Robust Design in Monotonic Matching Markets: A Case for Firm-Proposing Deferred-Acceptance*

Lars Ehlers† Jordi Massó‡

May 2018

Abstract

We study two-sided matching markets among workers and firms. Workers seek one position at a firm but firms may employ several workers. In many applications those markets are monotonic: leaving positions unfilled is costly as for instance, for hospitals this means not being able to provide full service to its patients. A huge literature has advocated the use of stable mechanisms for clearinghouses. The interests among workers and firms are polarized among stable mechanisms, most famously the firm-proposing DA and the worker-proposing DA. We show that for the firm-proposing DA ex-ante incentive compatibility and ex-post incentive compatibility are equivalent whereas this is not necessarily true for the worker-proposing DA. The firm-proposing DA turns out to be more robust than the worker-proposing DA under incomplete information when incentives of both sides of the market are important.

JEL classification number: C78; D81; J44.

Keywords: Many-to-one matching market; Stability; Incomplete information; Monotonic responsive extensions; Robust mechanism design.

*L. Ehlers acknowledges financial support from the FRQSC (Québec) and SSHRC (Canada). The work of J. Massó is partially supported by the Spanish Ministry of Economy and Competitiveness, through the Grant ECO2017-83534-P and the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2015-0563), and from the Generalitat de Catalunya, through Grant SGR2014-515. Part of this research was done while J. Massó was a visiting fellow of CIREQ, at the Université de Montréal.

†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, Universitat Autònoma de Barcelona, and Barcelona GSE. 08193 Cerdanyola del Vallès (Barcelona), Spain; e-mail: `jordi.masso@uab.es`

1 Introduction

Centralized many-to-one matching markets operate as follows: each participant submits a ranked list of potential partners and a mechanism matches firms and workers on the basis of the submitted ranked lists. The participants belong to two sides: the firms (colleges, hospitals, schools, etc.) and the workers (students, medical interns, children, etc.). In applications many of the successful mechanisms are stable.[1],[2] Most markets are monotonic meaning it is (very) costly to leave positions unfilled and a firm desires to fill as many positions as possible with acceptable workers. For instance, in medical markets for a certain speciality hospitals may not be able to provide full medical service in case a position is left vacant and in school choice funding of a school may depend on the number of students admitted.

Stability of a matching (in the sense that all agents are matched to acceptable partners and no unmatched pair of a firm and a worker prefer each other to their proposed partners) has been considered to be the main feature in order to survive.[3] This is puzzling because there exists no stable mechanism which is ex-post incentive compatible on the full domain of preferences (Roth, 1982). Therefore, an agent's (submitted) ranked lists of potential partners are not necessarily his true ones and the implemented matching may not be stable for the true profile.[4]

Complete information is unrealistic and we use the (ordinally) Bayesian approach where nature selects a preference profile according to a common belief. Since real-life matching markets require to report ranked lists and not their specific utility representations, we follow the ordinal setting. Probability distributions are evaluated according to the first-order stochastic dominance criterion. Then a mechanism is monotonic ordinal Bayesian incentive compatible (monotonic OBIC) iff truth-telling is a monotonic ordinal Bayesian Nash equilibrium (monotonic OBNE) for any realized preference profile.[5]

Our main result shows that for the firm-proposing deferred-acceptance mechanism (firm-proposing DA) monotonic OBIC is equivalent to ex-post incentive compatibility on the support of the common belief. Hence, Bayesian incentive compatibility of the firm-

[1] See Roth (1984a), Roth and Peranson (1999), and Roth (2002) for a careful description and analysis of the American entry-level medical market. Roth (1991), Kesten (2005), Ünver (2005), and Ehlers (2008) describe and analyze the equivalent UK markets.

[2] Abdulkadiroğlu and Sönmez (2003) introduce school choice which is further studied by Chen and Sönmez (2006), Abdulkadiroğlu, Pathak, and Roth (2009) and Abdulkadiroğlu, Che, and Yosuda (2011). For college admissions, see Roth and Sotomayor (1989, 1990) and recently Chen and Kesten (2017).

[3] See, for instance, Roth (1984a).

[4] The literature has studied intensively Nash equilibria of direct preference revelation games induced by different stable mechanisms under complete information. See Dubins and Freedman (1981), Roth (1982, 1984b, 1985), Shin and Suh (1996), Sönmez (1997), Ma (1995, 2002), and Alcalde (1996).

