

Matching Markets under (In)complete Information*

Lars Ehlers[†]

Jordi Massó[‡]

November 2008

*We are especially grateful to three anonymous referees and the Managing Editor Andrea Prat for their detailed comments and suggestions. L. Ehlers acknowledges financial support from 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 GSE research network. 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 CREA, 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.massó@uab.es

Abstract

We introduce incomplete information to centralized many-to-one matching markets such as those of 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 link between Nash equilibria under complete information and ordinal Bayesian Nash equilibria under incomplete information. That is, given a common prior, a strategy profile is an ordinal Bayesian Nash equilibrium under incomplete information in a stable mechanism if and only if, for any true profile in the support of the common prior, 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 8th 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 rank list of potential partners and proposes, after processing the profile of submitted rank lists, a final matching between firms and workers.

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

²See Roth (1984a), Roth and Peranson (1999), and Roth (2002) for a careful description and analysis of this market. Roth (1991), Kesten (2004), Ünver (2005), and Ehlers (2008) describe and analyze the equivalent UK markets.

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

⁴Chen 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.

⁵Abdulkadiroğlu, 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) rank 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 prior (or 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 rank lists and not their specific utility representations, we

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

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

stick to the 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 (OBNE) if, for every von Neumann-Morgenstern utility function of an agent's preference ordering (his type), the submitted rank list maximizes his expected utility in the direct preference revelation game induced by the common prior 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 link between Nash equilibria under complete information and ordinal Bayesian Nash equilibria under incomplete information. More precisely, Theorem 1 states that, given a common prior, 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 prior, 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 orderings are a Nash equilibrium under complete information. This also implies that

⁸This 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 (2007), and Pais (2005) have already used this ordinal equilibrium notion in one-to-one matching markets.

truth-telling is a dominant strategy for all agents in the stable mechanism under incomplete information (given by a common prior) if and only if for any profile in the support of the common prior, truth-telling is a best reply in the stable mechanism under complete information if all agents truth-tell. In other words, a stable mechanism is Bayesian incentive compatible for a given common prior if and only if the stable mechanism restricted to the support of the common prior is incentive compatible. For matching markets and stable mechanisms, this is an important connection between ex-ante incentive compatibility and ex-post incentive compatibility.

Another corollary of Theorem 1 is that for any stable mechanism, the set of ordinal Bayesian Nash equilibria is identical for any two common priors with equal support. Therefore, any equilibrium is robust to perturbations of the common prior which do not change the support of the common prior and agents may have different priors with equal support. Note that in matching markets it may not be appropriate to consider arbitrary perturbations of the common prior because in empirical applications certain firms are unambiguously perceived better than others and any worker's true preference relation never reverses those perceptions. At the (realistic) extreme, for any realization of the common prior all workers rank all firms identically (according to some objective criterion). We show in this case that truth-telling is an OBNE in the stable mechanism under incomplete information and all realized matchings are stable with respect to the true profile. Furthermore, Theorem 1 permits an important extension of the result of Roth (1984b) from complete information to incomplete information. He shows for one-to-one matching markets under complete information that for the stable mechanism, called the workers-proposing deferred-acceptance algorithm, the outcome of any Nash equilibrium, where the workers truthfully reveal their preferences, is stable with respect to the true profile. We show for many-to-one matching markets under incomplete information, if the workers preferences are correlated, then the outcome of any realization of an OBNE in the workers-proposing deferred-acceptance algorithm, where the workers truthfully reveal their preferences,

is stable with respect to the true profile.

Another important consequence is that the set of ordinal Bayesian Nash equilibria for common priors with full support remain equilibria for any common prior. 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 prior. Finally, we will argue that as a consequence of our main result the description of incomplete information by a common belief can be weakened by using preference type spaces (like Bergemann and Morris (2005)) that admit less common knowledge about beliefs and beliefs about beliefs (or higher order beliefs).

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.

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 $v \in V \equiv F \cup W$. Both the set of workers W and the set of firms F are assumed to be finite. 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 to 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 $w \in \mu(f)$, $wP_f\emptyset$; and

(s3) there is no pair $(w, f) \in W \times F$ such that $w \notin \mu(f)$, $fP_w\mu(w)$, and either wP_fw' for some $w' \in \mu(f)$ or $wP_f\emptyset$ 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 (*direct revelation*) *mechanism* is a function $\varphi : \mathcal{P} \rightarrow \mathcal{M}$ mapping each preference profile $P \in \mathcal{P}$ to a matching $\varphi[P] \in \mathcal{M}$. 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 revelation mechanisms, 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 $P_f : w_1 w_2 w_3 w_4 \emptyset^9$ 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,¹⁰ 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

⁹We will use the convention that $P_f : w_1 w_2 w_3 w_4 \emptyset$ means $w_1 P_f w_2 P_f w_3 P_f w_4 P_f \emptyset$.

¹⁰See for instance, Roth and Sotomayor (1990).

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 $W \cup \{\emptyset\}$ 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) $(S \setminus \{w\}) \cup \{w'\} P_f^* S$ 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 under Responsive Extensions

The core of a many-to-one matching market where firms have responsive preferences has a special structure. The following well-known properties will be useful in the sequel:¹¹

(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 $DA_F : \mathcal{P} \rightarrow \mathcal{M}$ and $DA_W : \mathcal{P} \rightarrow \mathcal{M}$, 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 $DA_W [P] = \mu_W$.¹²

¹¹See Roth and Sotomayor (1990) for a detailed presentation of these properties.

¹²Strictly 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.

(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 blocking¹³. 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 $\mu(f) = \mu'(f)$.

(P4) Starting from the workers-optimal matching of any college admissions problem, once new workers become available any firm weakly prefers any stable matching of the enlarged market to the workers-optimal matching of the smaller market. More precisely, consider a college admissions problem (F, W, q, P) and suppose that new workers enter the market. Let (F, W', q, P') be this new market where $W \subseteq W'$ and P' agrees with P over F and W . Let $DA_W[P] = \mu_W$. Then, for all $f \in F$ and all responsive extensions $R_f'^*$ of R_f , $\mu'(f)R_f'^*\mu_W(f)$ for all $\mu' \in C(F, W', q, P')$ (Gale and Sotomayor, 1985; Crawford, 1991).

None of these four properties holds if firms have non-responsive preferences on 2^W . In particular, the set of stable matchings for an arbitrary profile may be empty, and if it is non-empty it may not coincide with the core. For many applications, however, responsiveness is a meaningful restriction since the goodness of a set of workers comes from the specific quality of each of its workers (*i.e.*, responsiveness excludes complementarities among workers). In addition, it allows to focus on real life centralized stable matching mechanisms which are much simpler in terms of the information

¹³A 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 $v \in S$.

agents have to submit—preference orderings on $F \cup \{\emptyset\}$ and $W \cup \{\emptyset\}$ —compared with mechanisms that would admit as inputs preference orderings on $F \cup \{\emptyset\}$ and 2^W , particularly when the number of workers is large. Nevertheless, our analysis could be similarly performed if stable mechanisms were using preference orderings on 2^W as inputs of the firms, as long as they were responsive to their induced rankings on $W \cup \{\emptyset\}$.

3 Incomplete Information

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

Definition 2 (Nash Equilibrium) Let P be a profile.

- (a) Truth-telling is a *Nash equilibrium (NE)* in the mechanism φ under complete information P 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 $P_f^* \in \text{resp}(P_f)$, $\varphi[P](f)R_f^*\varphi[\hat{P}_f, P_{-f}](f)$ for all $\hat{P}_f \in \mathcal{P}_f$.
- (b) A profile P' is a *Nash equilibrium (NE)* in the mechanism φ under complete information P 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 $P_f^* \in \text{resp}(P_f)$, $\varphi[P'](f)R_f^*\varphi[\hat{P}_f, P'_{-f}](f)$ 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 prior about the true profile, which is shared among all agents. A *common prior* is a probability distribution \tilde{P} over \mathcal{P} . Given a profile P and the common prior \tilde{P} , $\Pr\{\tilde{P} = P\}$ is the probability that \tilde{P} assigns to

the event that the true profile is P .¹⁴ 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 prior \tilde{P} describes agents' uncertainty about the true profile before agents learn their types. Now, given a common prior \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 .¹⁵ This formulation does not require symmetry nor independence of priors; conditional priors 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 $\tilde{\eta}$ is a probability distribution over the set of matchings \mathcal{M} . Given a matching μ and the random matching $\tilde{\eta}$, $\Pr\{\tilde{\eta} = \mu\}$ is the probability that $\tilde{\eta}$ assigns to matching μ . But the uncertainty important for agent v is not over matchings but over v 's set of potential partners. Let $\tilde{\eta}(w)$ denote the probability distribution which $\tilde{\eta}$ induces over worker w 's set of potential partners $F \cup \{\emptyset\}$ and let $\tilde{\eta}(f)$ denote the probability distribution which $\tilde{\eta}$ induces over firm f 's set of potential partners 2^W . Namely, for $w \in W$ and all $v \in F \cup \{\emptyset\}$,

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

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

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

A mechanism φ and a common prior \tilde{P} define a direct (ordinal) preference revelation game under incomplete information as follows. Before submitting a list to

¹⁴Strictly speaking \tilde{P} cannot be set equal to P because \tilde{P} is not a random variable but a probability distribution on \mathcal{P} . However, for convenience we use this notation as if \tilde{P} were a random variable.

