under $(P''_v, s_{-v}(P_{-v}))$. Since $\varphi[s(P)](v) = 3D v$ and $\varphi[P'_v, s_{-v}(P_{-v})](v) \neq 3D v$, this contradicts the fact that the set of unmatched agents is identical at any two stable matchings.

Finally we show that

$$\Pr\{\varphi[P_v, s_{-v}(\tilde{P}_{-v}|_{P_v})](v) \in A(P_v)\} > \Pr\{\varphi[s_v(P_v), s_{-v}(\tilde{P}_{-v}|_{P_v})](v) \in A(P_v)\}, \quad (8)$$

which contradicts $s(P)$ being an OBNE. Note that $\Pr\{\tilde{P}_{-v}|_{P_v} = 3D \emptyset\} > 0$, and by (6), $\varphi[P_v, s_{-v}(P_{-v})](v) \neq 3D v = 3D \varphi[s(P)](v)$. Let \mathcal{P}'_{-v} be such that $\Pr\{\tilde{P}_{-v}|_{P_v} = 3D P'_{-v}\} > 0$ and $\varphi[s_v(P_v), s_{-v}(P'_{-v})](v) \neq 3D v$. Then by the fact that (\mathcal{P}_v) is a truncation of P_v , $\varphi[s_v(P_v), s_{-v}(P'_{-v})]$ is stable under $(P_v, s_{-v}(P'_{-v}))$. Because the set of unmatched agents is identical under any two stable matchings, $\varphi[P_v, s_{-v}(P'_{-v})](v) \neq 3D v$ and (8) follows. \blacksquare

We also show that in Theorem 2 the necessary and sufficient condition for a profile of truncation strategies to be an OBNE is independent of the stable mechanism.

Let s be a profile of truncation strategies and P be a profile. Let

$$\mathcal{T}(s(P)) \equiv \{\{w, f\} \mid w \in A(s_f(P_f)) \text{ or } f \in A(s_w(P_w))\}.$$

Let $C^{\mathcal{T}(s(P))}(P)$ denote the set of matchings which are not blocked under P by a coalition belonging to $\mathcal{T}(s(P))$.

Lemma 1 *Let φ be a stable mechanism, P be a profile, and $s(P)$ be a profile of truncations of P . Then $s(P)$ is a Nash equilibrium under complete information P in the direct preference revelation game induced by φ if and only if $|C(s(P))| = 3D 1$ and $C(s(P)) \subseteq C^{\mathcal{T}(s(P))}(P)$.*

Proof. (\Rightarrow) From the proof of Theorem 2 we know $|C(s(P))| = 3D 1$. Let $C(s(P)) = 3D \{\mu\}$. Suppose $\mu \notin C^{\mathcal{T}(s(P))}(P)$. Then there is some $\{w, f\} \in \mathcal{T}(s(P))$ such that (w, f) blocks μ under P . If both $w \in A(s_f(P_f))$ and $f \in A(s_w(P_w))$, then by the fact that $s(P)$ is a truncation of P , (w, f) blocks μ under $s(P)$, a contradiction to $\mu \in C(s(P))$. Without loss of generality, suppose $w \in A(s_f(P_f))$ and $f \notin A(s_w(P_w))$. Since (w, f) blocks μ under P and $s_w(P_w)$ is a truncation of P_w , we have $\mu(w) = 3D w$ and $f \in A(P_w) \setminus A(s_w(P_w))$.

Consider $(P_w, s_{-w}(P_{-w}))$. If $\varphi[P_w, s_{-w}(P_{-w})](w) \neq 3D w$, then $\varphi[P_w, s_{-w}(P_{-w})](w) P_w w$. Thus, by $\mu(w) = 3D w$ and $\varphi[s(P)] = 3D \mu$, $s(P)$ is not a Nash equilibrium under P in the preference revelation game induced by φ . Hence, $\varphi[P_w, s_{-w}(P_{-w})](w) = 3D w$. Since $s_w(P_w)$ is a truncation of P_w , $\varphi[P_w, s_{-w}(P_{-w})]$ is stable under $s(P)$. By $C(s(P)) = 3D \{\mu\}$, $\varphi[P_w, s_{-w}(P_{-w})] = 3D \mu$. However, by $w \in A(s_f(P_f))$ and (w, f) blocks μ under P , μ is blocked by (w, f) under $(P_w, s_{-w}(P_{-w}))$, a contradiction to $\mu \in C(P_w, s_{-w}(P_{-w}))$.