[5] This means that for every von Neumann Morgenstern (vNM)-utility function of an agent's preference ordering (his type), submitting the true ranking maximizes his expected utility in the direct preference revelation game induced by the common prior and the mechanism given that all other agents truth-tell. This notion was introduced by d'Aspremont and Peleg (1988). Majumdar and Sen (2004) use it to relax strategy-proofness in the Gibbard-Satterthwaite Theorem and Ehlers and Massó (2007, 2015) use it to study strategic behavior in matching markets.

proposing DA does not depend on exact probabilities of the common belief for profiles in the support (and is detail-free à la Wilson's doctrine). We show that this does not necessarily hold for the worker-proposing DA, and the firm-proposing DA is more robust than the worker-proposing DA when incentives of both sides of the market are taken into account.

The paper is organized as follows. Section 2 describes monotonic many-to-one matching markets with incomplete information. Section 3 states our main result, Theorem 1. It shows that Theorem 1 does not hold for the worker-proposing DA and that the firm-proposing DA is more robust. Section 4 concludes and the Appendix contains all proofs.

2 Monotonic Many-To-One Matching Markets

Let F denote the set of firms, W denote the set of workers, and $V \equiv F \cup W$ denote the set of agents. For each firm f, there is a maximum number $q_f \geq 1$ of workers that f may hire, f's *quota*. Let $q = (q_f)_{f \in F}$ denote the vector of quotas. Each worker w has a strict preference ordering P_w over $F \cup \{\emptyset\}$, where $\emptyset$ stands for being unmatched. Each firm f has a strict preference ordering P_f over $W \cup \{\emptyset\}$, where $\emptyset$ stands for leaving a position unfilled. A *profile* $P = (P_v)_{v \in V}$ is a list of preference orderings. Given $S \subseteq V$, we sometimes write (P_S, P_{-S}) instead of P. 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 $\mathcal{P}_{-v} = \times_{v' \in V \setminus \{v\}} \mathcal{P}_{v'}$. Let R_v denote the weak preference associated with P_v. Given $w \in W$, $P_w \in \mathcal{P}_w$, and $v \in F \cup \{\emptyset\}$, let $B(v, P_w)$ denote the *weak upper contour set* of P_w at v; *i.e.*, $B(v, P_w) = \{v' \in F \cup \{\emptyset\} \mid v' R_w v\}$. 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 $P_f|S$. Given P_f and $w, w' \in W$, let $P_f^{w \leftrightarrow w'}$ stand for f's ordering where w and w' switch positions in P_f.[6] For example, if $P_f : w_1 w_2 w_3 w_4 \emptyset$, then $P_f^{w_1 \leftrightarrow w_3} : w_3 w_2 w_1 w_4 \emptyset$.[7]

A *many-to-one matching market (or college admissions problem)* is a quadruple (F, W, q, P). Because F, W and q remain fixed, a problem is simply a profile $P \in \mathcal{P}$. If $q_f = 1$ for all $f \in F$, (F, W, q, P) is called a *one-to-one matching market*.

A *matching* is a function $\mu : V \to 2^V$ satisfying the following: (m1) for all $w \in W$, $\mu(w) \subseteq F$ and $|\mu(w)| \leq 1$; (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 $w \in \mu(f)$. We will write $\mu(w) = f$ instead of $\mu(w) = \{f\}$. If $\mu(w) = \emptyset$, we say that w is *unmatched* at μ. If $|\mu(f)| < q_f$, we say that f has $q_f - |\mu(f)|$ *unfilled* positions at μ. Let $\mathcal{M}$ denote the set of all matchings. Given $P \in \mathcal{P}$ and $\mu \in \mathcal{M}$, μ is *stable (at P)* if (s1) for all $v \in V$, $\mu(v) \subseteq A(P_v)$ (*individual rationality*); and (s2) there exists no pair $(w, f) \in W \times F$ such that $f P_w \mu(w)$ and either $[w P_f \emptyset$ and $|\mu(f)| < q_f]$ or $[w P_f w'$ for some $w' \in \mu(f)]$ (*pairwise stability*). Let $C(P)$ denote the set of stable

[6] Formally, (i) $P_f^{w \leftrightarrow w'}|W \cup \{\emptyset\} \setminus \{w, w'\} = P_f|W \cup \{\emptyset\} \setminus \{w, w'\}$, (ii) $w P_f w'$ iff $w' P_f^{w \leftrightarrow w'} w'$, (iii) for all $v \in W \cup \{\emptyset\} \setminus \{w, w'\}$, $v P_f w$ iff $v P_f^{w \leftrightarrow w'} w'$, and (iv) for all $v \in W \cup \{\emptyset\} \setminus \{w, w'\}$, $v P_f w'$ iff $v P_f^{w \leftrightarrow w'} w$.

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

matchings at P (or the core of P). A (direct) mechanism is a function $\varphi : \mathcal{P} \to \mathcal{M}$. A mechanism φ is *stable* if for all $P \in \mathcal{P}$, $\varphi[P]$ is stable at P. The most popular stable mechanisms are the *deferred-acceptance algorithms (DA)* (Gale and Shapley, 1962): the firm-proposing DA is denoted by DA_F and the worker-proposing DA is denoted by DA_W.

