Matching Markets under (In)complete Information*

Lars Ehlers[†] Jordi Massó[‡]

February 2007

*L. Ehlers acknowledges financial support form the SSHRC (Canada). The work of J. Massó is partially supported by the Spanish Ministry of Education and Science, through grant SEJ2005-01481/ECON, FEDER, and CONSOLIDER-INGENIO 2010 (CDS2006-00016), and by the Generalitat de Catalunya, through grant SGR2005-00454 and through the Barcelona Economics Program (XREA). Part of this research was done while J. Massó was visiting the Université de Montréal and while L. Ehlers was visiting the Universitat Autònoma de Barcelona; the visits were financed by CIREQ and XREA, respectively.

†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.masso@uab.es

Abstract

We are the first to introduce incomplete information to centralized many-to-one matching markets such as those to entry-level labor markets or college admissions. This is important because in real life markets (i) any agent is uncertain about the other agents' true preferences and (ii) most entry-level matching is many-to-one (and not one-to-one). We show that for stable (matching) mechanisms there is a strong and surprising link between Nash equilibria under complete information and Bayesian Nash equilibria under incomplete information. That is, given a common belief, a strategy profile is a Bayesian Nash equilibrium under incomplete information in a stable mechanism if and only if, for any true profile in the support of the common belief, the submitted profile is a Nash equilibrium under complete information at the true profile in the direct preference revelation game induced by the stable mechanism. This result may help to explain the success of stable mechanisms in these markets.

JEL Classification: C78, D81, J44.

Keywords: Many-To-One Matching Market, Stability, Incomplete Information.

1 Introduction

Both empirical and theoretical studies of two-sided matching markets have been useful in applications. Many such markets have developed centralized market clearing mechanisms (in response to various failures of the decentralized market) to match the agents from the two sides: the institutions (*firms*, colleges, hospitals, schools, etc.) and the individuals (*workers*, students, medical interns, children, etc.).[1] The National Resident Matching Program is the most well-studied example of this kind of two-sided matching markets. Each year around 20,000 medical students look for a four-years position in American hospital programs to undertake their medical internships.[2] In many countries, each year thousands of students seek for positions in colleges,[3] six years old children have to be assigned to public schools,[4] 8^{th} graders high school students to high schools,[5] as well as civil servants to similar jobs in public positions scattered in different cities across a country.

All of these entry-level matching markets share two specific features. The first one is the *many-to-one* nature of the problem: the workers enter the market by cohorts (often once per year) and each worker has to be matched to at most one firm while each firm might be matched to many workers. The second one is the *centralized* way of reaching a solution: a centralized institution (clearinghouse) collects, for each participant, a ranked list of potential partners and proposes, after processing the profile of submitted ranked lists, a final matching between firms and workers.

[1]Roth and Sotomayor (1990) give a masterful overview of two-sided matching markets.

[2]See Roth (1984a), Roth and Peranson (1999), and Roth (2002) for a careful description and analysis of this market. Roth (1991), Ehlers (2002), Kesten (2004), and Ünver (2005) describe and analyze the equivalent UK markets.

[3]Romero-Medina (1998) studies the case of Spain.

[4]Chen and Sönmez (2006) and Ergin and Sönmez (2006) study the case of public schools in Boston. Abdulkadiroğlu and Sönmez (2003) studies the cases of public schools in Boston, Lee County (Florida), Minneapolis, and Seattle.

[5]Abdulkadiroğlu, Pathak, and Roth (2005) studies the case of public high schools in New York City.

Yet, and in order to survive, the proposed matching has to be *stable* (relative to the true preference profile) in the sense that all agents have to be matched to acceptable partners and no unmatched pair of a firm and a worker prefer each other rather than the proposed partners. Stability constitutes a minimal requirement that a matching has to fulfill if the assignment is voluntary rather than compulsory. The literature has considered stability of a matching to be its main characteristic in order to survive.[6] Indeed, many of the successful mechanisms are stable. This is puzzling because there exists no stable mechanism which makes truth-telling a dominant strategy for all agents (Roth, 1982). Therefore, an agent's (submitted) ranked lists of potential partners are not necessarily his true ones and the implemented matching may not be stable for the true profile. As a consequence, the literature has studied intensively Nash equilibria of direct preference revelation games induced by different stable mechanisms for a given true preference profile.[7] Not only that, there is also a fair amount of agreement that these studies have provided us with a very good understanding of the strategic incentives that participants face in these markets under complete information.

Nevertheless all this strategic analysis might be marred by the assumption that the true profile of preferences is both certain and common knowledge among all agents; the very definition of Nash equilibrium under complete information requires it. Indeed, participants in these markets perceive the outcome of the mechanism as being uncertain because the submitted preferences of the other participants are unknown. To model this uncertainty and to overcome the limitation of the complete information set up, we follow the Bayesian approach by assuming that participants share a common belief; namely, nature selects a preference profile according to a commonly known probability distribution on the set of profiles. Since matching markets require to report ranked lists and not their specific utility representations, we stick to the

[6]See, for instance, Roth (1984a) and Niederle and Roth (2003).

[7]See Dubins and Freedman (1981), Roth (1982, 1984b, 1985a), Gale and Sotomayor (1985), Shin and Suh (1996), Sönmez (1997), Ma (1995, 2002), and Alcalde (1996).

ordinal setting and assume that probability distributions are evaluated according to the first-order stochastic dominance criterion. Then, a strategy profile is an ordinal Bayesian Nash equilibrium if, for every von Neumann-Morgenstern utility function of an agent's preference ordering (his type), the submitted ranked list maximizes his expected utility in the direct preference revelation game induced by the common belief and the mechanism.[8]

Investigating many-to-one matching markets under incomplete information is important for applications because in real life markets (i) any agent is uncertain about the other agents' true preferences and (ii) most entry-level matching is many-to-one (and not one-to-one). More precisely, we study in many-to-one matching markets direct preference revelation games under incomplete information induced by a stable mechanism. Our main result shows that there is a strong and surprising link between Nash equilibria under complete information and ordinal Bayesian Nash equilibria under incomplete information. More precisely, Theorem 1 states that, given a common belief, a strategy profile is an ordinal Bayesian Nash equilibrium under incomplete information in a stable mechanism if and only if for any profile in the support of the common belief, the submitted profile is a Nash equilibrium under complete information at the true profile in the direct preference revelation game induced by the stable mechanism.

Theorem 1 has many important consequences and applications. The most important consequence of this result is that it points out that, after all, the former strategic analysis under complete information is meaningful, relevant, and essential to undertake the corresponding analysis under incomplete information. Furthermore, for determining whether a strategy profile is an equilibrium under incomplete information, we only need to check whether for each realization the submitted preference

[8] 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. Majumdar (2003), Ehlers and Massó (2004), and Pais (2005) have already used this ordinal equilibrium notion in one-to-one matching markets.

orderings are a Nash equilibrium under complete information. This also implies that for any stable mechanism, the set of ordinal Bayesian Nash equilibria is identical for any two common beliefs with equal support. Therefore, any equilibrium is robust to perturbations of the common belief which do not change the support of the common belief and agents may have different beliefs with equal support.

Another important consequence is that the set of ordinal Bayesian Nash equilibria for common beliefs with full support remain equilibria for any common belief. We show that full support equilibria provide a foundation why any agent submits only preference orderings which rank acceptable only partners which are acceptable according to his true preference relation and the reported ranking over the acceptable partners is truthful. This may help to explain why in markets using stable mechanisms most agents truthfully reveal their preferences over their partners reported acceptable (Roth and Peranson, 1999). It also gives some insight into the success of stable mechanisms since exactly these equilibria are robust to arbitrary changes of the (non-)common belief. Furthermore, we apply our main result to obtain conclusions about the stability of the outcomes realized under any ordinal Bayesian Nash equilibrium and when truth-telling is an ordinal Bayesian Nash equilibrium.[9]

The paper is organized as follows. Section 2 describes the many-to-one matching market with responsive preferences. Section 3 introduces the incomplete information framework to the many-to-one matching market and the notion of ordinal Bayesian Nash equilibrium. Section 4 states our main result, Theorem 1, and its applications. Section 5 concludes with some final remarks and the Appendix contains the proof of Theorem 1.

[9]We will describe these conclusions in detail later in the main text.

2 Many-To-One Matching Markets

2.1 Agents, Quotas, and Preferences

The *agents* of a college admissions problem (or many-to-one matching market) consist of two disjoint sets, the set of *firms* F and the set of *workers* W. A generic firm will be denoted by f, a generic worker by w, and a generic agent by $v \in V \equiv F \cup W$. While workers can only work for at most one firm, firms may hire different numbers of workers. For each firm f, there is a maximum number $q_f \geq 1$ of workers that f may hire, f's *quota*. Let $q = (q_f)_{f \in F}$ be the vector of quotas. To emphasize the quotas of a subset of firms $S \subseteq F$ we sometimes write (q_S, q_{-S}) instead of q. Each worker w has a strict preference ordering P_w over $F \cup \{\emptyset\}$, where $\emptyset$ means the prospect of not being hired by any firm. Each firm f has a strict preference ordering P_f over $W \cup \{\emptyset\}$, where $\emptyset$ means the prospect of not hiring any worker. A *profile* $P = (P_v)_{v \in V}$ is a list of preference orderings. To emphasize the preference orderings of a subset of agents $S \subseteq V$ we often denote a profile P by (P_S, P_{-S}). Let $\mathcal{P}_v$ be the set of all preference orderings of agent v. Let $\mathcal{P} = \times_{v \in V} \mathcal{P}_v$ be the set of all profiles and let $\mathcal{P}_{-v}$ denote the set $\times_{v' \in V \setminus \{v\}} \mathcal{P}_{v'}$. Since agent v might have to compare potentially the same partner, we denote by R_v the weak preference ordering corresponding to P_v; namely, for $v', v'' \in V \cup \{\emptyset\}$, $v' R_v v''$ means either $v' = v''$ or $v' P_v v''$. Momentarily fix a worker w and his preference ordering P_w. Given $v \in F \cup \{\emptyset\}$, let $B(v, P_w)$ be the *weak upper contour set* of P_w at v; *i.e.*, $B(v, P_w) = \{v' \in F \cup \{\emptyset\} \mid v' R_w v\}$. Let $A(P_w)$ be the set of *acceptable* firms for w according to P_w; *i.e.*, $A(P_w) = \{f \in F \mid f P_w \emptyset\}$. Given a subset $S \subseteq F \cup \{\emptyset\}$, let $P_w|S$ denote the restriction of P_w to S. Similarly, given $P_f \in \mathcal{P}_f$, $v \in W \cup \{\emptyset\}$, and $S \subseteq W \cup \{\emptyset\}$, we define $B(v, P_f)$, $A(P_f)$, and $P_f|S$. A *college admissions problem (or many-to-one matching market)* is a quadruple (F, W, q, P).