¹⁵Notice that we rule out interdependent preferences where the preferences of the other agents influence agent v 's preference. Chakraborty, Citanna, and Ostrovsky (2007) study two-sided matching with interdependent preferences.

the mechanism, agents learn their types. Thus, a strategy of agent v is a function $s_v : \mathcal{P}_v \rightarrow \mathcal{P}_v$ 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 $\varphi : \mathcal{P} \rightarrow \mathcal{M}$ and a common prior \tilde{P} over \mathcal{P} , a strategy profile $s : \mathcal{P} \rightarrow \mathcal{P}$ induces a random matching $\varphi[s(\tilde{P})]$ in the following way: for all $\mu \in \mathcal{M}$,

$$\sum_{P \in \mathcal{P}: \varphi[s(P)] = \mu} \Pr\{\tilde{P} = P\}$$

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. Observe that this uncertainty is held by each agent v , given his type P_v , before the mechanism proposes a matching. Then, after collecting the preference profile $s(P)$, the mechanism φ proposes a unique matching $\varphi[s(P)]$, the one that agents have to carry out. Our stability condition on the mechanism applies to this proposed matching (*i.e.*, it is the standard notion of stability under complete information). This is why we do not need an *ex ante* notion of stability relative to the incomplete information environment when the partners are still uncertain and no particular matching has been proposed yet.

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

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

(fo2)¹⁶ A random matching $\tilde{\eta}$ *first-order stochastically P_f -dominates* a random match-

¹⁶Observe that this definition requires that $\tilde{\eta}$ first-order stochastically dominates $\tilde{\eta}'$ according to

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

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

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 rank 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 well-defined if an agent only observes his ordinal ranking and may have (still) little information about his utilities of his potential partners. Below we define the notion of ordinal Bayesian Nash equilibrium and formally state that for any ordinal Bayesian Nash equilibrium, each agent's strategy maximizes his expected utility for all utility representations of his true preference.

Definition 4 (Ordinal Bayesian Nash Equilibrium) Let \tilde{P} be a common prior.

- (a) Truth-telling 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 $\Pr\{\tilde{P}_v = P_v\} > 0$,

$$\varphi[P_v, \tilde{P}_{-v}|P_v](v) \succ_{P_v} \varphi[P'_v, \tilde{P}_{-v}|P_v](v) \quad \text{for all } P'_v \in \mathcal{P}_v. \quad (1)$$

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.

¹⁷In 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 prior $\tilde{P}_{-v}|P_v$ cannot be derived from \tilde{P} . However, we could complete the prior 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 .

- (b) 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 $\Pr\{\tilde{P}_v = P_v\} > 0$,

$$\varphi[s_v(P_v), s_{-v}(\tilde{P}_{-v}|P_v)](v) \succ_{P_v} \varphi[P'_v, s_{-v}(\tilde{P}_{-v}|P_v)](v) \quad \text{for all } P'_v \in \mathcal{P}_v. \quad (2)$$

The following terminology is standard from expected utility theory. Let P_i be agent i 's strict preference relation on a finite set of alternatives A and $u_i : A \rightarrow \mathbb{R}$ be a (vNM) utility function. Then u_i is a utility representation of P_i if for all $a, b \in A$: $aP_ib \Leftrightarrow u_i(a) > u_i(b)$. For any $w \in W$ and any $P_w \in \mathcal{P}_w$, let $\mathcal{U}_w(P_w)$ denote the set of all utility representations $u_w : F \cup \{\emptyset\} \rightarrow \mathbb{R}$ of P_w ; and for any $f \in F$, any $P_f \in \mathcal{P}_f$, and any $P_f^* \in \text{resp}(P_f)$, let $\mathcal{U}_f(P_f^*)$ denote the set of all utility representations $u_f : 2^W \rightarrow \mathbb{R}$ of P_f^* . Let $\mathcal{U}_f(\text{resp}(P_f)) = \cup_{P_f^* \in \text{resp}(P_f)} \mathcal{U}_f(P_f^*)$ denote the set of all utility representations of all responsive extensions of P_f . We omit the proof of the following lemma.¹⁸

Lemma 1 *Let \tilde{P} be a common prior, φ be a mechanism and s be a strategy profile. Then s is an OBNE in the mechanism φ under incomplete information \tilde{P} if and only if conditions (i) and (ii) hold:*

- (i) *For all $w \in W$, all $P_w \in \mathcal{P}_w$ such that $\Pr\{\tilde{P}_w = P_w\} > 0$, and all $u_w \in \mathcal{U}_w(P_w)$:*

$$\sum_{P_{-w} \in \mathcal{P}_{-w}} \Pr\{\tilde{P} = (P_w, P_{-w})\} u_w(\varphi[s(P)](w)) \geq \sum_{P_{-w} \in \mathcal{P}_{-w}} \Pr\{\tilde{P} = (P_w, P_{-w})\} u_w(\varphi[P'_w, s_{-w}(P_{-w})](w))$$

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

- (ii) *For all $f \in F$, all $P_f \in \mathcal{P}_f$ such that $\Pr\{\tilde{P}_f = P_f\} > 0$, and all $u_f \in \mathcal{U}_f(\text{resp}(P_f))$:*

$$\sum_{P_{-f} \in \mathcal{P}_{-f}} \Pr\{\tilde{P} = (P_f, P_{-f})\} u_f(\varphi[s(P)](f)) \geq \sum_{P_{-f} \in \mathcal{P}_{-f}} \Pr\{\tilde{P} = (P_f, P_{-f})\} u_f(\varphi[P'_f, s_{-f}(P_{-f})](f))$$

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

¹⁸The proof is easy and we refer the interested reader to Theorem 3.11 in d'Aspremont and Peleg (1988).

Lemma 1 formally states that in any OBNE, each agent's strategy maximizes his expected utility for any utility representation of his true preference. The following example establishes that given a common prior \tilde{P} , the set of OBNE in a stable mechanism is non-empty.

Example 1 Imagine that the workers and the firms are divided into “local” matching markets as follows: let $(W_f)_{f \in F}$ be a partition of the set of workers (allowing $W_f = \emptyset$ for some firms f) where W_f denotes the set of workers belonging to the “local” market of f . Loosely speaking, any worker belonging to the local market of firm f applies to f if and only if firm f is acceptable for the worker (and she never applies to other firms), and any firm chooses from its applicants according to its true preference relation restricted to its local market.

Formally, given the partition $(W_f)_{f \in F}$ of W , let the strategy profile s be defined in the following way: (i) for any $w \in W$ and any $P_w \in \mathcal{P}_w$, $A(s_w(P_w)) = \{f\}$ if both $f \in A(P_w)$ and $w \in W_f$, and $A(s_w(P_w)) = \emptyset$ otherwise; and (ii) for all $f \in F$ and all $P_f \in \mathcal{P}_f$, let $A(s_f(P_f)) = A(P_f) \cap W_f$ and $s_f(P_f)|_{(A(P_f) \cap W_f)} = P_f|_{(A(P_f) \cap W_f)}$.

In other words, if worker w belongs to the local market of firm f , then w ranks f uniquely acceptable if f is preferred to being unmatched and otherwise w ranks no firm acceptable. Any firm f ranks as acceptable all workers which both belong to its local market and are acceptable according to its true preference relation.

Claim: For any stable mechanism φ and any profile P , $s(P)$ is a NE in the mechanism φ under complete information P .

Proof: To see that, first consider any $w \in W$. Let $f \in F$ be such that $w \in W_f$. By the definition of s , we have $w \notin A(s_{f'}(P_{f'}))$ for all $f' \in F \setminus \{f\}$. Hence, by stability of φ ,

$$\varphi[P'_w, s_{-w}(P_{-w})](w) \notin F \setminus \{f\} \text{ for all } P'_w \in \mathcal{P}_w. \quad (3)$$

Now if $f \notin A_w(P_w)$, then $A_w(s_w(P_w)) = \emptyset$ and by stability of φ , $\varphi[s(P)](w) = \emptyset$. Thus, from (3), $\varphi[P'_w, s_{-w}(P_{-w})](w) \in \{\emptyset, f\}$ and $\varphi[s(P)](w) R_w \varphi[P'_w, s_{-w}(P_{-w})](w)$ for all $P'_w \in \mathcal{P}_w$. If $f \in A_w(P_w)$, then by (3), the stability of φ and both $A(s_f(P_f)) =$

$A(P_f) \cap W_f$ and $s_f(P_f)|(A(P_f) \cap W_f) = P_f|(A(P_f) \cap W_f)$, $\varphi[P'_w, s_{-w}(P_{-w})](w) = \varphi[s(P)](w)$ for all $P'_w \in \mathcal{P}_w$ such that $f \in A(P'_w)$ and $\varphi[P'_w, s_{-w}(P_{-w})](w) = \emptyset$ for all $P'_w \in \mathcal{P}_w$ such that $f \notin A(P'_w)$. Thus, $\varphi[s(P)](w) R_w \varphi[P'_w, s_{-w}(P_{-w})](w)$ for all $P'_w \in \mathcal{P}_w$.

Second, consider now any $f \in F$. By the definition of s , we have $f \notin A(s_w(P_w))$ for all $w \in W \setminus W_f$. Hence, by stability of φ , $\varphi[P'_f, s_{-f}(P_{-f})](f) \subseteq W_f$ for all $P'_f \in \mathcal{P}_w$. Furthermore, by $A_w(s_w(P_w)) \subseteq \{f\}$ for all $w \in W_f$ and both $A(s_f(P_f)) = A(P_f) \cap W_f$ and $s_f(P_f)|(A(P_f) \cap W_f) = P_f|(A(P_f) \cap W_f)$, we obtain $\varphi[s(P)](f) R_f^* \varphi[P'_f, s_{-f}(P_{-f})](f)$ for all $P'_f \in \mathcal{P}_f$ and all $P_f^* \in \text{resp}(P_f)$. \square

Since for any profile P , $s(P)$ is a NE in any stable mechanism φ under complete information P , it follows that s is an OBNE in any stable mechanism φ under any common prior \tilde{P} .