A mechanism matches each firm f to a *set* of workers, taking into account only f's preference ordering P_f over individual workers. To study firms' incentives, preference orderings of firms over individual workers have to be extended to preference orderings over subsets of workers. The preference extension P_f^* over 2^W is *monotonic responsive* to $P_f \in \mathcal{P}_f$ if for all $S \in 2^W$, all $w \in S$, and all $w' \notin S$: (r1) $S \cup \{w'\} P_f^* S \Leftrightarrow |S| < q_f$ and $w' P_f \emptyset$; (r2) $(S \backslash \{w\}) \cup \{w'\} P_f^* S \Leftrightarrow w' P_f w$; and (r3) $S P_f^* S'$ if $|S'| < |S| \leq q_f$ and $S \subseteq A(P_f)$. Let R_f^* denote the weak preference associated with P_f^* and $mresp(P_f)$ denote the set of all monotonic responsive extensions of P_f. Moreover, given $S \in 2^W$, let $B(S, P_f^*)$ be the *weak upper contour set of* P_f^* at S; *i.e.*, $B(S, P_f^*) = \{S' \in 2^W \mid S' R_f^* S\}$.

Any mechanism and any true profile define a direct (ordinal) preference revelation game under complete information for which we can define the natural (ordinal) notion of Nash equilibrium. We denote such a game by $(\mathcal{P}, \varphi, P)$ where P is the true profile, φ is a mechanism and any agent v's set of strategies is $\mathcal{P}_v$. Given a mechanism φ and $P, P' \in \mathcal{P}$, P' is a *monotonic Nash equilibrium (monotonic NE) in the mechanism* φ *under complete information* P if (n1) 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 (n2) for all $f \in F$ and all $P_f^* \in mresp(P_f)$, $\varphi[P'](f) R_f^* \varphi[\hat{P}_f, P'_{-f}](f)$ for all $\hat{P}_f \in \mathcal{P}_f$. Truth-telling is a monotonic NE in φ under P if P is a monotonic NE in φ under P.

A *common prior* is a probability distribution $\tilde{P}$ over $\mathcal{P}$. Given $P \in \mathcal{P}$, let $\Pr\{\tilde{P} = P\}$ denote the probability that $\tilde{P}$ assigns to P. Given $v \in V$, let $\tilde{P}_v$ denote the marginal distribution of $\tilde{P}$ over $\mathcal{P}_v$. Given a common prior $\tilde{P}$ and $P_v \in \mathcal{P}_v$, 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 (Bayesian) uncertainty about the preferences of the other agents, given that v's preference ordering is P_v.[8]

A random matching $\tilde{\eta}$ is a probability distribution over the set of matchings $\mathcal{M}$. Given $\mu \in \mathcal{M}$, let $\Pr\{\tilde{\eta} = \mu\}$ denote the probability that $\tilde{\eta}$ assigns to μ. Let $\tilde{\eta}(w)$ denote the distribution which $\tilde{\eta}$ induces over w's set of potential partners $F \cup \{\emptyset\}$, and let $\tilde{\eta}(f)$ denote the distribution which $\tilde{\eta}$ induces over f's set of potential partners 2^W. Given two random matchings $\tilde{\eta}$ and $\tilde{\eta}'$, (fo1) for $w \in W$ and $P_w \in \mathcal{P}_w$ we say that $\tilde{\eta}(w)$ *first-order stochastically* P_w*−dominates* $\tilde{\eta}'(w)$, denoted by $\tilde{\eta}(w) P_w^{sd} \tilde{\eta}'(w)$, if for all $v \in F \cup \{\emptyset\}$, $\sum_{v' \in F \cup \{\emptyset\} : v' R_w v} \Pr\{\tilde{\eta}(w) = v'\} \geq \sum_{v' \in F \cup \{\emptyset\} : v' R_w v} \Pr\{\tilde{\eta}'(w) = v'\}$; and (fo2) for $f \in F$ and $P_f \in \mathcal{P}_f$, $\tilde{\eta}(f)$ *first-order stochastically* P_f*−dominates* $\tilde{\eta}'(f)$, denoted by $\tilde{\eta}(f) P_f^{sd} \tilde{\eta}'(f)$, if for all $P_f^* \in mresp(P_f)$ and all $S \in 2^W$, $\sum_{S' \in 2^W : S' R_f^* S} \Pr\{\tilde{\eta}(f) = S'\} \geq \sum_{S' \in 2^W : S' R_f^* S} \Pr\{\tilde{\eta}'(f) = S'\}$.[9]

---

[8]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.).