2.2 Stable Matchings

The assignment problem consists of matching workers with firms keeping the bilateral nature of their relationship, complying with firms' capacities given by their quotas, and allowing for the possibility that both workers and firms may remain unmatched. Formally, given a college admissions problem (F, W, q, P), a *matching* μ is a mapping from the set V into the set of all subsets of V such that:

(m1) for all $w \in W$, either $|\mu(w)| = 1$ and $\mu(w) \subseteq F$ or else $\mu(w) = \emptyset$;

(m2) for all $f \in F$, $\mu(f) \subseteq W$ and $|\mu(f)| \leq q_f$; and

(m3) $\mu(w) = \{f\}$ if and only if $w \in \mu(f)$.

Abusing notation, we will often write $\mu(w) = f$ instead of $\mu(w) = \{f\}$. If $\mu(w) = \emptyset$ we say that w is *unmatched* at μ and if $|\mu(f)| < q_f$ we say that f has $q_f - |\mu(f)|$ *unfilled* positions at μ; f is *unmatched* at μ when it has q_f unfilled positions at μ. Let $\mathcal{M}$ denote the set of all matchings. A college admissions problem (F, W, q, P) in which $q_f = 1$ for all $f \in F$ is called a *marriage market* or a *one-to-one matching market*.

Not all matchings are equally likely. Stability of a matching is considered to be its main characteristic in order to survive. A matching is stable if no agent is matched to an unacceptable partner (*individual rationality*) and no unmatched worker-firm pair mutually prefers each other to (one of) their current assignments (*pair-wise stability*). That is, given a college admissions problem (F, W, q, P), a matching $\mu \in \mathcal{M}$ is *stable* (at P) if

(s1) for all $w \in W$, $\mu(w) R_w \emptyset$;

(s2) for all $f \in F$ and all $w \in \mu(f)$, $w P_f \emptyset$; and

(s3) there is no pair $(w, f) \in W \times F$ such that $w \notin \mu(f)$, $f P_w \mu(w)$, and either $w P_f w'$ for some $w' \in \mu(f)$ or $w P_f \emptyset$ if $|\mu(f)| < q_f$.

Notice that this definition declares a matching to be stable if it is not blocked (in the sense of the core) by either individual agents or unmatched pairs. Gale and Shapley (1962) established that all college admissions problems have a non-empty set of stable matchings and Roth (1985b) showed that larger coalitions do not have additional (weak) blocking power because the set of stable matchings coincides with the core. We denote by $C(F, W, q, P)$ the non-empty core of the college admissions problem (F, W, q, P). Since sometimes everything but P remains fixed we will often write P instead of (F, W, q, P); then, for instance, $C(P)$ denotes the set of stable matchings at P (or the core of P).

2.3 Matching Mechanisms

Whether or not a matching is stable depends on the preference orderings of agents, and since they are private information, agents have to be asked about them. A mechanism requires each agent v to report some preference ordering P_v and associates a matching with any reported profile P. Namely, a *mechanism* is a function $\varphi : \mathcal{P} \to \mathcal{M}$ mapping each preference profile $P \in \mathcal{P}$ to a matching $\varphi[P] \in \mathcal{M}$. Then $\varphi[P](v)$ is the match of agent v at preference profile P under mechanism φ. Note that, for all $w \in W$, $\varphi[P](w) \in F \cup \{\emptyset\}$ and, for all $f \in F$, $\varphi[P](f) \in 2^W$. A mechanism φ is stable if for all $P \in \mathcal{P}$, $\varphi[P] \in C(P)$.

2.4 Responsive Extensions

The notion of a mechanism in which firms (like workers) only submit rankings on individual agents fits with most of the mechanisms used in real life centralized matching markets. But a mechanism matches each firm f to a *set* of workers, taking into account only f's preference ordering P_f over individual workers. Thus, to study firms' incentives in direct preference revelation games induced by a mechanism, preference orderings of firms over individual workers have to be extended to preference orderings over subsets of workers. But a firm f may have different rankings over subsets of workers respecting its quota q_f and the ranking P_f over individual workers. For instance, let $W = \{w_1, w_2, w_3, w_4\}$ be the set of workers and let P_f be such that $P_f : w_1w_2w_3w_4\emptyset$[10] and $q_f = 2$. While it is reasonable to assume that, under the absence of very strong complementarities among workers, the set $\{w_1, w_2\}$ is preferred by f to the set $\{w_3, w_4\}$ or to the set $\{w_1, w_3\}$, firm f's preference between the sets $\{w_1, w_4\}$ and $\{w_2, w_3\}$ is ambiguous since P_f does not convey this information. Following the literature,[11] we will only require these extensions to be responsive in the sense that replacing a worker in a set (or an unfilled position) by a better worker (or an acceptable worker) makes a set more preferred; for example, in all extensions $\{w_1, w_2\}$ is preferred to $\{w_1\}$, to $\{w_3, w_4\}$ and to $\{w_1, w_3\}$ but for some extensions $\{w_1, w_4\}$ is preferred to $\{w_2, w_3\}$ while for other extensions $\{w_2, w_3\}$ is preferred to $\{w_1, w_4\}$.

Definition 1 (Responsive Extensions) The preference extension P_f^* over 2^W is *responsive* to the preference ordering P_f over $W \cup \{f\}$ if for all $S \in 2^W$, all $w \in S$, and all $w' \notin S$:
(r1) $S \cup \{w'\} P_f^* S$ if and only if $|S| < q_f$ and $w' P_f \emptyset$.
(r2) $S \cup \{w'\} P_f^* S \backslash \{w\}$ if and only if $w' P_f w$.

Given a responsive extension P_f^* of P_f, let R_f^* denote its corresponding weak preference ordering on 2^W. Moreover, given $S \in 2^W$, let $B(S, P_f^*)$ be the *weak upper contour set of* P_f^* at S; *i.e.*, $B(S, P_f^*) = \{S' \in 2^W \mid S' R_f^* S\}$. Given $P_f \in \mathcal{P}_f$, we denote by $resp(P_f)$ the set of responsive extensions of P_f.

2.5 Properties of the Core

The core of a college admissions problem has a special structure. The following well-known properties will be useful in the sequel:[12]

[10] We will use the convention that $P_f : w_1w_2w_3w_4\emptyset$ means $w_1P_fw_2P_fw_3P_fw_4P_f\emptyset$.
[11] See for instance, Roth and Sotomayor (1990).
[12] See Roth and Sotomayor (1990) for a detailed presentation of these properties.

(P1) For each profile $P \in \mathcal{P}$, $C(P)$ contains two stable matchings, the firms-optimal stable matching μ_F and the workers-optimal stable matching μ_W, with the property that for all $\mu \in C(P)$, $\mu_W(w) R_w \mu(w) R_w \mu_F(w)$ for all $w \in W$, and for all $f \in F$, $\mu_F(f) R_f^* \mu(f) R_f^* \mu_W(f)$ for all $P_f^* \in resp(P_f)$. The *deferred-acceptance algorithms (DA-algorithms)*, introduced by Gale and Shapley (1962) and denoted by $DA_F : \mathcal{P} \to \mathcal{M}$ and $DA_W : \mathcal{P} \to \mathcal{M}$, are two stable mechanisms that select, for each profile P, μ_F and μ_W, respectively; *i.e.*, for all $P \in \mathcal{P}$, $DA_F[P] = \mu_F$ and $DA_W[P] = \mu_W$.[13]

(P2) For each profile $P \in \mathcal{P}$ and any responsive extensions $P_F^* = (P_f^*)_{f \in F}$ of $P_F = (P_f)_{f \in F}$, $C(P)$ coincides with the set of group stable matchings at (P_W, P_F^*), where group stability corresponds to the usual cooperative game theoretical notion of weak blocking[14]. This is important because it means that the set of group stable matchings (relative to P) is invariant with respect to any specific responsive extensions of P_F.

(P3) For each $P \in \mathcal{P}$, the set of unmatched agents is the same for all stable matchings and if a firm does not fill all its positions at some stable matching, then this firm is matched to the same set of workers at all stable matchings; namely, for all $\mu, \mu' \in C(P)$, and for all $w \in W$ and all $f \in F$, (i) $\mu(w) = \emptyset$ if and only if $\mu'(w) = \emptyset$, (ii) $|\mu(f)| = |\mu'(f)|$, and (iii) if $|\mu(f)| < q_f$, then $\mu(f) = \mu'(f)$.

3 Incomplete Information

Clearly any mechanism and any true profile define a direct (ordinal) preference revelation game under complete information.

Definition 2 (Nash Equilibrium) A profile P' is a *Nash equilibrium (NE)* un-

[13]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.