In the special case where $W_f = W$ for some firm f , firm f has a monopolistic market in Example 1.

Both under complete and incomplete information there is a multiplicity of OBNE and the existence of OBNE is guaranteed. Observe that complete information is the particular instance of incomplete information where the common prior puts probability one on a unique profile. Thus, the notion of OBNE inherits all properties of NE and like in NE, there is no reason to expect that agents play undominated strategies in OBNE.

4 The Main Result and Its Applications

The support of a common prior \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. This link holds for any stable mechanism and not only for the deferred-

acceptance algorithms.

Theorem 1 *Let \tilde{P} be a common prior, 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 in the stable mechanism φ under complete information P .*

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 prior 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 prior and 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.¹⁹

Remark 1 Theorem 1 provides for stable mechanisms a strong link between OBNE under incomplete information and NE under complete information. Neither for (ordinal) games of incomplete information nor for many-to-one matching markets using unstable mechanisms this link is true. It is the stability of mechanisms which drives our main result. Later we exhibit an example of an OBNE in an unstable mechanism where for some profiles in the support of the common prior, the submitted profile is not a Nash equilibrium under complete information at the true profile in the direct preference revelation game induced by the unstable mechanism.

While the proof of Theorem 1's (If)-part is straightforward, its (Only if)-part proceeds roughly as follows. If for some profile P in the support of \tilde{P} , $s(P)$ does

¹⁹Of course, constructing OBNE requires more than just choosing a NE for every profile in the support of the common prior because we need to assure that the chosen NE yield a strategy for each agent in the direct preference revelation game under incomplete information.

not constitute a NE, then some agent v has a profitably deviation from $s(P)$ under complete information P . Using this deviation we then construct another deviation and show that agent v profitably manipulates given his type P_v and his prior $\tilde{P}_{-v}|P_v$ which implies that strategy profile s cannot be an OBNE. This step uses repeatedly the following peculiarities of stable matchings in college admissions problems: (1) *invariance of unmatched agents and unfilled positions*: the set of unmatched agents and any firm's number of unfilled positions are the same for all stable matchings; and (2) *comparative statics*: 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 smaller problem.

Remark 2 Theorem 1 restricts attention to mechanisms which choose for each profile one of its stable matchings. Alternatively one may consider *random* stable mechanisms choosing for each profile a lottery over its stable matchings. It is easy to adapt the proof of Theorem 1 for random stable mechanisms.²⁰ Hence, it is without loss of generality in Theorem 1 to consider deterministic stable mechanisms. Furthermore, we focus on stable mechanisms because those are used in many real-life matching markets.²¹

Below we turn to the applications of Theorem 1.

²⁰Details are available from the authors upon request. Pais (2008) provides a strategic analysis of random stable mechanisms under complete information.

²¹Due to this reason, we do not consider mechanisms with larger message spaces which seem to be unrealistic for matching markets such as “shoot them all” mechanisms where under complete information every agent has to reveal a complete preference profile and everybody remains unmatched in case of disagreement of the reported profiles. This guarantees truth-telling to be a NE under complete information. We do not know of any real-life centralized matching market where these general mechanisms are used.

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 prior. Therefore, (a) the set of OBNE is identical for any two common priors with equal support and (b) the set of OBNE shrinks if the support of the common prior 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 priors 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 priors 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 prior which leave its support unchanged. Therefore, any OBNE remains an equilibrium if agents have different priors with equal support, *i.e.* each agent v may have a private prior \tilde{P}^v but all private priors have identical (or common) support.²² This consequence is especially important for applications since for many of them, the common prior assumption might be too strong.

By (b) of Corollary 1, the set of OBNE with full support (*i.e.* all common priors which put positive probability on all profiles) is contained in the set of OBNE of any arbitrary common prior (or support). It turns out that OBNE with full support

²²Then in Definition 4 of OBNE the common prior \tilde{P} is replaced for each agent v by his private prior \tilde{P}^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 prior. If all private priors have equal support, then it follows that a strategy profile s is an OBNE with private priors (with common support) if and only if for any profile P in the common support, $s(P)$ is a Nash equilibrium in the mechanism φ under complete information P .

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, \dots, w_{q_v} , we have for all $w' \in A(P'_v)$ and all $w \in A(P'_v) \setminus \{w_1, \dots, w_{q_v}\}$, $P'_v|\{w, w'\} = P_v|\{w, w'\}$. For example, if $q_v = 2$ and $P_v : w_1w_2w_3w_4\emptyset \dots$, then $P'_v : w_3w_2w_4\emptyset \dots$ and $P''_v : w_2w_1w_4\emptyset \dots$ are essentially P_v -truthful. Observe that condition (i) above will require in addition that $A(P'_v) \subseteq A(P_v)$ and $A(P''_v) \subseteq A(P_v)$.

Notice that the example of the local markets after Definition 4 is an OBNE where each agent's reported rankings contain acceptable matches only and are essentially truthful.

Corollary 2 (Essential Truthfulness for Full Support) *Let \tilde{P} be a common prior 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).

First we show that $A(s_v(P_v)) \subseteq A(P_v)$. Suppose that $A(s_v(P_v)) \setminus A(P_v) \neq \emptyset$. Let $w \in A(s_v(P_v)) \setminus A(P_v)$ and $P_{-v} \in \mathcal{P}_{-v}$ be such that $A(P_w) = \{v\}$ and for all $v' \in V \setminus \{v, w\}$, $A(P_{v'}) = \emptyset$. 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 $v' \in V \setminus \{v, w\}$, $A(P_{v'}) = \emptyset$ 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, \dots, w_{q_v} be the q_v most preferred workers under $s_v(P_v)$. Let $W' = \{w_1, \dots, w_{q_v}\}$. 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 $w \in A(s_v(P_v)) \setminus W'$, $w's_v(P_v)ws_v(P_v)\emptyset$ and $wP_vw'P_v\emptyset$.²³ 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 $v' \in V \setminus (\{v, w, w'\} \cup W')$, $A(P_{v'}) = \emptyset$. 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 $v' \in V \setminus (\{v, w, w'\} \cup W')$, $A(P_{v'}) = \emptyset$ 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 $A(P'_v) = (W' \setminus \{w'\}) \cup \{w\}$ because by $v \in A(s_w(P_w))$, $\varphi[P'_v, s_{-v}(P_{-v})](v) = (W' \setminus \{w'\}) \cup \{w\}$ and both wP_vw' and responsiveness imply $(W' \setminus \{w'\}) \cup \{w\}P_v^*W' = \varphi[s(P)](v)$ for all $P_v^* \in \text{resp}(P_v)$. This means that $s(P)$ is not a NE in φ for P , a contradiction. \square

Note that any OBNE for a common prior with full support is an OBNE for any arbitrary prior. Hence, such OBNE are invariant with respect to the common prior

²³Observe 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 $fP_vf'P_v\emptyset$.

and remain OBNE if the agents' priors are not necessarily derived from the same common prior. Of course, by Corollary 2, those OBNE are robust to changes of the common prior(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: Truth-Telling and Realized Matchings

When agents' preferences are private information, we would like to design a mechanism which elicits the true preferences from the agents. Essential truthfulness only partially achieves this because agents are not required to submit preference relations containing all acceptable matches. For real-life environments we are interested whether agents truth-tell and 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.

In order to guarantee that agents truthfully report their preferences, incentive-compatible mechanisms make truth-telling a (weakly) best reply when the other agents truth-tell.

Definition 5 (Incentive Compatibility) A mechanism φ is *incentive compatible* if for each profile $P \in \mathcal{P}$, we have for all $w \in W$, $\varphi[P_w, P_{-w}](w) R_w \varphi[\hat{P}_w, P_{-w}](w)$ for all $\hat{P}_w \in \mathcal{P}_w$, and for all $f \in F$ and all $P_f^* \in \text{resp}(P_f)$, $\varphi[P_f, P_{-f}](f) R_f^* \varphi[\hat{P}_f, P_{-f}](f)$ for all $\hat{P}_f \in \mathcal{P}_f$.

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 priors.

Since incentive-compatibility is a strong condition, our incomplete information environment allows a weaker (but still natural) condition. Given a common prior and a mechanism, Bayesian incentive-compatibility requires that all agents truthfully reveal their preferences at any profile belonging to the support of the common prior.

Definition 6 (Bayesian Incentive Compatibility) Let \tilde{P} be a common prior. A mechanism φ is *Bayesian incentive compatible under incomplete information \tilde{P}* if for all $v \in V$ and all $P_v \in \mathcal{P}_v$ such that $\Pr\{\tilde{P}_v = P_v\} > 0$,

$$\varphi[P_v, \tilde{P}_{-v}|P_v](v) \succ_{P_v} \varphi[P'_v, \tilde{P}_{-v}|P_v](v) \quad \text{for all } P'_v \in \mathcal{P}_v. \quad (4)$$

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 prior.

Corollary 3 *Let \tilde{P} be a common prior. Then a stable mechanism φ is Bayesian incentive compatible under incomplete information \tilde{P} if and only if for any profile P in the support of \tilde{P} , P is a Nash equilibrium in the mechanism φ under complete information P .*

In other words, a stable mechanism φ is Bayesian incentive compatible under the common prior \tilde{P} if and only if φ restricted to the support of \tilde{P} is incentive compatible. This result yields a strong connection between Bayesian incentive compatibility and incentive compatibility for stable mechanisms in matching markets. Furthermore, it can be easily seen that for P to be a Nash equilibrium in the mechanism φ under complete information P is independent of the stable mechanism. All what matters is the stability of the mechanism.