[9]It is well-known that (fo1) is equivalent to that for any vNM-representation of P_w the expected utility of $\tilde{\eta}$ is greater than or equal to the expected utility of $\tilde{\eta}'$ (and similarly for (fo2) and all vNM-representations

A mechanism φ and a common prior $\tilde{P}$ define a direct (ordinal) preference revelation game under incomplete information. Given a mechanism $\varphi : \mathcal{P} \to \mathcal{M}$ and a common prior $\tilde{P}$ over $\mathcal{P}$, truth-telling induces a random matching $\varphi[\tilde{P}]$ in the following way: for all $\mu \in \mathcal{M}$,

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

Using Bayesian updating, the relevant random matching for agent v, given his type P_v and report P'_v, is $\varphi[P'_v, \tilde{P}_{-v}|_{P_v}]$; namely $\Pr\{\varphi[P'_v, \tilde{P}_{-v}|_{P_v}] = \mu\} = \sum_{P_{-v} \in \mathcal{P}_{-v}: \varphi[P'_v, P_{-v}]=\mu} \Pr\{\tilde{P}_{-v}|_{P_v} = P_{-v}\}$.

Definition 1 (Monotonic OBIC) Let $\tilde{P}$ be a common prior. Then mechanism φ is *monotonic ordinal Bayesian incentive compatible (monotonic OBIC)* under incomplete information $\tilde{P}$ if and only if truth-telling is a *monotonic ordinal Bayesian Nash equilibrium (monotonic OBNE)* under incomplete information $\tilde{P}$, namely for all $v \in V$ and all $P_v \in \mathcal{P}_v$ such that $\Pr\{\tilde{P}_v = P_v\} > 0$,[10]

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

3 Ex-ante and Ex-post Incentive Compatibility

The support of a common prior $\tilde{P}$ is the set of profiles on which $\tilde{P}$ puts positive probability: $P \in \mathcal{P}$ belongs to the support of $\tilde{P}$ if and only if $\Pr\{\tilde{P} = P\} > 0$.

Our main result shows that for monotonic matching markets there is a strong link between complete and incomplete information for the firm-proposing DA.

Theorem 1. *Let $\tilde{P}$ be a common belief. Then DA_F is monotonic OBIC under incomplete information $\tilde{P}$ if and only if the support of $\tilde{P}$ is contained in the set of all profiles where truth-telling is a monotonic NE in DA_F under complete information.*

For DA_F to be ex-ante incentive compatible under a common belief, DA_F restricted to the support of the common belief must be ex-post incentive compatible. This in turn implies that exact probabilities of the common belief do not matter and any agent may have a private belief as long as its support is contained in the "common support". This means that for DA_F to be monotonic OBIC under common belief is robust subject to perturbations of the probabilities as along as the support(s) of the private beliefs coincide.

of any monotonic responsive extension of P_f). See for instance, Theorem 3.11 in d'Aspremont and Peleg (1988).

[10]In Definition 1 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.

3.1 Truth-telling under Stable Mechanisms

For one-to-one matching markets, Ehlers and Massó (2007, Theorem 1) implies that a stable mechanism is monotonic OBIC under $\tilde{P}$ if and only if the support of $\tilde{P}$ is contained in the set of profiles with singleton core. Theorem 1 implies that if DA_F is monotonic OBIC under common prior $\tilde{P}$, then the support of $\tilde{P}$ must be contained in the set of preference profiles having a singleton core.[11]

We show that for any stable mechanism to be monotonic OBIC, it is necessary for the common belief to put positive probability only on profiles with singleton core. In monotonic matching markets the core is unique for any realized (or observed) profile under ex-ante incentive compatibility.

Theorem 2. *Let $\tilde{P}$ be a common belief. If the stable mechanism φ is monotonic OBIC under incomplete information $\tilde{P}$, then the support of $\tilde{P}$ is contained in the set of all profiles with a singleton core.*

It is natural to ask whether the link in Theorem 1 breaks for stable mechanisms other than DA_F. We provide an example of a market in which our main result is not true.

Recall that Roth (1985) exhibits an example of a profile in a many-to-one matching market with a singleton core where a firm profitably manipulates any stable mechanism and truth-telling is not a monotonic NE under complete information. Below we exhibit an example of a many-to-one matching market where (i) DA_W is monotonic OBIC under the common belief (and any profile in the support has a singleton core) and (ii) there is a profile belonging to the support of the common belief at which truth-telling is not a monotonic NE under complete information for DA_W. Therefore, parallel to Roth (1985), the link in Theorem 1 is broken in one direction[12] when all firms have only monotonic responsive extensions and one firm has a capacity of at least two.