[14]A matching μ is weakly blocked by coalition $S \subseteq V$ under (P_W, P_F^*) if there exists a matching μ' such that (b1) for all $v \in S$, $\mu'(v) \subseteq S$, (b2) for all $w \in W \cap S$, $\mu'(w) R_w \mu(w)$, and (b3) for all $f \in F \cap S$, $\mu'(f) R_f^* \mu(f)$, with strict preference holding for at least one $v \in S$.

der complete information P in the direct preference revelation game induced by the mechanism φ if for all $w \in W$, $\varphi[P'](w) R_w \varphi[\hat{P}_w, P'_{-w}](w)$ for all $\hat{P}_w \in \mathcal{P}_w$, and for all $f \in F$ and all $P^*_f \in resp(P_f)$, $\varphi[P'](f) R^*_f \varphi[\hat{P}_f, P'_{-f}](f)$ for all $\hat{P}_f \in \mathcal{P}_f$.

A large literature on matching studies Nash equilibrium and its refinements under complete information in direct preference revelation games induced by stable mechanisms; in particular, for the mechanisms DA_F and DA_W. However, for many applications the assumption that the true profile is common knowledge is extremely unrealistic. We depart from it and consider the Bayesian direct preference revelation games induced by a mechanism and a belief about the true profile, which is shared among all agents. A *common belief* is a probability distribution $\tilde{P}$ over $\mathcal{P}$. Given a profile P and the common belief $\tilde{P}$, $\Pr\{\tilde{P} = P\}$ is the probability that $\tilde{P}$ assigns to the event that the true profile is P. Given $v \in V$, let $\tilde{P}_v$ denote the marginal distribution of $\tilde{P}$ over $\mathcal{P}_v$. Observe that, following the Bayesian approach, the common belief $\tilde{P}$ describes agents' uncertainty about the true profile before agents learn their types. Now, given a common belief $\tilde{P}$ and a preference ordering P_v (agent v's type), let $\tilde{P}_{-v}|_{P_v}$ denote the probability distribution which $\tilde{P}$ induces over $\mathcal{P}_{-v}$ conditional on P_v. It describes agent v's uncertainty about the preferences of the other agents, given that his preference ordering is P_v. This formulation does not require symmetry nor independence of beliefs; conditional beliefs might be very correlated if agents use similar sources to form them (*i.e.*, rankings, grades, recommendation letters, etc.).

An agent with incomplete information about the others' preference orderings (more importantly, about their submitted lists) will perceive the outcome of a mechanism as being uncertain. A random matching $\tilde{\mu}$ is a probability distribution over the set of matchings $\mathcal{M}$. Given a matching μ and the random matching $\tilde{\mu}$, $\Pr\{\tilde{\mu} = \mu\}$ is the probability that $\tilde{\mu}$ assigns to matching μ. But the uncertainty important for agent v is not over matchings but over v's set of potential partners. Let $\tilde{\mu}(w)$ denote the probability distribution which $\tilde{\mu}$ induces over worker w's set of potential partners $F \cup \{\emptyset\}$ and let $\tilde{\mu}(f)$ denote the probability distribution which $\tilde{\mu}$ induces over firm

f's set of potential partners 2^W. Namely, for $w \in W$ and all $v \in F \cup \{\emptyset\}$,

$$\Pr\{\tilde{\mu}(w) = v\} = \sum_{\mu \in \mathcal{M}:\mu(w)=v} \Pr\{\tilde{\mu} = \mu\}$$

and for $f \in F$ and all $S \in 2^W$,

$$\Pr\{\tilde{\mu}(f) = S\} = \sum_{\mu \in \mathcal{M}:\mu(f)=S} \Pr\{\tilde{\mu} = \mu\}.$$

A mechanism φ and a common belief $\tilde{P}$ define a direct (ordinal) preference revelation game under incomplete information as follows. Before submitting a list to the mechanism, agents learn their types. Thus, a strategy of agent v is a function $s_v : \mathcal{P}_v \to \mathcal{P}_v$ specifying for each type of agent v, P_v, a list that v submits to the mechanism, $s_v(P_v)$. A *strategy profile* is a list $s = (s_v)_{v \in V}$ of strategies specifying for each true profile P a submitted profile $s(P)$. 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 $\varphi[s(\tilde{P})]$ 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} and $\tilde{P}$ induce conditional on P_v). But again, the relevant uncertainty that agent v faces is given by $\varphi[s_v(P_v), s_{-v}(\tilde{P}_{-v}|_{P_v})](v)$, the probability 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.

Definition 3 (First-Order Stochastic Dominance) (fo1) A random matching $\tilde{\mu}$ *first-order stochastically* P_w-*dominates* a random matching $\tilde{\mu}'$, denoted by $\tilde{\mu}(w) \succ_{P_w} \tilde{\mu}'(w)$, if for all $v \in F \cup \{\emptyset\}$,

$$\sum_{v' \in F \cup \{\emptyset\}:v' R_w v} \Pr\{\tilde{\mu}(w) = v'\} \geq \sum_{v' \in F \cup \{\emptyset\}:v' R_w v} \Pr\{\tilde{\mu}'(w) = v'\}.$$

(fo2)[15] A random matching $\tilde{\mu}$ *first-order stochastically* P_f-*dominates* a random match-

[15]Observe that this definition requires that $\tilde{\mu}$ first-order stochastically dominates $\tilde{\mu}'$ according to

ing $\tilde{\mu}'$, denoted by $\tilde{\mu}(f) \gtrdot_{P_f} \tilde{\mu}'(f)$, if for all $P_f^* \in resp(P_f)$ and all $S \in 2^W$,

$$\sum_{S' \in 2^W : S' R_f^* S} \Pr\{\tilde{\mu}(f) = S'\} \geq \sum_{S' \in 2^W : S' R_f^* S} \Pr\{\tilde{\mu}'(f) = S'\}.$$

All mechanisms used in centralized matching markets are ordinal. In other words the only information available for a clearinghouse are the agents' ordinal preferences over potential partners. In such an environment a strategy profile is an ordinal Bayesian Nash equilibrium whenever, for any agent's true ordinal preference, submitting the ranked list specified by his strategy maximizes his expected utility for every von Neumann-Morgenstern (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 (if any). Moreover, 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 4 (Ordinal Bayesian Nash Equilibrium) Let $\tilde{P}$ be a common belief. Then a strategy profile s is an *ordinal Bayesian Nash equilibrium* (*OBNE*) in the mechanism φ under incomplete information $\tilde{P}$ if and only if for all $v \in V$ and all $P_v \in \mathcal{P}_v$ such that $\Pr\{\tilde{P}_v = P_v\} > 0$,

$$\varphi[s_v(P_v), s_{-v}(\tilde{P}_{-v}|_{P_v})](v) \gtrdot_{P_v} \varphi[P_v', s_{-v}(\tilde{P}_{-v}|_{P_v})](v) \quad \text{for all } P_v' \in \mathcal{P}_v.^{16} \qquad (1)$$

Observe that, given a common belief $\tilde{P}$, the set of OBNE in a stable mechanism is non-empty. For instance, the strategy profile in which all agents declare that no agent in the other side of the market is acceptable is an OBNE under any common belief

all responsive extensions of P_f. Note that this requirement is meaningful since the clearinghouse observes firms' rankings over individual workers only and not which responsive extension they use to compare sets of workers.

[16] In the definition of 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.

since a stable mechanism selects, at all profiles P and (P_{-v}, P'_v), the empty matching. Furthermore, any matching μ can be connected to an OBNE s^μ in a stable mechanism in the following way: for any $v \in V$ and any $P_v \in \mathcal{P}_v$, let $A(s^\mu_v(P_v)) = \mu(v) \cap A(P_v)$. Then s^μ is an OBNE in a stable mechanism under any common belief because for any true preference relation, agent v reports acceptable exactly the partner(s) which are both specified by μ and are acceptable under his true preference relation. 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 matchings (Roth, 1985a). Both under complete and incomplete information there is a multiplicity of OBNE and the existence of OBNE is guaranteed.

4 The Main Result and Its Applications

The support of a common belief $\tilde{P}$ is the set of profiles on which $\tilde{P}$ puts a positive weight; namely, profile P belongs to the support of $\tilde{P}$ if and only if $\Pr\{\tilde{P} = P\} > 0$.

We will show that for stable mechanisms there is a strong and surprising link between equilibria under incomplete information and equilibria under complete information. Note that this link holds for any stable mechanism and not only for the deferred-acceptance algorithms.

Theorem 1 *Let $\tilde{P}$ be a common belief, s be a strategy profile, and φ be a stable mechanism. Then, s is an OBNE in the stable mechanism φ under incomplete information $\tilde{P}$ 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 1 has several important consequences and applications. One immediate consequence is that for determining whether a strategy profile is an OBNE, we only need to check whether for each realization of the common belief the submitted preference orderings constitute a Nash equilibrium under complete information. This means that the uniquely relevant information for an OBNE is the support of the common belief. Therefore, no calculations of probabilities are necessary. This consequence is very important for applications because we need to check equilibrium play only for the realized (or observed) profiles. Furthermore, by Theorem 1, we can use properties of NE (under complete information) to deduce characteristics of OBNE. Below we turn to the applications of Theorem 1.

4.1 Application I: Structure of OBNE

By Theorem 1, a strategy profile is an OBNE if and only if the agents play a Nash equilibrium for any profile in the support of the common belief. Therefore, (i) the set of OBNE is identical for any two common beliefs with equal support and (ii) the set of OBNE shrinks if the support of the common belief becomes larger.

Corollary 1 (Invariance) *Let s be a strategy profile and φ be a stable mechanism.*

(a) Let $\tilde{P}$ and $\tilde{P}'$ be two common beliefs with equal support. Then, s is an OBNE in the stable mechanism φ under $\tilde{P}$ if and only if s is an OBNE in the stable mechanism φ under $\tilde{P}'$.