If each firm has exactly one position, then Ehlers and Massó (2007) show that a singleton core is necessary and 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 (2007) as a corollary from Theorem 1.

Corollary 4 *[Theorem 1 in Ehlers and Massó (2007)] Let $q_f = 1$ for all $f \in F$ and \tilde{P} be a common prior. 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.*

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 prior, 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.

This is also the reason why Roth (1984b) does not apply to college admissions problems. Roth (1984b) shows that when each firm has exactly one position, the outcome of any NE under complete information of DA_W where workers truth-tell (*i.e.* they play their weakly dominant strategy) is stable with respect to the true profile (see also Theorem 4.16 of Roth and Sotomayor (1990)). The analogous statement is true for DA_F when each firm has exactly one position. Once at least one firm has several positions, truth-telling is not a weakly dominant strategy for the firms in DA_F and there is no reason to expect that firms truth-tell for a given prior. Furthermore, it is easily seen that in the example of Roth (1985a) for DA_W , if workers truth-tell, then there are NE where the outcome is not stable with respect to the true profile.

Of course, by Theorem 1, Roth (1984b) applies when each firm has exactly one position.

Corollary 5 *Let $q_f = 1$ for all $f \in F$ and \tilde{P} be a common prior. Let s be a strategy profile such that $s_w(P_w) = P_w$ for all $w \in W$ and all $P_w \in \mathcal{P}_w$. If s is an OBNE in the stable mechanism DA_W under incomplete information \tilde{P} , then for all profiles P in the support of \tilde{P} , $DA_W[s(P)]$ is stable with respect to P . The analogous statement is true for the stable mechanism DA_F .*

Proof. Let $P \in \mathcal{P}$ be such that $\Pr\{\tilde{P} = P\} > 0$. By Theorem 1, $s(P)$ is a NE in DA_W under complete information P . Hence, by $s_w(P_w) = P_w$ for all $w \in W$ and Roth (1984b), $DA_W[s(P)] \in C(P)$, the desired conclusion. \square

Note that in the above applications the preferences of each side of the market are allowed to be uncorrelated. However, in empirical applications the preferences of one side of the market may be perfectly correlated. For example, each firm may rank all workers according to an objective criterion such as their degree of qualifications or each college may rank all students according to their grades. Furthermore, it is common in labor economics or search theory to often assume that all workers have identical preferences over firms.²⁴ One-sided perfect correlation is an extreme case of interdependence of preferences where an agent's preference may depend on the preferences of the other agents on his side.

We say that a common prior \tilde{P} is *F-correlated* if for any profile P in the support of \tilde{P} , all firms have identical preferences.²⁵ Similarly we say that a prior \tilde{P} is *W-correlated* if for any profile P in the support of \tilde{P} , all workers have identical preferences.

Corollary 6 *Let \tilde{P} be a common prior.*

- (a) *If \tilde{P} is F-correlated or W-correlated, then truth-telling is an OBNE in any stable mechanism under incomplete information \tilde{P} .*
- (b) *Let s be a strategy profile such that $s_w(P_w) = P_w$ for all $w \in W$ and all $P_w \in \mathcal{P}_w$. If \tilde{P} is W-correlated and s is an OBNE in the stable mechanism DA_W under incomplete information \tilde{P} , then for all profiles P in the support of \tilde{P} , $DA_W[s(P)]$ is stable with respect to P . The analogous statement is true for the stable mechanism DA_F .*

Proof. (a) Let φ be a stable mechanism and \tilde{P} be a common prior. Without loss of generality, let \tilde{P} be *F-correlated*. The case where \tilde{P} is *W-correlated* is analogous to the case where \tilde{P} is *F-correlated* and all firms have quota 1. Let P be in the

²⁴For instance, Shi (2002) provides a long list of papers on directed search models in labor markets where at least one side of the market is homogenous.

²⁵Formally this means for all $f, f' \in F$, $A(P_f) = A(P_{f'})$ and $P_f|W = P_{f'}|W$.

support of \tilde{P} . Because all firms' preferences are identical at P , we have $|C(P)| = 1$, say $C(P) = \{\mu\}$. By stability of φ , $\varphi[P] = \mu$. By Theorem 1, it suffices to show that P is a NE in the mechanism φ under complete information P .

Because all firms have identical preferences, say $P_f : w_1 w_2 \cdots w_k \emptyset w_{k+1} \cdots$ for all $f \in F$, $\mu(w_1)$ is w_1 's most preferred firm (if any) under P_{w_1} . Then $\mu(w_2)$ is w_2 's most preferred firm (if any) from $F \setminus \{\mu(w_1)\}$ under P_{w_2} , and in general for $i = 1, \dots, |W|$, $\mu(w_i)$ is w_i 's most preferred firm (if any) from $F \setminus \{\mu(w_1), \dots, \mu(w_{i-1})\}$ under P_{w_i} . By stability of φ , obviously no worker can profitably manipulate.

Let $f \in F$, $P'_f \in \mathcal{P}_f$, and $\mu' = \varphi[P'_f, P_{-f}]$. Suppose that for some $P_f^* \in \text{resp}(P_f)$ we have $\mu'(f) P_f^* \mu(f)$. Hence, $\mu'(f) \neq \mu(f)$. By stability of φ , without loss of generality we may suppose $A(P'_f) = \mu'(f)$ and because any firm's set of acceptable workers is $\{w_1, \dots, w_k\}$, $A(P'_f) \subseteq A(P_f)$. Again, by stability of φ , $\mu(f) \subseteq A(P_f)$. First, suppose that $|\mu'(f)| > |\mu(f)|$. Note that $\mu \in C(P)$ and $\mu' \in C(P'_f, P_{-f})$. Then, without loss of generality, we may suppose $|\mu'(f)| = q_f$.²⁶ Let $P''_f \in \mathcal{P}_f$ be such that $A(P''_f) = A(P_f)$ and $P''_f | A(P''_f) = P'_f | A(P''_f)$. By $|\mu(f)| < q_f$, $A(P''_f) = A(P_f)$ and (P3), we obtain $\mu \in C(P''_f, P_{-f})$. By $|\mu(f)| \neq |\mu'(f)|$ and (P3), we must have $\mu' \notin C(P''_f, P_{-f})$. Thus, μ' is blocked by some pair (w', f') under (P''_f, P_{-f}) . By $\mu' \in C(P'_f, P_{-f})$ and $A(P'_f) \subseteq A(P''_f)$, we must have $f' = f$ and $w' \notin \mu'(f)$. But then by $A(P'_f) = \mu'(f)$ and $P''_f | A(P''_f) = P'_f | A(P''_f)$, we must have $w P_f w'$ for all $w \in \mu'(f)$. Now f must have an unfilled slot under μ' and $|\mu'(f)| < q_f$, a contradiction.

Hence, $|\mu'(f)| \leq |\mu(f)|$. Let \underline{w} be the least P_f -preferred worker in $\mu(f)$, i.e. $w R_f \underline{w}$ for all $w \in \mu(f)$. If for all $w \in \mu'(f) \setminus \mu(f)$ we have $\underline{w} P_f w$, then by $|\mu'(f)| \leq |\mu(f)|$ and responsiveness of P_f^* , we have $\mu(f) R_f^* \mu'(f)$, a contradiction. Let $w_l \in \mu'(f) \setminus \mu(f)$ be such that $w_l P_f \underline{w}$. We show that there exists an index $i(l) < l$ such that $w_{i(l)} \in \mu(f) \setminus \mu'(f)$. Let $\mu(w_l) = f_l$. Note that by $f_l \neq f$, $w_l P_f \underline{w}$ and $\mu \in C(P)$ we have $f_l P_{w_l} f$. Thus, by $\mu' \in C(P'_f, P_{-f})$, we must have $|\mu'(f_l)| = q_{f_l}$ and $w P_{f_l} w_l$ for all

²⁶If $|\mu'(f)| < q_f$, then set $q'_f = |\mu'(f)|$. From (P3) (where we specify both the profile and the quotas), $\mu \in C(P; q)$ and $|\mu(f)| < q'_f$ imply $\mu \in C(P; q'_f, q_{-f})$, and similarly $\mu' \in C(P'_f, P_{-f}; q)$ and $\mu'(f) = A(P'_f)$ imply $\mu' \in C(P'_f, P_{-f}; q'_f, q_{-f})$.

$w \in \mu'(f_l)$. But then w_l 's position at firm f_l is filled with some new worker w' , i.e. $w' \in \mu'(f_l) \setminus \mu(f_l)$ and $w'P_{f_l}w_l$. Now by $\mu \in C(P)$ and $\mu(w_l) = f_l$, $\mu(w')P_{w'}f_l$. If $\mu(w') = f$, then w' has an index $i(l) < l$ such that $w' = w_{i(l)}$ and $w_{i(l)} \in \mu(f) \setminus \mu'(f)$. Otherwise, let $\mu(w') = f' \neq f$. Then again as above from $f'P_{w'}f_l$ we have $|\mu'(f')| = q_{f'}$ and $wP_{f'}w'$ for all $w \in \mu'(f')$, and w' 's position at firm f' is filled with some new worker w'' , i.e. $w'' \in \mu'(f') \setminus \mu(f')$ and $w''P_{f'}w'$. By $P'_{-f} = P_{-f}$ and the finiteness of W and F , in the end for $w_l \in \mu'(f) \setminus \mu(f)$ there must exist $w_{i(l)} \in \mu(f) \setminus \mu'(f)$ with $i(l) < l$. Furthermore, from the above arguments, we can choose $i(l) \neq i(l')$ for all $l \neq l'$ such that $w_l, w_{l'} \in \mu'(f) \setminus \mu(f)$. Since $\mu(f) \subseteq A(P_f)$ and $|\mu'(f)| \leq |\mu(f)|$, responsiveness of P_f^* implies $\mu(f)R_f^*\mu'(f)$, a contradiction.