Example 1. *Consider a many-to-one matching market with three firms $F = \{f_1, f_2, f_3\}$ and four workers $W = \{w_1, w_2, w_3, w_4\}$. Firm f_1 has capacity $q_{f_1} = 2$ and firms f_2 and f_3 have capacity $q_{f_2} = q_{f_3} = 1$. Consider the common belief $\tilde{P}$ with* $\Pr\{\tilde{P} = P\} = p$ *and* $\Pr\{\tilde{P} = \bar{P}\} = 1 - p$*, where $p < 1/2$, and P and $\bar{P}$ are the following profiles:*

P_{f_1}	P_{f_2}	P_{f_3}	P_{w_1}	P_{w_2}	P_{w_3}	P_{w_4}	$\bar{P}_{f_1}$	$\bar{P}_{f_2}$	$\bar{P}_{f_3}$	$\bar{P}_{w_1}$	$\bar{P}_{w_2}$	$\bar{P}_{w_3}$	$\bar{P}_{w_4}$
w_1	w_1	w_3	f_3	f_2	f_1	f_1	w_1	w_2	w_4	f_1	f_2	f_1	f_3
w_2	w_2	w_1	f_1	f_1	f_3	f_2	w_2						
w_3	w_3	w_2	f_2	f_3	f_2	f_3	w_3						
w_4	w_4	w_4					w_4						

.

[11] If for some P belonging to the support of $\tilde{P}$, we have $|C(P)| \geq 2$, then there exist $\mu \in C(P)$ and $w \in W$ such that $\mu(w) P_w DA_F[P](w)$. Then $\mu(w) \neq w$. Let $P'_w \in \mathcal{P}_w$ be such that $A(P'_w) = \{\mu(w)\}$. Then $\mu \in C(P'_w, P_{-w})$ and by the fact that the set of unmatched workers is identical for any two stable matchings, $\mu(w) = DA_F[P'_w, P_{-w}](w) P_w DA_F[P](w)$. Hence, truth-telling is not a monotonic NE under P.

[12] It is clear that the other direction is true for any arbitrary game of incomplete information.

Note that $P_{f_1} = \bar{P}_{f_1}$*. It is straightforward to verify that both profiles have a singleton core and* $C(P) = \{\mu\}$ *and* $C(\bar{P}) = \{\bar{\mu}\}$*, where*

$$\mu = \begin{pmatrix} f_1 & f_2 & f_3 \\ \{w_3, w_4\} & \{w_2\} & \{w_1\} \end{pmatrix} \quad \text{and} \quad \bar{\mu} = \begin{pmatrix} f_1 & f_2 & f_3 \\ \{w_1, w_3\} & \{w_2\} & \{w_4\} \end{pmatrix}.$$

Let φ *be a stable mechanism. Thus, by stability of* φ, $\varphi[P] = \mu$ *and* $\varphi[\bar{P}] = \bar{\mu}$*. On the one hand, we will show that truth-telling is not a monotonic NE under complete information* P*. Let* $P'_{f_1} \in \mathcal{P}_{f_1}$ *be such that* $P'_{f_1} : w_1 w_2 w_4 \emptyset w_3$*. Then* $C(P'_{f_1}, P_{-f_1}) = \{\mu'\}$ *where*

$$\mu' = \begin{pmatrix} f_1 & f_2 & f_3 \\ \{w_1, w_4\} & \{w_2\} & \{w_3\} \end{pmatrix}.$$

Hence, by stability of φ, $\varphi[P'_{f_1}, P_{-f_1}] = \mu'$*. Obviously, for all monotonic responsive extensions* $P^*_{f_1}$ *of* P_{f_1} *we have* $\{w_1, w_4\} P^*_{f_1} \{w_3, w_4\}$*, which is equivalent to* $\varphi[P'_{f_1}, P_{-f_1}](f_1) P^*_{f_1} \varphi[P](f_1)$*. Therefore, truth-telling is not a monotonic NE in any stable mechanism* φ *under complete information* P *(and profile* P *belongs to the support of the common belief* $\tilde{P}$*).*

On the other hand, we will show that DA_W *is monotonic OBIC under incomplete information* $\tilde{P}$*. Note that for all* $v \in V \setminus \{f_1\}$*, if* v *observes his preference relation, then* v *knows whether* P *was realized or* $\bar{P}$ *was realized. Since at both of* P *and* $\bar{P}$ *the core is a singleton and firms* f_2 *and* f_3 *have quota one, it follows that* v *cannot gain by a deviation.*