(b) Let $\tilde{P}$ and $\tilde{P}'$ be two common beliefs such that the support of $\tilde{P}'$ is contained in the support of $\tilde{P}$. If s is an OBNE in the stable mechanism φ under $\tilde{P}$, then s is an OBNE in the stable mechanism φ under $\tilde{P}'$.

Now by (a) of Corollary 1, for stable mechanisms any OBNE is robust to perturbations of the common belief which leave its support unchanged. Therefore, any OBNE remains an equilibrium if agents have different beliefs with equal support, i.e. each agent v may have a private belief $\tilde{P}^v$ but all private beliefs have identical (or common) support.[17] This consequence is especially important for applications since for many of them, the common belief assumption might be too strong.

[17] Then in Definition 4 of OBNE $\tilde{P}$ is replaced by $\tilde{P}^v$ for each agent v. Theorem 1 and its proof show that for any OBNE s, each agent's strategy s_v chooses a best response to the other reported preferences for any profile belonging to the support of his private belief. If all private beliefs have

By (b) of Corollary 1, the set of OBNE with full support (*i.e.* all common beliefs which put positive probability on all profiles) is contained in the set of OBNE of any arbitrary common belief (or support). It turns out that OBNE with full support provide a foundation of why any agent submits only rankings which according to his true preference relation (i) contain only acceptable matches and (ii) report the true ranking over the reported acceptable matches. For the firms (ii) requires an inessential modification: because we consider only stable mechanisms it is irrelevant for a firm in which order it ranks its first q_f acceptable matches. For OBNE with full support any firm submits only rankings which are essentially truthful: the first q_f reported workers are the q_f truthfully most preferred workers among all workers reported acceptable and the reported ranking over the remaining workers reported acceptable is truthful.

Formally, given $v \in F$ and $P_v, P'_v \in \mathcal{P}_v$, we call $P'_v|A(P'_v)$ essentially P_v-truthful if $|A(P'_v)| \leq q_v$ or for the q_v most preferred workers under P'_v, say $w_1, \ldots, w_{q_v}$, we have for all $w' \in A(P'_v)$ and all $w \in A(P'_v)\backslash\{w_1, \ldots, w_{q_v}\}$, $P'_v|\{w, w'\} = P_v|\{w, w'\}$. For example, if $q_v = 2$ and $P_v : w_1 w_2 w_3 w_4 \emptyset \ldots$, then $P'_v : w_3 w_2 w_4 \emptyset \ldots$ and $P''_v : w_2 w_1 w_4 \emptyset \ldots$ are essentially P_v-truthful. Observe that condition (i) above will require in addition that $A(P'_v) \subseteq A(P_v)$ and $A(P''_v) \subseteq A(P_v)$.

Corollary 2 (Essential Truthfulness for Full Support) *Let $\tilde{P}$ be a common belief with full support, s be a strategy profile, and φ be a stable mechanism. Then, s is an OBNE in the stable mechanism φ under $\tilde{P}$ only if for all $v \in V$ and all $P_v \in \mathcal{P}_v$, (i) $A(s_v(P_v)) \subseteq A(P_v)$ and (ii) $s_v(P_v)|A(s_v(P_v)) = P_v|A(s_v(P_v))$ (if $v \in W$) and $s_v(P_v)|A(s_v(P_v))$ is essentially P_v-truthful (if $v \in F$).*

Proof. Let s be an OBNE in the mechanism φ under $\tilde{P}$. Let $v \in V$ and $P_v \in \mathcal{P}_v$. Assume $v \in F$ (if $v \in W$ the proof follows a similar argument).

equal support, then it follows that a strategy profile s is an OBNE with private beliefs (with common support) if and only if for any profile P in the common support, $s(P)$ is a Nash equilibrium under complete information P in the direct preference revelation game induced by φ.

First we show that $A(s_v(P_v)) \subseteq A(P_v)$. Suppose that $A(s_v(P_v))\backslash A(P_v) \neq \emptyset$. Let $w \in A(s_v(P_v))\backslash A(P_v)$ and $P_{-v} \in \mathcal{P}_{-v}$ be such that $A(P_w) = \{v\}$ and for all $v' \in V\backslash\{v, w\}$, $A(P_{v'}) = \emptyset$. Let $P = (P_v, P_{-v})$. Because $\tilde{P}$ has full support, we have $\Pr\{\tilde{P} = P\} > 0$. Thus, by Theorem 1, $s(P)$ must be a NE in φ for P. But then for all $v' \in V\backslash\{v, w\}$, $A(P_{v'}) = \emptyset$ implies $\varphi[s(P)](v') = \emptyset$. This and $w \notin A(P_v)$ implies $\varphi[s(P)](v) = \emptyset$ and $\varphi[s(P)](w) = \emptyset$. Hence, by stability of φ, we have $v \notin A(s_{v'}(P_{v'}))$ for all $v' \in A(s_v(P_v))$. But now w profitably deviates by reporting $P'_w \in \mathcal{P}_w$ such that $A(P'_w) = \{v\}$ because by $w \in A(s_v(P_v))$, $\varphi[P'_w, s_{-w}(P_{-w})](w) = v$ and $vP_w\emptyset = \varphi[s(P)](w)$. This means that $s(P)$ is not a NE in φ for P, a contradiction.

Second we show that $s_v(P_v)|A(s_v(P_v))$ is essentially P_v-truthful. If $|A(s_v(P_v))| \leq q_v$, then nothing has to be shown. Let $|A(s_v(P_v))| > q_v$ and $w_1, \ldots, w_{q_v}$ be the q_v most preferred workers under $s_v(P_v)$. Let $W' = \{w_1, \ldots, w_{q_v}\}$. By $A(s_v(P_v)) \subseteq A(P_v)$, if (ii) does not hold, then for some $w' \in A(s_v(P_v))$ and some $w \in A(s_v(P_v))\backslash W'$, $w's_v(P_v)ws_v(P_v)\emptyset$ and $wP_vw'P_v\emptyset$.[18] Without loss of generality, let $w' \in W'$ (if $w' \notin W'$, then the proof is analogous). Let $P_{-v} \in \mathcal{P}_{-v}$ be such that (a) $A(P_w) = \{v\}$, (b) $A(P_{w'}) = \{v\}$, (c) for all $w'' \in W'$, $A(P_{w''}) = \{v\}$, and (d) for all $v' \in V\backslash(\{v, w, w'\} \cup W')$, $A(P_{v'}) = \emptyset$. Let $P = (P_v, P_{-v})$. Because $\tilde{P}$ has full support, we have $\Pr\{\tilde{P} = P\} > 0$. Thus, by Theorem 1, $s(P)$ must be a NE in φ for P. But then for all $v' \in V\backslash(\{v, w, w'\} \cup W')$, $A(P_{v'}) = \emptyset$ implies $\varphi[s(P)](v') = \emptyset$. Furthermore, because $\tilde{P}$ has full support and s is an OBNE in φ under $\tilde{P}$, it is easy to verify that $v \in A(s_{w''}(P_{w''}))$ for $w'' \in W' \cup \{w\}$. Then by stability of φ, $w's_v(P_v)ws_v(P_v)\emptyset$, $W' \subseteq A(P_v)$, $A(P_{w'}) = \{v\}$, and the fact that $s(P)$ is a NE in φ for P, we must have $\varphi[s(P)](v) = W'$. Since $v \in A(s_w(P_w))$, now v profitably deviates by reporting $P'_v \in \mathcal{P}_v$ such that $A(P'_v) = (W'\backslash\{w'\}) \cup \{w\}$ because by $v \in A(s_w(P_w))$, $\varphi[P'_v, s_{-v}(P_{-v})](v) = (W'\backslash\{w'\}) \cup \{w\}$ and both wP_vw' and responsiveness imply $(W'\backslash\{w'\}) \cup \{w\}P^*_vW' = \varphi[s(P)](v)$ for all $P^*_v \in resp(P_v)$. This means that $s(P)$ is not a NE in φ for P, a contradiction. □

[18]Observe that if $v \in W$ the contradiction hypothesis would be that for some $f, f' \in A(s_v(P_v))$, $f's_v(P_v)fs_v(P_v)\emptyset$ and $fP_vf'P_v\emptyset$.

Note that any OBNE for a common belief with full support is an OBNE for any arbitrary belief. Hence, such OBNE are invariant with respect to the common belief and remain OBNE if the agents' beliefs are not necessarily derived from the same common belief. Of course, by Corollary 2, those OBNE are robust to changes of the common belief(s) only if each agent's strategy ranks acceptable only matches which are acceptable according to the true ranking and the reported ranking over the acceptable matches is essentially truthful.

4.2 Application II: Realized Matchings

The previous application described properties of strategy profiles which constitute an OBNE in a stable mechanism. For real-life environments we are also interested in which outcomes will be observed. Or in other words, for a given OBNE which matchings are realized ex-post, *i.e.* after each realization of a profile and its submitted rankings. Since we consider stable mechanisms, any realized matching is stable for the submitted profile. It turns out that all agents unanimously agree that the realized matching is truthfully most preferred among all matchings which are stable for the submitted profile.