(b) Let P be in the support of \tilde{P} . Since $s_w(P_w) = P_w$ for all w and \tilde{P} is W -correlated, we have by (a) that no worker can gain by manipulation. Furthermore, by Theorem 1, $s(P)$ must be a NE in DA_W under P . Because \tilde{P} is W -correlated, all workers have identical preferences, say $P_w : f_1f_2 \cdots f_l \emptyset f_{l+1} \cdots$ for all $w \in W$. Suppose that $DA_W[s(P)]$ is not stable with respect to P . Since $s(P)$ is a NE in DA_W under P , no agent is matched to any partner under $DA_W[s(P)]$ which is unacceptable according to its true preference relation. Suppose that some unmatched worker-firm pair (w, f) blocks $DA_W[s(P)]$. Then $f \in A(P_w)$ and by $s_w(P_w) = P_w$, $fP_wDA_W[s(P)](w)$. But then, along the DA_W -algorithm which produces $DA_W[s(P)]$, worker w proposed to f before proposing to $DA_W[s(P)](w)$ and because all workers' submitted lists are identical, at that step all unmatched workers proposed to f (and the set of unmatched workers shrinks from one step to the next one). Let w' be the least preferred worker according to P_f in $DA_W[s(P)](f)$. But now f profitably deviates from $s(P)$ in DA_W by submitting a list P'_f where $A(P'_f) = (DA_W[s(P)](f) \cup \{w\}) \setminus \{w'\}$. When in $DA_W[P'_f, s_{-f}(P_{-f})]$ worker w proposes to f , all unmatched workers propose to f in that step because all workers' submitted lists are identical. Firm f accepts $(DA_W[s(P)](f) \cup \{w\}) \setminus \{w'\}$ which is strictly preferred to $DA_W[s(P)](f)$ under any responsive extension P_f^* of P_f . Hence, $s(P)$ is not a NE in DA_W under P , a contra-

diction. □

Although Corollary 6 focuses on completely correlated priors, it is easy to extend it in the following direction. Suppose that each worker has a certain qualification and each firm only offers positions having the same job-specific qualification. Let all firms, which are interested in the same qualification, have identical preferences over all workers possessing this qualification for any realization in the common prior. Then the qualifications segregate the matching market and the conclusions of Corollary 6 apply. For example, each firm may represent a certain department in a hospital and they would like to fill their positions with physicians who studied the medical specialty of their department.

For a given OBNE we are interested in 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. Roth (1989) showed that ex-post stability cannot be ensured in (not necessarily ordinal) Bayesian Nash equilibrium of stable mechanisms.²⁷ For OBNE 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. We think that this is an extremely important property because it justifies ex-post the use of the particular stable mechanism φ .

Corollary 7 (Ex-Post Unanimity) *Let \tilde{P} be a common prior, 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 $v \in F$

²⁷As far as we know Roth (1989) is the first paper studying stable mechanisms under incomplete information.

(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 $P_v^* \in \text{resp}(P_v)$. 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 $\mu(v) \setminus \varphi[s(P)](v) \neq \emptyset$ and by Theorem 4 of Roth and Sotomayor (1989), for all $w \in \mu(v)$ and all $w' \in \varphi[s(P)](v) \setminus \mu(v)$, $wP_v w'$. 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. \square

Ehlers and Massó (2007, Theorem 2) showed Corollary 7 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. It follows directly from Corollary 7 that truth-telling is an OBNE only if the core is singleton at any realized profile.

4.3 Application III: Robust Mechanism Design

In this subsection we will relate our results to Bergemann and Morris (2005)'s equivalence result for all (payoff) type spaces on robust mechanism design.²⁸ Instead of the terminology of payoff type spaces, in (ordinal) matching markets it is natural to use the term “preference type spaces”.

Let $\mathcal{Q} \subseteq \mathcal{P}$ denote a set of possible preference type profiles. Then let $\mathcal{Q}_v = \{P_v | P \in \mathcal{Q}\}$ be the set of agent v 's possible preference types in \mathcal{Q} . Furthermore, let $\mathcal{Q}_{-v}|_{P_v} = \{P_{-v} | (P_v, P_{-v}) \in \mathcal{Q}\}$ denote the set of the other agents' preference types in \mathcal{Q} when v 's preference type is P_v . Let $\Delta(\mathcal{Q}_{-v}|_{P_v})$ denote the set of all probability

²⁸We are grateful to an anonymous referee and the managing editor to their suggestion to explore and clarify this connection.

distributions on $\mathcal{Q}_{-v|P_v}$. In our setting a *preference type space* is simply given by $(\mathcal{Q}, (\hat{\pi}_v)_{v \in V})$ where \mathcal{Q} denotes the agents' possible preference profiles and $\hat{\pi}_v$ describes agent v 's priors, *i.e.* for any $P_v \in \mathcal{Q}_v$, $\hat{\pi}_v(P_v) \in \Delta(\mathcal{Q}_{-v|P_v})$ is agent v 's prior about the other agents' preference types when v 's preference type is P_v . A preference type space $(\mathcal{Q}, (\hat{\pi}_v)_{v \in V})$ is a product space if $\mathcal{Q} = \times_{v \in V} \mathcal{Q}_v$. For product preference type spaces we sometimes write $(\mathcal{Q}_v, \hat{\pi}_v)_{v \in V}$ instead of $(\mathcal{Q}, (\hat{\pi}_v)_{v \in V})$. Although Bergemann and Morris (2005) only focussed on product preference type (or payoff type) spaces where $\mathcal{Q}_v = \mathcal{P}_v$ for all agents v , for later purposes we will allow for non-product preference type spaces. In particular, in some applications (as in matching markets), preference type profiles may be correlated and not necessarily be independent from each other, *i.e.* some preference type profiles might be regarded as impossible a priori. Furthermore, \mathcal{Q}_v may be a strict subset of \mathcal{P}_v for an agent v and thus, agent v 's set of possible true preference types may be strict subset of the rankings agent v may report to the mechanism.

A preference type space $(\mathcal{Q}, (\hat{\pi}_v)_{v \in V})$ is said to be “common prior” if there exists a common prior \tilde{P} on \mathcal{Q} such that $\hat{\pi}_v(P_v) = \tilde{P}_{-v|P_v}$ for all $v \in V$ and all $P_v \in \mathcal{Q}_v$. Obviously, for $\hat{\pi}_v$ to be well-defined, we must have $\Pr\{\tilde{P}_v = P_v\} > 0$ for all $P_v \in \mathcal{Q}_v$ because by Bayes' rule we have for all $P_v \in \mathcal{Q}_v$ and all $P_{-v} \in \mathcal{Q}_{-v|P_v}$,

$$\hat{\pi}_v(P_v)[P_{-v}] = \frac{\Pr\{\tilde{P} = (P_v, P_{-v})\}}{\Pr\{\tilde{P}_v = P_v\}}.$$

We denote a common prior preference type space simply by (\mathcal{Q}, \tilde{P}) . In order to make the relation clear, we adapt Bergemann and Morris (2005)'s definitions of interim incentive compatibility and ex-post incentive compatibility to our ordinal matching environment.

Definition 7 (Interim Incentive Compatibility) Let $(\mathcal{Q}, (\hat{\pi}_v)_{v \in V})$ be a preference type space. A mechanism φ is *interim incentive compatible on* $(\mathcal{Q}, (\hat{\pi}_v)_{v \in V})$ if for all $v \in V$ and all $P_v \in \mathcal{Q}_v$,

$$\varphi[P_v, \hat{\pi}_v(P_v)](v) \succ_{P_v} \varphi[P'_v, \hat{\pi}_v(P_v)](v) \quad \text{for all } P'_v \in \mathcal{P}_v. \quad (5)$$

Note that interim incentive compatibility reduces to Bayesian incentive compatibility for common prior preference type spaces. Instead of defining ex-post incentive compatibility for all possible profiles (or all types), for later purposes we define ex-post incentive compatibility for subdomains of the set of all profiles.

Definition 8 (Ex-post Incentive Compatibility on Subdomains) Let $\mathcal{Q} \subseteq \mathcal{P}$. A mechanism φ is *ex-post incentive compatible on \mathcal{Q}* if for each profile $P \in \mathcal{Q}$, we have for all $w \in W$, $\varphi[P_w, P_{-w}](w) R_w \varphi[P'_w, P_{-w}](w)$ for all $P'_w \in \mathcal{P}_w$, and for all $f \in F$ and all $P_f^* \in \text{resp}(P_f)$, $\varphi[P_f, P_{-f}](f) R_f^* \varphi[P'_f, P_{-f}](f)$ for all $P'_f \in \mathcal{P}_f$.

Note that ex-post incentive compatibility on \mathcal{P} is incentive compatibility. When $\mathcal{Q} \neq \mathcal{P}$, then both interim incentive compatibility and ex-post incentive compatibility are stronger than the corresponding versions of Bergemann and Morris (2005) because any agent v 's set of preference types \mathcal{Q}_v may be a strict subset of agent v 's set of possible reports \mathcal{P}_v (and not only \mathcal{Q}_v) to the mechanism.