(\Leftarrow) Let $C(s(P)) = 3D \{\mu\}$ and $\mu \in C^{\mathcal{T}(s(P))}(P)$. By stability of φ , $\varphi[s(P)] = 3D \mu$. Suppose $s(P)$ is not a Nash equilibrium under complete information P in the direct preference revelation game induced by φ . Then there is some $v \in V$ and $P'_v \in \mathcal{P}_v$ such that $\varphi[P'_v, s_{-v}(P_{-v})](v) P'_v \mu(v)$. If $\mu(v) \neq 3D v$, then by the fact that $s_v(P_v)$ is a truncation of P_v , $\varphi[P'_v, s_{-v}(P_{-v})](v)$ is strictly $s_v(P_v)$ -preferred to $\mu(v)$. This is a contradiction since $|C(s(P))| = 3D 1$ and similarly to the proof of Theorem 1, $s(P)$ is a Nash equilibrium

under complete information $s(P)$ in the direct preference revelation game induced by φ .

Hence, $\mu(v) \neq v$. Let $P_v'' \in \mathcal{P}_v$ be such that $A(P_v'') = \{ \varphi_v[B_{-v}(P_{-v})](v) \}$. Then $\varphi[P_v', s_{-v}(P_{-v})]$ is stable under $(P_v'', s_{-v}(P_{-v}))$. Since $\varphi[P_v', s_{-v}(P_{-v})](v) \neq v$ and the set of unmatched agents is identical at any two stable matchings, μ is not stable under $(P_v'', s_{-v}(P_{-v}))$ (by $\mu(v) \neq v$). Without loss of generality, let $v \in W$ and $f \equiv \varphi[P_v', s_{-v}(P_{-v})](v)$. If there is a pair (w', f') with $w' \neq v$ which blocks μ under $(P_v'', s_{-v}(P_{-v}))$, then (w', f') blocks μ under $s(P)$, a contradiction to $C(s(P)) = \{\mu\}$. Thus, (v, f) must block μ under $(P_v'', s_{-v}(P_{-v}))$. Therefore, $v \in A(s_f(P_f))$ and $f \in A(P_v)$. Hence, $\{v, f\} \in \mathcal{T}(s(P))$ and (v, f) blocks μ under P , a contradiction to $\mu \in C^{\mathcal{T}(s(P))}(P)$. \blacksquare

C Small Cores

Theorem 3 *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 $s_v(P_v) = f$. Let $P_v'' \in \mathcal{P}_v$ be such that (i) $A(P_v'') = A(P_v) \cap B(f, P_v)$ and (ii) $P_v'' \setminus A(P_v'') = A(P_v) \setminus A(P_v)$. We show that

$$\Pr\{\varphi[P_v'', s_{-v}(\tilde{P}_{-v}|P_v)](v) \in B(f, P_v)\} > \Pr\{\varphi[P_v', s_{-v}(\tilde{P}_{-v}|P_v)](v) \in B(f, P_v)\}, \quad (9)$$

which contradicts the fact that s is an OBNE.

Let $P'_{-v} \in \mathcal{P}_{-v}$ be such that $\Pr\{\tilde{P}_{-v}|_{P_v} = 3D_{-v} \mathcal{P}\} > 0$ 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, $\varphi[P'_{-v}, s_{-v}(P'_{-v})](v) \in B(f, P_v)$ implies that $\varphi[P'_{-v}, s_{-v}(P'_{-v})](v) \neq 3D v$, and hence $\varphi''[B_{-v}(P'_{-v})](v) \neq 3D v$. Thus, $\varphi''[B_{-v}(P'_{-v})](v) \in A(P''_v)$, 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 $\mu(v) = 3D f$ and the set of unmatched agents is identical under any two stable matchings, $\varphi[P''_v, s_{-v}(P_{-v})](v) \neq 3D v$. By $A(P''_v) \subseteq B(f, P_v)$, $\varphi[P''_v, s_{-v}(P_{-v})](v) \in B(f, P_v)$. Furthermore, $\Pr\{\tilde{P}_{-v}|_{P_v} = 3D_{-v} \mathcal{P}\} > 0$ and $\varphi[P'_v, s_{-v}(P_{-v})](v) \notin \mu \in B(f, P_v)$. Hence, (9) is true. ■