Corollary 3 (Ex-Post Unanimity) *Let $\tilde{P}$ be a common belief, s be a strategy profile, and φ be a stable mechanism. Then, s is an OBNE in the stable mechanism φ under $\tilde{P}$ only if for all profiles P belonging to the support of $\tilde{P}$, all $\mu \in C(s(P))$ and all $v \in V$, $\varphi[s(P)](v) R_v \mu(v)$ (if $v \in W$) and $\varphi[s(P)](v) R_v^* \mu(v)$ for all responsive extensions P_v^* of P_v (if $v \in F$).*

Proof. Let $P \in \mathcal{P}$ be such that $\Pr\{\tilde{P} = P\} > 0$. Without loss of generality, let $v \in F$ (the proof for $v \in W$ is analogous and easier). Suppose that for some $\mu \in C(s(P))$ we have $\mu(v) P_v^* \varphi[s(P)](v)$ for some $P_v^* \in resp(P_v)$. Since the number of filled positions is identical for all firms for any two stable matchings (property (P3) of the core and stable matchings), we have $|\mu(v)| = |\varphi[s(P)](v)|$. Then $\mu(v)\backslash\varphi[s(P)](v) \neq \emptyset$ and by Theorem 4 of Roth and Sotomayor (1989), for all $w \in \mu(v)$ and all $w' \in \varphi[s(P)](v)\backslash\mu(v)$, $wP_v w'$. Let $P'_v \in \mathcal{P}_v$ be such that $A(P'_v) = \mu(v)$. Then it is easy to check that $\mu \in C(s(P))$ implies $\mu \in C(P'_v, s_{-v}(P_{-v}))$. By stability of φ and $A(P'_v) = \mu(v)$, $\varphi[P'_v, s_{-v}(P_{-v})](v) = \mu(v)$. Since $\mu(v)P^*_v\varphi[s(P)](v)$, $s(P)$ is not a NE in φ for P and by Theorem 1, s is not an OBNE in φ under $\tilde{P}$, a contradiction. □

Ehlers and Massó (2004, Theorem 2) showed Corollary 3 for one-to-one matching markets. Note that they could not rely on our general result Theorem 1 which allows the use of simple arguments to show that whenever the agents do not unanimously agree that the realized matching is most preferred in the core of the reported profile, then the agents do not play a NE at this profile.

In the above corollary the core of the submitted profile and the realized matching were related in terms of the true profile. Below we give for one-to-one matching markets a necessary and sufficient condition for all realized matchings to belong to the core of the true profile. Then all realized matchings are ex-post stable, *i.e.* for any profile in the support of the common belief, the matching chosen for the submitted rankings is stable for the true profile.

A profile $P' \in \mathcal{P}$ is a *strong Nash equilibrium (SNE)* under complete information P in the direct preference revelation game induced by the mechanism φ if for all coalitions $S \subseteq V$ there exists no $P''_S \in \mathcal{P}_S$ such that (i) for all $w \in S \cap W$, $\varphi[P''_S, P'_{-S}](w)P_w\varphi[P'](w)$, and (ii) for all $f \in S \cap F$, $\varphi[P''_S, P'_{-S}](f)P^*_f\varphi[P'](f)$ for some $P^*_f \in resp(P_f)$.

Corollary 4 (Ex-Post Stability for Marriage Markets) *Let $q_f = 1$ for all $f \in F$, $\tilde{P}$ be a common belief, s be a strategy profile, and φ be a stable mechanism. Let s be an OBNE in φ under $\tilde{P}$. Then, s is ex-post stable (for all P in the support of $\tilde{P}$, $\varphi[s(P)] \in C(P)$) if and only if for all P in the support of $\tilde{P}$, $s(P)$ is a SNE under complete information P.*

Proof. Follows directly from Theorem 1 and the fact that for any profile P and any NE $s(P)$ in φ under complete information P, $\varphi[s(P)] \in C(P)$ if and only if $s(P)$ is a SNE in φ under complete information P (Shin and Suh, 1996; Sönmez, 1997). □

For college admissions problems, under complete information it is known that the outcome of a SNE might not be stable under the true profile (Ma, 2002, Example 1). Therefore, in general the requirement that for any profile in the support the agents play a SNE is not sufficient for an OBNE to be ex-post stable. However, since the set of stable matchings coincides with the core, in college admissions problems this condition remains necessary for OBNE to be ex-post stable.

4.3 Application III: Truth-Telling

When agents' preferences are private information, we would like to design a mechanism which elicits the true preferences from the agents. In order to guarantee that agents truthfully report their preferences, incentive-compatible mechanisms make it a (weakly) dominant strategy to report truthfully. Incentive-compatibility is equivalent to the requirement that for any profile truth-telling is a NE under complete information. Therefore, incentive-compatibility is equivalent to truth-telling being an OBNE for all common beliefs.

Since incentive-compatibility is a strong condition, our incomplete information environment allows a weaker (but still natural) condition. Given a common belief and a mechanism, Bayesian incentive-compatibility requires that all agents truthfully reveal their preferences at any profile belonging to the support of the common belief. By our powerful result Theorem 1, in many-to-one matching markets for stable mechanisms Bayesian incentive-compatibility is equivalent to the requirement that truth-telling is a NE under complete information for any profile belonging to the support of the common belief. Now it follows directly from Corollary 3 that truth-telling is an OBNE only if the core is singleton at any realized profile.

Corollary 5 *Let $\tilde{P}$ be a common belief. Then, truth-telling is an OBNE in a stable mechanism under incomplete information $\tilde{P}$ only if the support of $\tilde{P}$ is contained in the set of all profiles with a singleton core.*

Since in college admissions problems incentive compatibility is equivalent to Bayesian incentive-compatibility for a common belief with full support, Roth's (1982) result (there exists no mechanism which is both stable and incentive-compatible) follows from Corollary 5 because there exist profiles with non-singleton core.

By Theorem 1, singleton cores would be sufficient for truth-telling to be an OBNE if at any profile belonging to the support of the common belief, truth-telling is a NE under complete information. By Roth (1985a) we know that this is not the case since he provides an example with singleton core where truth-telling is not a NE under complete information. Specifically, in his example a firm with more than one position profitably manipulates.

If each firm has exactly one position, then Ehlers and Massó (2004) show that singleton core is sufficient for truth-telling to be a NE in any stable mechanism under complete information. Therefore, we obtain the principal result of Ehlers and Massó (2004) as a corollary from Theorem 1.

Corollary 6 *[Theorem 1 in Ehlers and Massó (2004)] Let $q_f = 1$ for all $f \in F$ and $\tilde{P}$ be a common belief. Then, truth-telling is an OBNE in a stable mechanism under incomplete information $\tilde{P}$ if and only if the support of $\tilde{P}$ is contained in the set of all profiles with singleton core.*

5 Final Remarks

In many-to-one matching markets Theorem 1 provides for stable mechanisms a strong link between OBNE under incomplete information and NE under complete information. The following peculiarities of college admissions problems are important for the main result: (p1) any firm fills the same number of positions under any two stable matchings; and (p2) starting from any college admissions problem and its workers-optimal matching, when new workers become available all firms weakly prefer any matching, which is stable for the enlarged problem, to the workers-optimal matching of the original problem.[19]

It is clear that the link in Theorem 1 is in general not true for BNE. For instance, in the two-player game of matching pennies we may interpret each agent's pure strategies (heads and tails) as his possible types. Now if the common belief comes from two independent marginal beliefs that put probability $\frac{1}{2}$ on each type, and hence, the common belief puts probability $\frac{1}{4}$ on each strategy profile (as in the unique NE in mixed strategies), then truth-telling is a BNE under this common belief whereas the game does not have any NE (in pure strategies) under complete information.

It would be interesting to identify other economic environments where a similar link between BNE under incomplete information and NE under complete information holds. In those environments the strategic analysis under complete information is essential to undertake the corresponding analysis under incomplete information. For determining whether a strategy profile is an equilibrium under incomplete information, we only need to check whether for each realization the submitted preference orderings are a Nash equilibrium under complete information. Furthermore, if this link holds, then any BNE is robust to perturbations of the common belief which do not change the support of the common belief and agents may have private beliefs with equal support.

References

[1] A. Abdulkadiroğlu, P. Pathak, and A. Roth. "The New York City High School Match," *American Economic Review, Papers and Proceedings* **95**, 364-367 (2005).

[19] These two properties will be used frequently in the Appendix for the proof of Theorem 1.

[2] A. Abdulkadiroğlu and T. Sönmez. "School Choice: A Mechanism Design Approach," *American Economic Review* **93**, 729-747 (2003).

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

[4] J. Alcalde (1996). "Implementation of Stable Solutions to Marriage Problems," *Journal of Economic Theory* **69**, 240-254.

[5] Y. Chen and T. Sönmez. "School Choice: An Experimental Study," *Journal of Economic Theory* **127**, 202-231 (2006).

[6] V.P. Crawford. "Comparative Statics in Matching Markets," *Journal of Economic Theory* **54**, 389-400 (1991).

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

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

[9] L. Ehlers and J. Massó. "Incomplete Information and Singleton Cores in Matching Markets," *Journal of Economic Theory*, forthcoming (2004).

[10] H. Ergin and T. Sönmez. "Games of School Choice under the Boston Mechanism," *Journal of Public Economics* **90**, 215-237 (2006).

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

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

[13] O. Kesten. "An Advice to the Organizers of Entry-Level Labor Markets in the United Kingdom," mimeo, University of Rochester (2004).

[14] J. Ma. “Stable Matchings and Rematching-Proof Equilibria in a Two-Sided Matching Market,” *Journal of Economic Theory* **66**, 352–369 (1995).

[15] J. Ma. “Stable Matchings and the Small Core in Nash Equilibrium in the College Admissions Problem,” *Review of Economic Design* **7**, 117-134 (2002).

[16] D. Majumdar. “Ordinally Bayesian Incentive-Compatible Stable Matchings,” mimeo, CORE (2003).

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

[18] 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).

[19] J. Pais. *Incentives in Random Matching Markets.* Ph.D. Thesis, Universitat Autònoma de Barcelona (2005).

[20] A. Romero-Medina. “Implementation of Stable Solutions in a Restricted Matching Market,” *Review of Economic Design* **3**, 137-147 (1998).

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

[22] 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).

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

[24] A.E. Roth. “The College Admissions Problem is not Equivalent to the Marriage Problem,” *Journal of Economic Theory* **36**, 277-288 (1985a).

[25] A.E. Roth. "Common and Conflicting Interests in Two-Sided Matching Markets," *European Economic Review* **27**, 75-96 (1985b).

[26] A.E. Roth. "A Natural Experiment in the Organization of Entry Level Labor Markets: Regional Markets for New Physicians and Surgeons in the U.K.," *American Economic Review* **81**, 415-440 (1991).