Corollary 8 *Let $\mathcal{Q} \subseteq \mathcal{P}$ and φ be a stable mechanism. Then the following are equivalent.*

- (a) φ is interim incentive compatible on all preference type spaces $(\mathcal{Q}, (\hat{\pi}_v)_{v \in V})$.
- (b) φ is interim incentive compatible on all common prior preference type spaces (\mathcal{Q}, \tilde{P}) .
- (c) φ is ex-post incentive compatible on \mathcal{Q} .

Proof. (a) \Rightarrow (b) follows by definition because we are asking for interim incentive compatibility on a smaller collection of preference type spaces. (b) \Rightarrow (c) follows from Theorem 1 by considering a common prior preference type space (\mathcal{Q}, \tilde{P}) where \tilde{P} has support \mathcal{Q} . (c) \Rightarrow (a) is trivial for matching markets. \square

Now Corollary 8 is the matching analogue of Corollary 1 of Bergemann and Morris (2005) for single-valued social choice correspondences and all payoff type spaces. The

two important differences between Corollary 1 of Bergemann and Morris (2005) and Corollary 8 are that (i) our equivalence result allows for non-product preference type spaces whereas their result does not and (ii) our incentive compatibility notions are stronger than theirs.

Indeed, when $\mathcal{Q} = \mathcal{P}$ their result implies Corollary 8. However, for $\mathcal{Q} = \mathcal{P}$, ex-post incentive compatible becomes equivalent to incentive compatibility and from Roth (1982) we know that there exists no stable and incentive compatible mechanism. Therefore, by their corollary and Corollary 8, there does not exist any stable mechanism which is interim incentive compatible on all preference type spaces.

Below we will develop for matching markets an interesting positive variant of the equivalence result by Bergemann and Morris (2005).

We say that a preference profile P is *correlated* if either all workers' preferences are correlated ($P_w = P_{w'}$ for all $w, w' \in W$) or all firms' preferences are correlated ($P_f = P_{f'}$ for all $f, f' \in F$). Let $\mathcal{C} \subseteq \mathcal{P}$ denote the set of all correlated preference profiles. Note that for all $v \in V$ we have $\mathcal{C}_v = \mathcal{P}_v$. Now we call the preference type space $(\mathcal{P}_v, \hat{\pi}_v)_{v \in V}$ *correlated* if for all $v \in V$ and all $P_v \in \mathcal{P}_v$, for all Q_{-v} belonging to the support of $\hat{\pi}_v(P_v)$, (P_v, Q_{-v}) is correlated.

Corollary 9 *Let φ be a stable mechanism. Then the following are equivalent.*

- (a) φ is interim incentive compatible on all correlated preference type spaces $(\mathcal{P}_v, \hat{\pi}_v)_{v \in V}$.
- (b) φ is interim incentive compatible on all correlated common prior preference type spaces (\mathcal{C}, \tilde{P}) .
- (c) φ is ex-post incentive compatible on \mathcal{C} .

Proof. (a) \Rightarrow (b) follows by definition because we asking for interim incentive compatibility on a smaller collection of preference type spaces. (b) \Rightarrow (c) follows from Theorem 1 by considering a common prior preference type space (\mathcal{C}, \tilde{P}) where \tilde{P} has support \mathcal{C} . (c) \Rightarrow (a) follows from the fact that by Theorem 1 and by part (a) of Corollary 6, truth-telling is a Nash equilibrium in any stable mechanism under complete

information for any correlated profile. □

By Corollary 9, any stable mechanism is interim incentive compatible on all correlated preference type space if and only if the stable mechanism is ex-post incentive compatible on the domain of all correlated preference profiles. Note that a mechanism designer might not know the exact preferences of agents but only certain properties about profiles (such as correlation). Now arbitrary preferences are admissible for any agent and any direct mechanism will need to be defined for arbitrary preference profiles and hence for arbitrary individual deviations from the true profile. Any stable mechanism is both ex-post and interim incentive compatible on correlated preference type spaces and any agent's prior may put simultaneously positive probability on workers' correlated preference profiles and firms' correlated preference profiles. Restrictions (like correlation) on the domain of preference profiles may be well-defined in applications because certain preference profiles may be regarded impossible. It would be interesting to investigate whether a similar result (with restrictions on the type space) as Corollary 9 can be obtained in the more general setting of separable environments of Bergemann and Morris (2005).

Below we will reformulate Corollary 9 in terms of implementability of the stable matching correspondence C , *i.e.* in our matching environment the equivalence holds for both (single-valued) stable mechanisms and the stable matching correspondence.

Definition 9 (Implementability)

- (i) Let $(\mathcal{P}_v, \hat{\pi}_v)_{v \in V}$ be a preference type space. A matching correspondence $M : \mathcal{P} \rightarrow 2^{\mathcal{M}} \setminus \{\emptyset\}$ is *interim incentive compatible* on $(\mathcal{P}_v, \hat{\pi}_v)_{v \in V}$ if there exists a mechanism $\varphi : \mathcal{P} \rightarrow \mathcal{M}$ such that φ is interim incentive compatible on $(\mathcal{P}_v, \hat{\pi}_v)_{v \in V}$ and for each profile $P \in \mathcal{P}$, $\varphi[P] \in M(P)$.
- (ii) A matching correspondence $M : \mathcal{P} \rightarrow 2^{\mathcal{M}} \setminus \{\emptyset\}$ is *ex-post implementable* on \mathcal{Q} if there exists a mechanism $\varphi : \mathcal{P} \rightarrow \mathcal{M}$ such that φ is ex-post incentive compatible on \mathcal{Q} and $\varphi[P] \in M(P)$ for all $P \in \mathcal{Q}$.

Let $C : \mathcal{P} \rightarrow 2^{\mathcal{M}} \setminus \{\emptyset\}$ denote the stable matching correspondence.

Corollary 10 *The following are equivalent.*

- (a) C is interim implementable on all correlated preference type spaces $(\mathcal{P}_v, \hat{\pi}_v)_{v \in V}$.
- (b) C is interim implementable on all correlated common prior preference type spaces (\mathcal{C}, \tilde{P}) .
- (c) C is ex-post implementable on \mathcal{C} .

Proof. Note that by Theorem 1 and by part (a) of Corollary 6, for any correlated profile, truth-telling is a Nash equilibrium in any stable mechanism under complete information. Now Corollary 10 follows from applying Corollary 9 to a stable mechanism φ . □

Corollary 10 is a positive equivalence result for matching markets because it says that stable matchings can be implemented (via any stable mechanism) in correlated environments.

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. Theorem 1 is in general not true for OBNE for arbitrary mechanisms. For instance, suppose that the common prior \tilde{P}^u is uniform in the sense that it puts equal probability on all preference profiles. Furthermore, suppose that the mechanism φ matches a worker and a firm if and only if they rank each other as their most preferred choice (and φ leaves all other positions unfilled and all other workers unmatched). Then it is easy to verify that truth-telling is an OBNE in the mechanism

φ under the uniform prior \tilde{P}^u .²⁹ However, truth-telling is not always a NE in the mechanism φ under complete information since for some profiles, a firm may rank a worker first and a worker the firm second, and if the worker is unmatched, then she profitably manipulates by moving the firm to the first position of her submitted ranking. Hence, the reason for the link between the ex-ante and ex-post equilibrium in our main result relies on the stability of mechanisms.

In our framework one may be tempted to apply the revelation principle. This means that for a given OBNE s in a stable mechanism φ under \tilde{P} , we define another mechanism φ^s such that truth-telling is an OBNE in φ^s under \tilde{P} . However, there is no reason to expect that φ^s is a stable mechanism. For example, for any of the “local market” OBNE after Definition 4 the induced mechanism is unstable. And again, both stable mechanisms are frequently used in real-life matching markets and we are interested in those. Finally, the above example also shows that after invoking the revelation principle, the strong link of Theorem 1 is not true for truth-telling and arbitrary mechanisms. It is not clear which properties matching mechanisms possess where truth-telling is an OBNE.

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.

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).

²⁹For any agent v and any $P_v \in \mathcal{P}_v$, $\tilde{P}_{-v}^u|_{P_v}$ is uniform over \mathcal{P}_{-v} . For all agents belonging to the opposite side of the market, the probability that she ranks v first is identical. Hence, v cannot do better than submitting the true preference relation.

- [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. “Implementation of Stable Solutions to Marriage Problems,” *Journal of Economic Theory* **69**, 240-254 (1996).
- [5] D. Bergemann and S. Morris. “Robust Mechanism Design,” *Econometrica* **73**, 1771-1813 (2005).
- [6] A. Chakraborty, A. Citanna, and M. Ostrovsky. “Two-Sided Matching with Interdependent Values,” mimeo, Stanford University (2007).
- [7] Y. Chen and T. Sönmez. “School Choice: An Experimental Study,” *Journal of Economic Theory* **127**, 202-231 (2006).
- [8] V.P. Crawford. “Comparative Statics in Matching Markets,” *Journal of Economic Theory* **54**, 389-400 (1991).
- [9] L. Dubins and D. Freedman. “Machiavelli and the Gale-Shapley Algorithm,” *American Mathematical Monthly* **88**, 485-494 (1981).
- [10] L. Ehlers. “Truncation Strategies in Matching Markets,” *Mathematics of Operations Research* **33**, 327-335 (2008).
- [11] L. Ehlers and J. Massó. “Incomplete Information and Singleton Cores in Matching Markets,” *Journal of Economic Theory* **136**, 587-600 (2007).
- [12] H. Ergin and T. Sönmez. “Games of School Choice under the Boston Mechanism,” *Journal of Public Economics* **90**, 215-237 (2006).
- [13] D. Gale and L. Shapley. “College Admissions and the Stability of Marriage,” *American Mathematical Monthly* **69**, 9-15 (1962).