Next consider firm f_1*. All arguments except for the last one apply to any stable mechanism* φ*. Observe that* $P_{f_1} = \bar{P}_{f_1}$ *and the random matching* $\varphi[P_{f_1}, \tilde{P}_{-f_1}|_{P_{f_1}}]$ *assigns to* f_1 *the set* $\{w_3, w_4\}$ *with probability* p *and the set* $\{w_1, w_3\}$ *with probability* $1 - p$*. Let* $P^*_{f_1} \in mresp(\mathrm{P}_{f_1})$ *and* $P''_{f_1} \in \mathcal{P}_{f_1}$ *be arbitrary. We show*

$$\varphi[P_{f_1}, \tilde{P}_{-f_1}|_{P_{f_1}}](f_1) P^{sd}_{f_1} \varphi[P''_{f_1}, \tilde{P}_{-f_1}|_{P_{f_1}}](f_1). \tag{2}$$

We distinguish two cases. First, suppose that $|\varphi[P''_{f_1}, P_{-f_1}](f_1)| = 1$ *or* $|\varphi[P''_{f_1}, \bar{P}_{-f_1}](f_1)| = 1$*. Now if (2) does not hold, then by monotonicity of* $P^*_{f_1}$ *and the fact that when submitting* P_{f_1}*,* f_1 *is matched to the set* $\{w_3, w_4\}$ *with probability* p *and to the set* $\{w_1, w_3\}$ *with probability* $1 - p$ *(where* $1 - p > 1/2$*), we must have that* $\varphi[P''_{f_1}, P_{-f_1}](f_1) P^*_{f_1} \{w_1, w_3\}$ *or* $\varphi[P''_{f_1}, \bar{P}_{-f_1}](f_1) P^*_{f_1} \{w_1, w_3\}$*. Obviously, from the definition of* $\bar{P}_{-f_1}$*, the last is impossible. Thus,* $\varphi[P''_{f_1}, P_{-f_1}](f_1) P^*_{f_1} \{w_1, w_3\}$ *and, by responsiveness of* $P^*_{f_1}$*, we must have* $\varphi[P''_{f_1}, P_{-f_1}](f_1) = \{w_1, w_2\}$*. But then, without loss of generality, we would have* $DA_F[P''_{f_1}, P_{-f_1}](f_1) = \{w_1, w_2\}$ *(because* DA_F *chooses the most preferred stable matching from the firms' point of view).*[13] *Since we have* $C(P) = \{\mu\}$ *and* $DA_F[P_{f_1}, P_{-f_1}](f_1) = \{w_3, w_4\}$*, this would imply that in the corresponding one-to-one matching problem* DA_F *is group manipulable by the two copies of* f_1 *(with each copy gaining strictly), a contradiction to the result of Dubins and Freedman (1981).*[14]

[13] If $DA_F[P''_{f_1}, P_{-f_1}](f_1) \neq \{w_1, w_2\}$, then choose $P'''_{f_1} \in \mathcal{P}_{f_1}$ such that $A(P'''_{f_1}) = \{w_1, w_2\}$. Then we obtain $DA_F[P'''_{f_1}, P_{-f_1}](f_1) = \{w_1, w_2\}$.

[14] Their result says that in a marriage market no group of firms can profitably manipulate DA_F at the true profile under complete information (with strict preference holding for all firms belonging to the group).

Second, suppose that $|\varphi[P''_{f_1}, P_{-f_1}](f_1)| = 2$ *and* $|\varphi[P''_{f_1}, \bar{P}_{-f_1}](f_1)| = 2$. *Then by definition of* $\bar{P}_{-f_1}$ *and* $|\varphi[P''_{f_1}, \bar{P}_{-f_1}](f_1)| = 2$, *we must have* $\varphi[P''_{f_1}, \bar{P}_{-f_1}](f_1) = \{w_1, w_3\}$. *Thus, by stability of* φ, $\{w_1, w_3\} \subseteq A(P''_{f_1})$.

If for all $\mu'' \in C(P''_{f_1}, P_{-f_1})$, $\mu''(w_4) = \emptyset$, *then* $w_4 \notin A(P''_{f_1})$. *By definition of* P_{-f_1} *and* $w_3 \in A(P''_{f_1})$, $DA_W[P''_{f_1}, P_{-f_1}](f_1) = \{w_3\}$. *Then* f_1 *does not fill all its positions at the worker-optimal matching and by Roth and Sotomayor (1990),* f_1 *is matched to the same set of workers at all stable matchings. Hence,* $\varphi[P''_{f_1}, P_{-f_1}](f_1) = \{w_3\}$ *and (2) holds. To see that observe first that when submitting* P''_{f_1}, *firm* f_1 *is matched with probability* p *to* $\{w_3\}$ *and with probability* $1-p$ *to* $\{w_1, w_3\}$ *while when submitting* P_{f_1}, *firm* f_1 *is matched with probability* p *to* $\{w_3, w_4\}$ *and with probability* $1-p$ *to* $\{w_1, w_3\}$. *Since* $\{w_3, w_4\} P^*_{f_1} \{w_3\}$ *for all monotonic responsive extensions of* P_{f_1}, *(2) holds.*