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

[28] 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).

[29] A.E. Roth and M. Sotomayor. "The College Admissions Problem Revisited," *Econometrica* **57**, 559-570 (1989).

[30] 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).

[31] S. Shin and S.-C. Suh. "A Mechanism Implementing the Stable Rule in Marriage Problems," *Economics Letters* **51**, 185-189 (1996).

[32] T. Sönmez. "Games of Manipulation in Marriage Problems," *Games and Economic Behaviour* **20**, 169-176 (1997).

[33] U. Ünver. "On the Survival of Some Unstable Two-Sided Matching Mechanisms," *International Journal of Game Theory* **33**, 239-254 (2005).

APPENDIX

Before we prove Theorem 1, we recall the following properties of the core of a college admissions problem. These properties will be used frequently in the proof. It will be convenient to write $(F, W, P; q)$ for any college admissions problem (F, W, q, P) in which $q_f = 1$ for all $f \in F$.

A.1 Properties of the Core

(I) For each $P \in \mathcal{P}$, the set of unmatched agents is the same for all stable matchings (see Roth and Sotomayor, 1990, Theorems 5.12 and 5.13); namely, for all $\mu, \mu' \in C(P)$, and for all $w \in W$ and $f \in F$, (i) if $\mu(w) = \emptyset$, then $\mu'(w) = \emptyset$; (ii) $|\mu(f)| = |\mu'(f)|$; and (iii) if $|\mu(f)| < q_f$, then $\mu(f) = \mu'(f)$.

(II) Given (F, W, q, P), split each firm f into q_f identical copies of itself (all having the same preference ordering P_f) and let F' be this new set of $\sum_{f \in F} q_f$ splitted firms. Set $q'_{f'} = 1$ for all $f' \in F'$ and replace f by its copies in F' (always in the same order) in each worker's preference relation P_w. Then, $(F', W, P; q')$ is a marriage market for which we can uniquely identify its matchings with the matchings of the original college admissions problem (F, W, q, P), and vice versa (Roth and Sotomayor, 1990, Lemma 5.6). Then, and using this identification, we write $C(F, W, q, P) = C(F', W, P; q')$.

(III) Consider a marriage market $(F, W, P; q)$ and suppose that new workers enter the market. Let $(F, W', P'; q)$ be this new marriage market where $W \subseteq W'$ and P' agrees with P over F and W. Let $DA_W[P] = \mu_W$. Then, for all $f \in F$, $\mu'(f) R'_f \mu_W(f)$ for all $\mu' \in C(F, W', P'; q)$ (Gale and Sotomayor, 1985; Crawford, 1991).

A.2 Proof of Theorem 1

Theorem 1 *Let $\tilde{P}$ be a common belief, s be a strategy profile, and φ be a stable mechanism. Then, s is an OBNE in the stable mechanism φ under incomplete information $\tilde{P}$ 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 $\tilde{P}$ be a common belief, s be a strategy profile and φ be a stable mechanism.

($\Leftarrow$) Suppose that 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. Let $v \in V$ and $P_v \in \mathcal{P}_v$ be such that $\Pr\{\tilde{P}_v = P_v\} > 0$. By the previous fact, then we have for all $P'_v \in \mathcal{P}_v$ and all $P_{-v} \in \mathcal{P}_{-v}$ such that $\Pr\{\tilde{P}_{-v}|_{P_v} = P_{-v}\} > 0$, $\varphi[s(P)](v) R^*_v \varphi[P'_v, s_{-v}(P_{-v})](v)$ for all $P^*_v \in resp(P_v)$ (if $v \in F$) and $\varphi[s(P)](v) R_v \varphi[P'_v, s_{-v}(P_{-v})](v)$ (if $v \in W$). Hence,

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

and s is an OBNE in φ under $\tilde{P}$, the desired conclusion.

($\Rightarrow$) Let s be an OBNE in the stable mechanism φ under incomplete information $\tilde{P}$. First we show that for all $P \in \mathcal{P}$ such that $\Pr\{\tilde{P} = P\} > 0$,

$$\varphi[s(P)](v) \subseteq A(P_v) \text{ for all } v \in V. \tag{2}$$

If for some P in the support of $\tilde{P}$ and for some $v \in V$, $\varphi[s(P)](v) \nsubseteq A(P_v)$, then choose $P'_v \in \mathcal{P}_v$ such that $A(P'_v) = A(P_v) \cap A(s_v(P_v))$ and $P'_v|A(P'_v) = s_v(P_v)|A(P'_v)$. By the stability of φ and our choice of P'_v, we have $\varphi[P'_v, s_{-v}(P'_{-v})](v) \subseteq A(P_v)$ for all $P'_{-v} \in \mathcal{P}_{-v}$. Let $v \in F$ (the case $v \in W$ is analogous and easier). We choose a responsive extension P^*_v of P_v such that for all $W' \in 2^W$, $W' R^*_v \emptyset$ if and only if $W' \subseteq A(P_v)$. Hence, by $\varphi[P'_v, s_{-v}(P_{-v})](v) \subseteq A(P_v)$ and $\varphi[s(P)](v) \nsubseteq A(P_v)$, $\varphi[P'_v, s_{-v}(P_{-v})](v) R^*_v \emptyset P^*_v \varphi[s(P)](v)$. Since $\Pr\{\tilde{P}_{-v}|_{P_v} = P_{-v}\} > 0$, it follows that

$$\Pr\{\varphi[P'_v, s_{-v}(\tilde{P}_{-v}|_{P_v})](v) \in B(\emptyset, P^*_v)\} = 1 > \Pr\{\varphi[s_v(P_v), s_{-v}(\tilde{P}_{-v}|_{P_v})](v) \in B(\emptyset, P^*_v)\},$$

which means that s is not an OBNE in the stable mechanism φ under $\tilde{P}$, a contradiction. Hence, (2) holds.

Second suppose that there is some $P \in \mathcal{P}$ such that $\Pr\{\tilde{P} = P\} > 0$ and $s(P)$ is not a Nash equilibrium under complete information P in the direct preference revelation game induced by φ. Then (w.l.o.g.) there exist $f \in F$, $P'_f \in \mathcal{P}_f$, and a responsive extension P^*_f of P_f such that

$$\varphi[P'_f, s_{-f}(P_{-f})](f) P^*_f \varphi[s(P)](f). \tag{3}$$

The case where a worker has a profitable deviation is analogous to the case where a firm with quota one has a profitable deviation.

Let $\varphi[P'_f, s_{-f}(P_{-f})] = \mu'$ and $\varphi[s(P)] = \mu$. Furthermore, let $\mu'(f) = \{w'_1, w'_2, \ldots, w'_{|\mu'(f)|}\}$ where $w'_1 P_f w'_2 P_f \cdots P_f w'_{|\mu'(f)|}$ and $\mu(f) = \{w_1, w_2, \ldots, w_{|\mu(f)|}\}$ where $w_1 P_f w_2 P_f \cdots P_f w_{|\mu(f)|}$.

Case 1: There exists $k \in \{1, \ldots, |\mu'(f)|\}$ such that $w'_k P_f w_k$ and $w_l R_f w'_l$ for all $l \in \{1, \ldots, k-1\}$.

Note that $w'_k \in A(P_f)$ because $w'_k P_f w_k$ and by (2), $w_k \in \mu(f) \subseteq A(P_f)$. Let $P''_f \in \mathcal{P}_f$ be such that $A(P''_f) = B(w'_k, P_f)$ and $P''_f | A(P''_f) = P'_f | A(P''_f)$.

First we show that $\varphi[P''_f, s_{-f}(P_{-f})](f)$ contains at least k workers. Note that any profile implicitly specifies the set of agents of the matching problem. For the time being, below we specify both the profile and the quota of the matching problem.

Because φ is stable and $\varphi[P'_f, s_{-f}(P_{-f})] = \mu'$, we have $\mu' \in C(P'_f, s_{-f}(P_{-f}); q)$. Let μ'' be the matching for the problem $(F, W \backslash \{w'_{k+1}, \ldots, w'_{|\mu'(f)|}\}, (k, q_{-f}), (P'_f, s_{-\{f\} \cup \{w'_{k+1}, \ldots, w'_{|\mu'(f)|}\}}(P_{-\{f\} \cup \{w'_{k+1}, \ldots, w'_{|\mu'(f)|}\}})))$ such that $\mu''(f) = \{w'_1, \ldots, w'_k\}$ and $\mu''(f') = \mu(f')$ for all $f' \in F \backslash \{f\}$. Then from $\mu' \in C(P'_f, s_{-f}(P_{-f}); q)$ it follows that

$$\mu'' \in C(P'_f, s_{-\{f\} \cup \{w'_{k+1}, \ldots, w'_{|\mu'(f)|}\}}(P_{-\{f\} \cup \{w'_{k+1}, \ldots, w'_{|\mu'(f)|}\}}); k, q_{-f}). \tag{4}$$

By our choice of P''_f, we have $\mu''(f) \subseteq A(P''_f)$ and $P''_f | A(P''_f) = P'_f | A(P''_f)$. Hence, we also have by (4),

$$\mu'' \in C(P''_f, s_{-\{f\} \cup \{w'_{k+1}, \ldots, w'_{|\mu'(f)|}\}}(P_{-\{f\} \cup \{w'_{k+1}, \ldots, w'_{|\mu'(f)|}\}}); k, q_{-f}). \tag{5}$$

Thus, by $\mu''(f) = \{w'_1, \ldots, w'_k\}$ and the fact that any firm is matched to the same number of workers under all stable matchings, firm f is matched to k workers for all matchings belonging to $C(P''_f, s_{-\{f\} \cup \{w'_{k+1}, \ldots, w'_{|\mu'(f)|}\}}(P_{-\{f\} \cup \{w'_{k+1}, \ldots, w'_{|\mu'(f)|}\}}); k, q_{-f})$.