- [14] D. Gale and M. Sotomayor. “Ms. Machiavelli and the Stable Matching Problem,” *American Mathematical Monthly* **92**, 261-268 (1985).
- [15] O. Kesten. “An Advice to the Organizers of Entry-Level Labor Markets in the United Kingdom,” mimeo, University of Rochester (2004).
- [16] J. Ma. “Stable Matchings and Rematching-Proof Equilibria in a Two-Sided Matching Market,” *Journal of Economic Theory* **66**, 352–369 (1995).
- [17] J. Ma. “Stable Matchings and the Small Core in Nash Equilibrium in the College Admissions Problem,” *Review of Economic Design* **7**, 117-134 (2002).
- [18] D. Majumdar. “Ordinally Bayesian Incentive-Compatible Stable Matchings,” mimeo, CORE (2003).
- [19] D. Majumdar and A. Sen. “Ordinally Bayesian Incentive-Compatible Voting Rules,” *Econometrica* **72**, 523-540 (2004).
- [20] 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).
- [21] J. Pais. *Incentives in Random Matching Markets*. Ph.D. Thesis, Universitat Autònoma de Barcelona (2005).
- [22] J. Pais. “Random Matching in the College Admissions Problem,” *Economic Theory* **35**, 99-116 (2008).
- [23] A. Romero-Medina. “Implementation of Stable Solutions in a Restricted Matching Market,” *Review of Economic Design* **3**, 137-147 (1998).
- [24] A.E. Roth. “The Economics of Matching: Stability and Incentives,” *Mathematics of Operations Research* **7**, 617-628 (1982).

- [25] 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).
- [26] A.E. Roth. “Misrepresentation and Stability in the Marriage Problem,” *Journal of Economic Theory* **34**, 383-387 (1984b).
- [27] A.E. Roth. “The College Admissions Problem is not Equivalent to the Marriage Problem,” *Journal of Economic Theory* **36**, 277-288 (1985a).
- [28] A.E. Roth. “Common and Conflicting Interests in Two-Sided Matching Markets,” *European Economic Review* **27**, 75-96 (1985b).
- [29] A.E. Roth. “Two-Sided Matching with Incomplete Information about Others’ Preferences,” *Games and Economic Behavior* **1**, 191-209 (1989).
- [30] 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).
- [31] A.E. Roth. “The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics,” *Econometrica* **70**, 1341-1378 (2002).
- [32] 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).
- [33] A.E. Roth and M. Sotomayor. “The College Admissions Problem Revisited,” *Econometrica* **57**, 559-570 (1989).
- [34] 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).

- [35] S. Shi. “Directed Search Model of Inequality with Heterogenous Skills and Skill-Biased Technology,” *Review of Economic Studies* **69**, 467-491 (2002).
- [36] S. Shin and S.-C. Suh. “A Mechanism Implementing the Stable Rule in Marriage Problems,” *Economics Letters* **51**, 185-189 (1996).
- [37] T. Sönmez. “Games of Manipulation in Marriage Problems,” *Games and Economic Behaviour* **20**, 169-176 (1997).
- [38] 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 prior, 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 in the stable mechanism φ under complete information P .*

Proof. Let \tilde{P} be a common prior, 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 in the mechanism φ under complete information P . 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 $P_v^* \in \text{resp}(P_v)$ (if $v \in F$) and $\varphi[s(P)](v)R_v\varphi[P'_v, s_{-v}(P_{-v})](v)$ (if $v \in W$). Hence,

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

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

$$\varphi[s(P)](v) \subseteq A(P_v) \text{ for all } v \in V. \quad (6)$$

If for some P in the support of \tilde{P} and for some $v \in V$, $\varphi[s(P)](v) \not\subseteq A(P_v)$, then choose $P'_v \in \mathcal{P}_v$ such that $A(P'_v) = A(P_v) \cap A(s_v(P_v))$ and $P'_v|A(P'_v) = s_v(P_v)|A(P'_v)$. Note that in the domain \mathcal{P}_v agent v can report any subset of his partners as acceptable and can rank his acceptable partners in any arbitrary linear order. Therefore, the existence of P'_v is guaranteed.

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 $\varphi[s(P)](v) \not\subseteq A(P_v)$, $\varphi[P'_v, s_{-v}(P_{-v})](v)R_v^*\emptyset P_v^*\varphi[s(P)](v)$. 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, (6) 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 in the mechanism φ under complete information P . Then,

without loss of generality, there exist $f \in F$, $P'_f \in \mathcal{P}_f$, and a responsive extension P_f^* of P_f such that

$$\varphi[P'_f, s_{-f}(P_{-f})](f)P_f^*\varphi[s(P)](f). \quad (7)$$

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 $\mu'(f) = \{w'_1, w'_2, \dots, w'_{|\mu'(f)|}\}$ where $w'_1 P_f w'_2 P_f \dots P_f w'_{|\mu'(f)|}$ and $\mu(f) = \{w_1, w_2, \dots, w_{|\mu(f)|}\}$ where $w_1 P_f w_2 P_f \dots P_f w_{|\mu(f)|}$. We now construct from P'_f another deviation P''_f and from $\mu'(f)$ both a responsive extension P_f^{**} of P_f and a subset of workers W^* , and prove that the random matching $\varphi[s_f(P_f), s_{-f}(\tilde{P}_{-f}|P_f)]$ does not first-order stochastically P_f -dominate the random matching $\varphi[P''_f, s_{-f}(\tilde{P}_{-f}|P_f)]$ since $\Pr\{\varphi[P''_f, s_{-f}(\tilde{P}_{-f}|P_f)](f) \in B(W^*, P_f^{**})\} > \Pr\{\varphi[s_f(P_f), s_{-f}(\tilde{P}_{-f}|P_f)](f) \in B(W^*, P_f^{**})\}$. We proceed by distinguishing between two mutually exclusive cases.

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

Note that $w'_k \in A(P_f)$ because $w'_k P_f w_k$ and by (6), $w_k \in \mu(f) \subseteq A(P_f)$. 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 $(F, W \setminus \{w'_{k+1}, \dots, w'_{|\mu'(f)|}\}, (k, q-f), (P'_f, s_{-f} \cup \{w'_{k+1}, \dots, w'_{|\mu'(f)|}\})(P_{-f} \cup \{w'_{k+1}, \dots, w'_{|\mu'(f)|}\}))$ such that $\mu''(f) = \{w'_1, \dots, w'_k\}$ and $\mu''(f') = \mu'(f')$ for all $f' \in F \setminus \{f\}$. Then from $\mu' \in C(P'_f, s_{-f}(P_{-f}); q)$ it follows that

$$\mu'' \in C(P'_f, s_{-f} \cup \{w'_{k+1}, \dots, w'_{|\mu'(f)|}\})(P_{-f} \cup \{w'_{k+1}, \dots, w'_{|\mu'(f)|}\}); k, q-f). \quad (8)$$

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 (8),

$$\mu'' \in C(P_f'', s_{-\{f\} \cup \{w'_{k+1}, \dots, w'_{|\mu'(f)|}\}}(P_{-\{f\} \cup \{w'_{k+1}, \dots, w'_{|\mu'(f)|}\}}); k, q_{-f}). \quad (9)$$

Thus, by $\mu''(f) = \{w'_1, \dots, w'_k\}$ 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 $C(P_f'', s_{-\{f\} \cup \{w'_{k+1}, \dots, w'_{|\mu'(f)|}\}}(P_{-\{f\} \cup \{w'_{k+1}, \dots, w'_{|\mu'(f)|}\}}); k, q_{-f})$. Now if firm f is matched to fewer than k workers in some matching belonging to $C(P_f'', s_{-\{f\} \cup \{w'_{k+1}, \dots, w'_{|\mu'(f)|}\}}(P_{-\{f\} \cup \{w'_{k+1}, \dots, w'_{|\mu'(f)|}\}}); q)$, then this matching is also stable for the problem $(P_f'', s_{-\{f\} \cup \{w'_{k+1}, \dots, w'_{|\mu'(f)|}\}}(P_{-\{f\} \cup \{w'_{k+1}, \dots, w'_{|\mu'(f)|}\}}); k, q_{-f})$, a contradiction to the previous fact. Hence, f is matched to at least k workers in any stable matching belonging to $C(P_f'', s_{-\{f\} \cup \{w'_{k+1}, \dots, w'_{|\mu'(f)|}\}}(P_{-\{f\} \cup \{w'_{k+1}, \dots, w'_{|\mu'(f)|}\}}); q)$. 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 $C(P_f'', s_{-f}(P_{-f}); q)$ to this matching. Since at least k copies of f are matched to a worker under the worker optimal matching in $C(P_f'', s_{-\{f\} \cup \{w'_{k+1}, \dots, w'_{|\mu'(f)|}\}}(P_{-\{f\} \cup \{w'_{k+1}, \dots, w'_{|\mu'(f)|}\}}); q)$, 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 $w \in B(w'_k, P_f) \setminus W^*$ and all $w^* \in W^*$, $w P_f w^*$. Let P_f^{**} be the responsive extension of P_f be such that for all $W'' \in 2^W$, $W'' P_f^{**} W^*$ if and only if the following three conditions hold: (i) $W'' \subseteq A(P_f)$, (ii) $|W''| \geq k$, and (iii) if $W'' = \{w''_1, w''_2, \dots, w''_{|W''|}\}$ where $w''_1 P_f \cdots P_f w''_{|W''|}$ and $W^* = \{w^*_1, \dots, w^*_k\}$ where $w^*_1 P_f \cdots P_f w^*_k$, then $w''_l R_f w^*_l$ for all $l \in \{1, \dots, k\}$. 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 $\varphi[s(P)](f)$ vio-

lates (iii) and our choice of P_f^{**} and W^* yields

$$\varphi[P_f'', s_{-f}(P_{-f})](f)R_f^{**}W^*P_f^{**}\varphi[s(P)](f). \quad (10)$$