If for some $\mu'' \in C(P''_{f_1}, P_{-f_1})$, $\mu''(w_4) \neq \emptyset$, *then by definition of* P_{-f_1}, $\mu''(w_4) = f_1$; *otherwise the pair* (w_2, f_2) *would block* μ'' *at* (P''_{f_1}, P_{-f_1}) *if* $\mu''(w_4) = f_2$ *and the pair* (w_1, f_3) *would block* μ'' *at* (P''_{f_1}, P_{-f_1}) *if* $\mu''(w_4) = f_3$. *Thus, by* $\{w_3, w_4\} \subseteq A(P''_{f_1})$, $\mu \in C(P''_{f_1}, P_{-f_1})$ *and* $DA_W[P''_{f_1}, P_{-f_1}] = \mu$. *Therefore, the probability distribution* $DA_W[P_{f_1}, \tilde{P}_{-f_1} \mid_{P_{f_1}}](f_1)$ *and* $DA_W[P''_{f_1}, \tilde{P}_{-f_1} \mid_{P_{f_1}}](f_1)$ *coincide with the distribution in which* f_1 *is matched to* $\{w_3, w_4\}$ *with probability* p *and to* $\{w_1, w_3\}$ *with probability* $1-p$. *Hence, (2) holds for* DA_W. □

Observe that in Example 1 DA_F is not monotonic OBIC under incomplete information $\tilde{P}$: for the preference $P''_{f_1} : w_1 w_2 w_4 w_3 \emptyset$ in Example 1 we have

$$DA_F[P''_{f_1}, P_{-f_1}] = \begin{pmatrix} f_1 & f_2 & f_3 \\ \{w_1, w_4\} & \{w_2\} & \{w_3\} \end{pmatrix} \text{ and } DA_F[P''_{f_1}, \bar{P}_{-f_1}] = \begin{pmatrix} f_1 & f_2 & f_3 \\ \{w_1, w_3\} & \{w_2\} & \{w_4\} \end{pmatrix}.$$

Thus,

$$\Pr\{DA_F[P''_{f_1}, \tilde{P}_{-f_1}|_{P_{f_1}}](f_1) \in B(\{w_1, w_4\}, P^*_{f_1})\} = 1$$
$$\Pr\{DA_F[P_{f_1}, \tilde{P}_{-f_1}|_{P_{f_1}}](f_1) \in B(\{w_1, w_4\}, P^*_{f_1})\} = 1 - p$$

for all monotonic responsive extensions $P^*_{f_1}$ of P_{f_1} (i.e. this is an unambiguous deviation for firm f_1 because it does not depend on the choice of the responsive extension of P_{f_1}). Thus, we do not have $DA_F[P_{f_1}, \tilde{P}_{-f_1}|_{P_{f_1}}](f_1) P^{sd}_{f_1} DA_F[P''_{f_1}, \tilde{P}_{-f_1}|_{P_{f_1}}](f_1)$ and DA_F is not monotonic OBIC under $\tilde{P}$. Roughly speaking, in Example 1 firm f_1 manipulates DA_F by reordering its acceptable workers whereas this is not possible for DA_W (when keeping the same set of acceptable workers).

3.2 Robust Design

For monotonic matching markets we establish an equivalence result for (payoff) type spaces on robust mechanism design à la Bergemann and Morris (2005).[15] Instead of the terminology of payoff type spaces, in (ordinal) matching markets it is natural to use the term "preference type spaces".

[15] We refer the interested reader to Bergemann and Morris (2012) for a comprehensive introduction to robust mechanism design.

Let $\mathcal{Q} \subseteq \mathcal{P}$ denote a set of possible preference type profiles. Let $\mathcal{Q}_v = \{P_v | P \in \mathcal{Q}\}$ be the set of agent v's possible preference types in $\mathcal{Q}$. 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 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 $\mathcal{Q} = \times_{v \in V} \mathcal{Q}_v$. 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, we allow for non-product preference type spaces. In applications (such 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 a 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 $(\mathcal{Q}, \tilde{P})$. 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 $\varphi : \mathcal{P} \to \mathcal{M}$ is *interim incentive compatible on* $(\mathcal{Q}, (\hat{\pi}_v)_{v \in V})$ if for all $v \in V$ and all $P_v \in \mathcal{Q}_v$,

$$\varphi[P_v, \hat{\pi}_v(P_v)](v) P_v^{sd} \varphi[P'_v, \hat{\pi}_v(P_v)](v) \quad \text{for all } P'_v \in \mathcal{P}_v. \tag{3}$$

Note that interim incentive compatibility reduces to monotonic OBIC for common prior preference type spaces. Instead of defining ex-post incentive compatibility for all possible profiles (or all types), 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 $\varphi : \mathcal{P} \to \mathcal{M}$ 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 $P^*_f \in mresp(P_f)$, $\varphi[P_f, P_{-f}](f) R^*_f \varphi[P'_f, P_{-f}](f)$ 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 1. *Let $\mathcal{Q} \subseteq \mathcal{P}$. Then the following are equivalent.*

(a) DA_F is interim incentive compatible on all preference type spaces $(\mathcal{Q}, (\hat{\pi}_v)_{v\in V})$.