Now if firm f is matched to fewer than k workers in some matching belonging to $C(P''_f, s_{-\{f\}\cup\{w'_{k+1},\ldots,w'_{|\mu'(f)|}\}}(P_{-\{f\}\cup\{w'_{k+1},\ldots,w'_{|\mu'(f)|}\}}); q)$, then this matching is also stable for the problem $(P''_f, s_{-\{f\}\cup\{w'_{k+1},\ldots,w'_{|\mu'(f)|}\}}(P_{-\{f\}\cup\{w'_{k+1},\ldots,w'_{|\mu'(f)|}\}}); k, q_{-f})$, a contradiction to the previous fact. Hence, f is matched to at least k workers in any stable matching belonging to $C(P''_f, s_{-\{f\}\cup\{w'_{k+1},\ldots,w'_{|\mu'(f)|}\}}(P_{-\{f\}\cup\{w'_{k+1},\ldots,w'_{|\mu'(f)|}\}}); q)$. Now when considering the worker optimal matching in this core, we may split firm f into q_f copies (all having the same preference P''_f) and each copy of firm f weakly prefers according to P''_f any matching in $C(P''_f, P_{-f}; q)$ to this matching. Since at least k copies of f are matched to a worker under the worker optimal matching in $C(P''_f, s_{-\{f\}\cup\{w'_{k+1},\ldots,w'_{|\mu'(f)|}\}}(P_{-\{f\}\cup\{w'_{k+1},\ldots,w'_{|\mu'(f)|}\}}); q)$, at least k copies of f must be also matched to a worker under any stable matching in $C(P''_f, s_{-f}(P_{-f}); q)$. Therefore, by $\varphi[P''_f, s_{-f}(P_{-f})] \in C(P''_f, s_{-f}(P_{-f}); q)$, $\varphi[P''_f, s_{-f}(P_{-f})](f)$ contains at least k workers.

Second we choose a responsive extension P^{**}_f of P_f. Let $W^* \subseteq B(w'_k, P_f)$ be such that W^* consists of the k lowest ranked workers (according to P_f) in the set $B(w'_k, P_f)$, *i.e.* $|W^*| = k$ and for all $w \in B(w'_k, P_f)\backslash W^*$ and all $w^* \in W^*$, wP_fw^*. Let P^{**}_f be the responsive extension of P_f be such that for all $W'' \in 2^W$, $W''P^{**}_fW^*$ if and only if the following three conditions hold: (i) $W'' \subseteq A(P_f)$, (ii) $|W''| \geq k$, and (iii) if $W'' = \{w''_1, w''_2, \ldots, w''_{|W''|}\}$ where $w''_1P_f \cdots P_fw''_{|W''|}$ and $W^* = \{w^*_1, \ldots, w^*_k\}$ where $w^*_1P_f \cdots P_fw^*_k$, then $w''_lR_fw^*_l$ for all $l \in \{1, \ldots, k\}$. Since $\varphi[P''_f, s_{-f}(P_{-f})](f)$ contains at least k workers and $A(P''_f) = B(w'_k, P_f)$, our construction implies that $\varphi[P''_f, s_{-f}(P_{-f})](f)P^{**}_f\varphi[s(P)](f)$. More precisely, for Case 1 the set $\varphi[s(P)](f)$ violates (iii) and our choice of P^{**}_f and W^* yields

$$\varphi[P''_f, s_{-f}(P_{-f})](f)R^{**}_fW^*P^{**}_f\varphi[s(P)](f). \tag{6}$$

Third we show that for all (P_f, P'_{-f}) in the support of $\tilde{P}$, if $\varphi[s_f(P_f), s_{-f}(P'_{-f})](f) \in B(W^*, P^{**}_f)$, then $\varphi[P''_f, s_{-f}(P'_{-f})](f) \in B(W^*, P^{**}_f)$. This then completes the proof for Case 1 because by $\Pr\{\tilde{P}_{-f}|_{P_f} = P_{-f}\} > 0$, and (6), it follows that

$$\Pr\{\varphi[P''_f, s_{-f}(\tilde{P}_{-f}|_{P_f})](f) \in B(W^*, P^{**}_f)\} > \Pr\{\varphi[s_f(P_f), s_{-f}(\tilde{P}_{-f}|_{P_f})](f) \in B(W^*, P^{**}_f)\},$$

which means that s is not an OBNE in φ under $\tilde{P}$.

Suppose that $\varphi[s_f(P_f), s_{-f}(P'_{-f})](f) R^{**}_f W^*$. By our choice of P^{**}_f, then

$$\varphi[s_f(P_f), s_{-f}(P'_{-f})](f) \cap B(w'_k, P_f) \text{ must contain at least } k \text{ workers.} \tag{7}$$

If $\varphi[P''_f, s_{-f}(P'_{-f})](f)$ contains at least k workers, then all these workers belong to $B(w'_k, P_f)$. Thus, by our choice of P^{**}_f and W^*, $\varphi[P''_f, s_{-f}(P'_{-f})](f) R^{**}_f W^*$, the desired conclusion.

Suppose that $\varphi[P''_f, s_{-f}(P'_{-f})](f)$ contains fewer than k workers. Let $\hat{\mu} = \varphi[s_f(P_f), s_{-f}(P'_{-f})]$. Let $\hat{\mu}(f) = \{\hat{w}_1, \ldots, \hat{w}_{|\hat{\mu}(f)|}\}$ where $\hat{w}_1 P_f \cdots P_f \hat{w}_{|\hat{\mu}(f)|}$. By (7), $\hat{\mu}(f) \cap B(w'_k, P_f)$ contains at least k workers. Thus, $k \leq |\hat{\mu}(f)|$. For the time being, below we specify both the profile and the quota of the matching problem. Then we have $\hat{\mu} \in C(s_f(P_f), s_{-f}(P'_{-f}); q)$. Let $\hat{\mu}'$ be the matching for the problem $(F, W \backslash \{\hat{w}_{k+1}, \ldots, \hat{w}_{|\hat{\mu}(f)|}\}, (k, q_{-f}), (s_f(P_f), s_{-\{f\} \cup \{\hat{w}_{k+1}, \ldots, \hat{w}_{|\hat{\mu}(f)|}\}}(P'_{-\{f\} \cup \{\hat{w}_{k+1}, \ldots, \hat{w}_{|\hat{\mu}(f)|}\}}))$ such that $\hat{\mu}'(f) = \{\hat{w}_1, \ldots, \hat{w}_k\}$ and $\hat{\mu}'(f') = \hat{\mu}(f')$ for all $f' \in F \backslash \{f\}$. Then, from $\hat{\mu} \in C(s_f(P_f), s_{-f}(P'_{-f}); q)$ it follows that

$$\hat{\mu}' \in C(s_f(P_f), s_{-\{f\} \cup \{\hat{w}_{k+1}, \ldots, \hat{w}_{|\hat{\mu}(f)|}\}}(P'_{-\{f\} \cup \{\hat{w}_{k+1}, \ldots, \hat{w}_{|\hat{\mu}(f)|}\}}); k, q_{-f}). \tag{8}$$

Let $\hat{w} \in \hat{\mu}'(f)$ be such that $\hat{\mu}'(f) \subseteq B(\hat{w}, s_f(P_f))$ (in other words, $\hat{w}$ is the worker who is least preferred in $\hat{\mu}'(f)$ according to $s_f(P_f)$). Let $\hat{P}_f \in \mathcal{P}_f$ be such that $A(\hat{P}_f) = B(\hat{w}_k, P_f) \cap B(\hat{w}, s_f(P_f))$ and $\hat{P}_f | A(\hat{P}_f) = P''_f | A(\hat{P}_f)$. Then we must have $\hat{\mu}' \in C(\hat{P}_f, s_{-\{f\} \cup \{\hat{w}_{k+1}, \ldots, \hat{w}_{|\hat{\mu}(f)|}\}}(P'_{-\{f\} \cup \{\hat{w}_{k+1}, \ldots, \hat{w}_{|\hat{\mu}(f)|}\}}); k, q_{-f})$ (otherwise there would exist a blocking pair for $\hat{\mu}'$;[20] then by (8) and the fact that only firm f's preference changed from $s_f(P_f)$ to $\hat{P}_f$, firm f needs to be part of this blocking pair; thus, (w, f) blocks $\hat{\mu}'$ which implies $w \notin \hat{\mu}'(f)$ and $w \neq \hat{w}$, and $w \in A(\hat{P}_f) = B(\hat{w}_k, P_f) \cap B(\hat{w}, s_f(P_f))$; therefore, $w \in B(\hat{w}, s_f(P_f)) \backslash \hat{\mu}'(f)$ and (w, f) must also block $\hat{\mu}'$ under $(s_f(P_f), s_{-\{f\} \cup \{\hat{w}_{k+1}, \ldots, \hat{w}_{|\hat{\mu}(f)|}\}}(P'_{-\{f\} \cup \{\hat{w}_{k+1}, \ldots, \hat{w}_{|\hat{\mu}(f)|}\}}); k, q_{-f})$, a contradiction to (8).)