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 (10), 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

$$\varphi[s_f(P_f), s_{-f}(P'_{-f})](f) \cap B(w'_k, P_f) \text{ must contain at least } k \text{ workers.} \quad (11)$$

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 $\hat{\mu}(f) = \{\hat{w}_1, \dots, \hat{w}_{|\hat{\mu}(f)|}\}$ where $\hat{w}_1 P_f \cdots P_f \hat{w}_{|\hat{\mu}(f)|}$. By (11), $\hat{\mu}(f) \cap B(w'_k, P_f)$ 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 $(F, W \setminus \{\hat{w}_{k+1}, \dots, \hat{w}_{|\hat{\mu}(f)|}\}, (k, q_{-f}), (s_f(P_f), s_{-\{f\} \cup \{\hat{w}_{k+1}, \dots, \hat{w}_{|\hat{\mu}(f)|}\}}(P'_{-\{f\} \cup \{\hat{w}_{k+1}, \dots, \hat{w}_{|\hat{\mu}(f)|}\}}))$ such that $\hat{\mu}'(f) = \{\hat{w}_1, \dots, \hat{w}_k\}$ and $\hat{\mu}'(f') = \hat{\mu}(f')$ for all $f' \in F \setminus \{f\}$. Then, from $\hat{\mu} \in C(s_f(P_f), s_{-f}(P'_{-f}); q)$ it follows that

$$\hat{\mu}' \in C(s_f(P_f), s_{-\{f\} \cup \{\hat{w}_{k+1}, \dots, \hat{w}_{|\hat{\mu}(f)|}\}}(P'_{-\{f\} \cup \{\hat{w}_{k+1}, \dots, \hat{w}_{|\hat{\mu}(f)|}\}}); k, q_{-f}). \quad (12)$$

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 $\hat{\mu}' \in C(\hat{P}_f, s_{-\{f\} \cup \{\hat{w}_{k+1}, \dots, \hat{w}_{|\hat{\mu}(f)|}\}}(P'_{-\{f\} \cup \{\hat{w}_{k+1}, \dots, \hat{w}_{|\hat{\mu}(f)|}\}}); k, q_{-f})$ (otherwise there would

exist a blocking pair for $\hat{\mu}'$;³⁰ then by (12) 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, $w \in B(\hat{w}, s_f(P_f)) \setminus \hat{\mu}'(f)$ and (w, f) must also block $\hat{\mu}'$ under $(s_f(P_f), s_{-\{f\} \cup \{\hat{w}_{k+1}, \dots, \hat{w}_{|\hat{\mu}'(f)|}\}}(P'_{-\{f\} \cup \{\hat{w}_{k+1}, \dots, \hat{w}_{|\hat{\mu}'(f)|}\}}); k, q_{-f})$, a contradiction to (12).)

Thus, since $|\hat{\mu}'(f)| = k$, firm f is matched to k workers for all matchings belonging to $C(\hat{P}_f, s_{-\{f\} \cup \{\hat{w}_{k+1}, \dots, \hat{w}_{|\hat{\mu}'(f)|}\}}(P'_{-\{f\} \cup \{\hat{w}_{k+1}, \dots, \hat{w}_{|\hat{\mu}'(f)|}\}}); k, q_{-f})$. Now if firm f is matched to fewer than k workers for some $\tilde{\mu} \in C(\hat{P}_f, s_{-\{f\} \cup \{\hat{w}_{k+1}, \dots, \hat{w}_{|\tilde{\mu}(f)|}\}}(P'_{-\{f\} \cup \{\hat{w}_{k+1}, \dots, \hat{w}_{|\tilde{\mu}(f)|}\}}); q)$, then $\tilde{\mu}$ is also stable under $(\hat{P}_f, s_{-\{f\} \cup \{\hat{w}_{k+1}, \dots, \hat{w}_{|\tilde{\mu}(f)|}\}}(P'_{-\{f\} \cup \{\hat{w}_{k+1}, \dots, \hat{w}_{|\tilde{\mu}(f)|}\}}); k, q_{-f})$, a contradiction to the previous fact. Hence, f is matched to at least k workers in any stable matching belonging to $C(\hat{P}_f, s_{-\{f\} \cup \{\hat{w}_{k+1}, \dots, \hat{w}_{|\hat{\mu}'(f)|}\}}(P'_{-\{f\} \cup \{\hat{w}_{k+1}, \dots, \hat{w}_{|\hat{\mu}'(f)|}\}}); q)$. 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 $C(\hat{P}_f, s_{-\{f\} \cup \{\hat{w}_{k+1}, \dots, \hat{w}_{|\hat{\mu}'(f)|}\}}(P'_{-\{f\} \cup \{\hat{w}_{k+1}, \dots, \hat{w}_{|\hat{\mu}'(f)|}\}}); q)$,

at least k copies of f are matched to a worker in any matching in $C(\hat{P}_f, s_{-f}(P'_{-f}); q)$. (13)

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 $(F, W \setminus (\tilde{\mu}(f) \setminus A(\hat{P}_f)), q, (P''_f, s_{-\{f\} \cup (\tilde{\mu}(f) \setminus A(\hat{P}_f))}(P'_{-\{f\} \cup (\tilde{\mu}(f) \setminus A(\hat{P}_f))})))$ such that $\tilde{\mu}'(f) = \tilde{\mu}(f) \cap A(\hat{P}_f)$ and $\tilde{\mu}'(f') = \tilde{\mu}(f')$ for all $f' \in F \setminus \{f\}$. 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 $\tilde{\mu}' \in C(P''_f, s_{-\{f\} \cup (\tilde{\mu}(f) \setminus A(\hat{P}_f))}(P'_{-\{f\} \cup (\tilde{\mu}(f) \setminus A(\hat{P}_f))}); q)$. 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 $\tilde{\mu}' \in C(\hat{P}_f, s_{-\{f\} \cup (\tilde{\mu}(f) \setminus A(\hat{P}_f))}(P'_{-\{f\} \cup (\tilde{\mu}(f) \setminus A(\hat{P}_f))}); q)$. 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 consider-

³⁰Note 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}).

ing the worker optimal matching in this core, we may split each firm $f' \in F \setminus \{f\}$ 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 $f' \in F \setminus \{f\}$ receive at least the same number of workers in $C(\hat{P}_f, s_{-f}(P'_{-f}); q)$ as they did previously. Since exactly $|\tilde{\mu}(f) \setminus A(\hat{P}_f)|$ 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 (13) 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 $l \in \{1, \dots, \min\{|\mu(f)|, |\mu'(f)|\}\}$. 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 (7). 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,

$$\text{firm } f \text{ is matched to } k \text{ workers under any matching in } C(P''_f, s_{-f}(P_{-f}); q). \quad (14)$$

On the other hand, let μ'' be the matching for the problem $(F, W \setminus \{w'_{k+2}, \dots, w'_{|\mu'(f)|}\}, (k+1, q_{-f}), (P''_f, s_{-\{f\} \cup \{w'_{k+2}, \dots, w'_{|\mu'(f)|}\}}(P_{-\{f\} \cup \{w'_{k+2}, \dots, w'_{|\mu'(f)|}\}})))$ such that $\mu''(f) = \{w'_1, \dots, w'_{k+1}\}$ and $\mu''(f') = \mu'(f')$ for all $f' \in F \setminus \{f\}$. Then from $\mu' \in C(P'_f, s_{-f}(P_{-f}); q)$ it follows that $\mu'' \in C(P'_f, s_{-\{f\} \cup \{w'_{k+2}, \dots, w'_{|\mu'(f)|}\}}(P_{-\{f\} \cup \{w'_{k+2}, \dots, w'_{|\mu'(f)|}\}}); k+1, q_{-f})$. Thus, by $\mu''(f) \subseteq B(w'_{k+1}, P_f) = A(P''_f)$ and $P''_f|A(P''_f) = P_f|A(P''_f)$, $\mu'' \in C(P''_f, s_{-\{f\} \cup \{w'_{k+2}, \dots, w'_{|\mu'(f)|}\}}(P_{-\{f\} \cup \{w'_{k+2}, \dots, w'_{|\mu'(f)|}\}}); k+1, q_{-f})$. Now if firm f is matched to fewer than $k+1$ workers in some matching belonging to $C(P''_f, s_{-\{f\} \cup \{w'_{k+2}, \dots, w'_{|\mu'(f)|}\}}(P_{-\{f\} \cup \{w'_{k+2}, \dots, w'_{|\mu'(f)|}\}}); q)$, then this matching is also stable for the problem $(P''_f, s_{-\{f\} \cup \{w'_{k+2}, \dots, w'_{|\mu'(f)|}\}}(P_{-\{f\} \cup \{w'_{k+2}, \dots, w'_{|\mu'(f)|}\}}); k+1, q_{-f})$, a contra-

diction to the previous fact. Hence, f is matched to at least $k + 1$ workers in any stable matching belonging to $C(P_f'', s_{-\{f\} \cup \{w'_{k+2}, \dots, w'_{|\mu'(f)|}\}}(P_{-\{f\} \cup \{w'_{k+2}, \dots, w'_{|\mu'(f)|}\}}); q)$. 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 $C(P_f'', s_{-\{f\} \cup \{w'_{k+2}, \dots, w'_{|\mu'(f)|}\}}(P_{-\{f\} \cup \{w'_{k+2}, \dots, w'_{|\mu'(f)|}\}}); q)$, 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 (14) 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