Thus, since $|\hat{\mu}'(f)| = k$, firm f is matched to k workers for all matchings belonging to $C(\hat{P}_f, s_{-\{f\} \cup \{\hat{w}_{k+1}, \ldots, \hat{w}_{|\hat{\mu}(f)|}\}}(P'_{-\{f\} \cup \{\hat{w}_{k+1}, \ldots, \hat{w}_{|\hat{\mu}(f)|}\}}); k, q_{-f})$. Now if firm f is matched

[20]Note that $\hat{\mu}'$ is individually rational because both $\hat{\mu}'(f) \subseteq B(\hat{w}_k, P_f)$ and $\hat{\mu}'(f) \subseteq B(\hat{w}, s_f(P_f))$ (by our choice of $\hat{w}$).

to fewer than k workers for some $\tilde{\mu} \in C(\hat{P}_f, s_{-\{f\}\cup\{\hat{w}_{k+1},\ldots,\hat{w}_{|\hat{\mu}(f)|}\}}(P'_{-\{f\}\cup\{\hat{w}_{k+1},\ldots,\hat{w}_{|\hat{\mu}(f)|}\}}); q)$, then $\tilde{\mu}$ is also stable under $(\hat{P}_f, s_{-\{f\}\cup\{\hat{w}_{k+1},\ldots,\hat{w}_{|\hat{\mu}(f)|}\}}(P'_{-\{f\}\cup\{\hat{w}_{k+1},\ldots,\hat{w}_{|\hat{\mu}(f)|}\}}); k, q_{-f})$, a contradiction to the previous fact. Hence, f is matched to at least k workers in any stable matching belonging to $C(\hat{P}_f, s_{-\{f\}\cup\{\hat{w}_{k+1},\ldots,\hat{w}_{|\hat{\mu}(f)|}\}}(P'_{-\{f\}\cup\{\hat{w}_{k+1},\ldots,\hat{w}_{|\hat{\mu}(f)|}\}}); q)$. Now when considering the worker optimal matching in this core, we may split firm f into q_f copies (all having the same preference $\hat{P}_f$) and each copy of firm f weakly prefers according to $\hat{P}_f$ any matching in $C(\hat{P}_f, s_{-f}(P'_{-f}); q)$ to this matching. Since at least k copies of f are matched to a worker under the worker optimal matching in $C(\hat{P}_f, s_{-\{f\}\cup\{\hat{w}_{k+1},\ldots,\hat{w}_{|\hat{\mu}(f)|}\}}(P'_{-\{f\}\cup\{\hat{w}_{k+1},\ldots,\hat{w}_{|\hat{\mu}(f)|}\}}); q)$,

$$\text{at least } k \text{ copies of } f \text{ are matched to a worker in any matching in } C(\hat{P}_f, s_{-f}(P'_{-f}); q). \tag{9}$$

On the other hand, $\varphi[P''_f, s_{-f}(P'_{-f})](f)$ contains fewer than k workers. Let $\tilde{\mu} = \varphi[P''_f, s_{-f}(P'_{-f})]$. Let $\tilde{\mu}'$ be the matching for the problem $(F, W\backslash(\tilde{\mu}(f)\backslash A(\hat{P}_f)), q, (P''_f, s_{-\{f\}\cup(\tilde{\mu}(f)\backslash A(\hat{P}_f))}(P'_{-\{f\}\cup(\tilde{\mu}(f)\backslash A(\hat{P}_f))})))$ such that $\tilde{\mu}'(f) = \tilde{\mu}(f)\cap A(\hat{P}_f)$ and $\tilde{\mu}'(f') = \tilde{\mu}(f')$ for all $f' \in F\backslash\{f\}$. Since $\tilde{\mu} \in C(P''_f, s_{-f}(P'_{-f}), q)$ and $\tilde{\mu}(f)$ contains fewer than q_f workers, we must have $\tilde{\mu}' \in C(P''_f, s_{-\{f\}\cup(\tilde{\mu}(f)\backslash A(\hat{P}_f))}(P'_{-\{f\}\cup(\tilde{\mu}(f)\backslash A(\hat{P}_f))}); q)$. Thus, by $\tilde{\mu}'(f) \subseteq A(\hat{P}_f)$ and $\hat{P}_f|A(\hat{P}_f) = P''_f|A(\hat{P}_f)$, we also obtain $\tilde{\mu}' \in C(\hat{P}_f, s_{-\{f\}\cup(\tilde{\mu}(f)\backslash A(\hat{P}_f))}(P'_{-\{f\}\cup(\tilde{\mu}(f)\backslash A(\hat{P}_f))}); q)$. Hence, in any matching belonging to this core firm f is matched to $|\tilde{\mu}'(f)| = |\tilde{\mu}(f)\cap A(\hat{P}_f)|$ workers. Now when considering the worker optimal matching in this core, we may split each firm $f' \in F\backslash\{f\}$ into $q_{f'}$ copies (all having the same preference $s_{f'}(P'_{f'})$) and each copy of firm f' weakly prefers according to $s_{f'}(P'_{f'})$ any matching in $C(\hat{P}_f, s_{-f}(P'_{-f}); q)$ to this matching. Thus, in total all the copies of all firms $f' \in F\backslash\{f\}$ receive at least the same number of workers in $C(\hat{P}_f, s_{-f}(P'_{-f}); q)$ as they did previously. Since exactly $|\tilde{\mu}(f)\backslash A(\hat{P}_f)|$ new workers are available and f was matched to $|\tilde{\mu}'(f)| = |\tilde{\mu}(f) \cap A(\hat{P}_f)|$ workers before, firm f can be matched to at most $|\tilde{\mu}(f)|$ workers under any stable matching in $C(\hat{P}_f, s_{-f}(P'_{-f}); q)$. Since $|\tilde{\mu}(f)|$ is smaller than k, this contradicts (9) and the fact that under responsive preferences, firm f is matched to the same number of workers for any two matchings in $C(\hat{P}_f, s_{-f}(P'_{-f}); q)$. Hence, $\varphi[P''_f, s_{-f}(P'_{-f})](f)$ cannot contain fewer than k workers.

Case 2: Otherwise.

Then we have $w_l R_f w'_l$ for all $l \in \{1, \ldots, \min\{|\mu(f)|, |\mu'(f)|\}\}$. Let $k = |\mu(f)|$. If $|\mu'(f)| \leq \mu(f)$, then by responsiveness of P^*_f and $\mu(f) \subseteq A(P_f)$, we have $\mu(f) R^*_f \mu'(f)$, which contradicts (3). Hence, we must have $|\mu'(f)| > |\mu(f)| = k$, $q_f > k$, and $w'_{k+1} \in A(P_f)$. Let $P''_f \in \mathcal{P}_f$ be such that $A(P''_f) = B(w'_{k+1}, P_f)$ and $P''_f|A(P''_f) = P'_f|A(P''_f)$. Since $\mu(f) \subseteq B(w'_{k+1}, P_f) = A(P''_f)$ and $\mu(f)$ does not fill the quota of firm f, we must have $\mu \in C(P''_f, s_{-f}(P_{-f}); q)$. Hence,

$$\text{firm } f \text{ is matched to } k \text{ workers under any matching in } C(P''_f, s_{-f}(P_{-f}); q). \qquad (10)$$

On the other hand, let μ'' be the matching for the problem $(F, W\backslash\{w'_{k+2}, \ldots, w'_{|\mu'(f)|}\}, (k+1, q_{-f}), (P''_f, s_{-\{f\}\cup\{w'_{k+2},\ldots,w'_{|\mu'(f)|}\}}(P_{-\{f\}\cup\{w'_{k+2},\ldots,w'_{|\mu'(f)|}\}})))$ such that $\mu''(f) = \{w'_1, \ldots, w'_{k+1}\}$ and $\mu''(f') = \mu'(f')$ for all $f' \in F\backslash\{f\}$. Then from $\mu' \in C(P'_f, s_{-f}(P_{-f}); q)$ it follows that $\mu'' \in C(P'_f, s_{-\{f\}\cup\{w'_{k+2},\ldots,w'_{|\mu'(f)|}\}}(P_{-\{f\}\cup\{w'_{k+2},\ldots,w'_{|\mu'(f)|}\}}); k+1, q_{-f})$. Thus, by $\mu''(f) \subseteq B(w'_{k+1}, P_f) = A(P''_f)$ and $P''_f|A(P''_f) = P'_f|A(P''_f)$, $\mu'' \in C(P''_f, s_{-\{f\}\cup\{w'_{k+2},\ldots,w'_{|\mu'(f)|}\}}(P_{-\{f\}\cup\{w'_{k+2},\ldots,w'_{|\mu'(f)|}\}}); k+1, q_{-f})$. Now if firm f is matched to fewer than $k+1$ workers in some matching belonging to $C(P''_f, s_{-\{f\}\cup\{w'_{k+2},\ldots,w'_{|\mu'(f)|}\}}(P_{-\{f\}\cup\{w'_{k+2},\ldots,w'_{|\mu'(f)|}\}}); q)$, then this matching is also stable for the problem $(P''_f, s_{-\{f\}\cup\{w'_{k+2},\ldots,w'_{|\mu'(f)|}\}}(P_{-\{f\}\cup\{w'_{k+2},\ldots,w'_{|\mu'(f)|}\}}); k+1, q_{-f})$, a contradiction to the previous fact. Hence, f is matched to at least $k+1$ workers in any stable matching belonging to $C(P''_f, s_{-\{f\}\cup\{w'_{k+2},\ldots,w'_{|\mu'(f)|}\}}(P_{-\{f\}\cup\{w'_{k+2},\ldots,w'_{|\mu'(f)|}\}}); q)$. Now when considering the worker optimal matching in this core, we may split firm f into $k+1$ copies (all having the same preference P''_f) and each copy of firm f weakly prefers according to P''_f any matching in $C(P''_f, s_{-f}(P_{-f}); q)$ to this matching. Since at least $k+1$ copies of f are matched to a worker under the worker optimal matching in $C(P''_f, s_{-\{f\}\cup\{w'_{k+2},\ldots,w'_{|\mu'(f)|}\}}(P_{-\{f\}\cup\{w'_{k+2},\ldots,w'_{|\mu'(f)|}\}}); q)$, at least $k+1$ copies of f must be also matched to a worker under any matching in $C(P''_f, s_{-f}(P_{-f}); q)$, which contradicts (10) and the fact that firm f is matched to the same number of workers under any matching in $C(P''_f, s_{-f}(P_{-f}); q)$. Hence, Case 2 cannot occur. $\square$