(b) DA_F is interim incentive compatible on all common prior preference type spaces $(\mathcal{Q}, \tilde{P})$.

(c) DA_F is interim incentive compatible on ***one*** *common prior preference type space $(\mathcal{Q}, (\hat{\pi}_v)_{v\in V})$.*

(d) DA_F is ex-post incentive compatible on $\mathcal{Q}$.

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

Note that Corollary 1 differs in the following sense from Bergemann and Morris (2005, Corollary 1) for social choice functions and all payoff type spaces: if DA_F is interim incentive compatible for *one* common prior preference type space,[16] then DA_F is interim incentive compatible on *all* preference type spaces (with the same support). This is a strong robustness feature for DA_F (and for DA_W, Corollary 1 does not hold by Example 1). Two other important qualifications between Bergemann and Morris (2005, Corollary 1)[17] and Corollary 1 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. Again, non-product preference type spaces are especially important for matching markets.

4 Conclusion

Ehlers and Massó (2015) study matching markets where all responsive extensions are allowed (i.e. where only (r1) and (r2) are imposed on the responsive preference extension).

[16]As one may check, instead we could have replaced (c) by DA_F is interim incentive compatible on one preference type space $(\mathcal{Q}, (\hat{\pi}_v)_{v\in V})$ such that for any $v \in V$ and $P_v \in \mathcal{Q}_v$, the support of $\hat{\pi}_v(P_v)$ coincides with $\mathcal{Q}_{-v}|_{P_v}$.

[17]Bergemann and Morris (2005) allow for larger type spaces which can be given here for $\mathcal{Q}$ by $(T_v, \hat{\theta}_v, \hat{\pi}_v)_{v\in V}$ where for each $v \in V$, T_v is a non-empty set, $\hat{\theta}_v : T_v \to \mathcal{Q}_v$ specifies for each $t_v \in T_v$ a payoff type $\hat{\theta}_v(t_v)$ in $\mathcal{Q}_v$ and a belief $\hat{\pi}_v(t_v) \in \Delta(T_{-v})$ such that for any t_{-v} in the support of $\hat{\pi}_v(t_v)$ we have $\hat{\theta}_{-v}(t_{-v}) \in \mathcal{Q}_{-v}|_{\hat{\theta}_v(t_v)}$. It is easy to see that in Corollary 1 for a fixed $\mathcal{Q}$, a stable mechanism φ is interim incentive compatible on all type spaces if and only if φ is interim incentive compatible on all preference type spaces. The reason is that preference type spaces are type spaces and larger type spaces can be replicated by a collection of preference type spaces.

From Ehlers and Massó (2015, Theorem 1) it follows that a stable mechanism φ is OBIC under incomplete information $\tilde{P}$ if and only if the support of $\tilde{P}$ is contained in the set of profiles where truth-telling is a NE under complete information (where all responsive extensions are considered). Ehlers and Massó (2015, Theorem 1) crucially depends on firms having responsive extensions which are not monotonic. Example 1 does not contradict their result: when considering the non-monotonic responsive extension $P^*_{f_1}$ such that $\{w_1\}P^*_{f_1}\{w_3, w_4\}$, then firm f_1 gains by submitting the list $\hat{P}_{f_1}$ where worker w_1 is the unique acceptable worker (*i.e.* $A(\hat{P}_{f_1}) = \{w_1\}$). Then, since $C(\hat{P}_f, P_{-f}) = C(\hat{P}_f, \bar{P}_{-f}) = \{\hat{\mu}\}$ and $\hat{\mu}(f_1) = \{w_1\}$, we have both $\varphi[\hat{P}_{f_1}, P_{-f_1}](f_1) = \{w_1\}$ and $\varphi[\hat{P}_{f_1}, \bar{P}_{-f_1}](f_1) = \{w_1\}$, which means that no stable mechanism φ is OBIC under incomplete information $\tilde{P}$.

Often in real-life matching markets it is costly to leave positions unfilled and responsive extensions are necessarily monotonic. For instance, in medical markets for a certain speciality hospitals may not be able to provide full medical service in case a position is left vacant. Once we focus on monotonic OBIC, the firm-proposing DA turns out to be more robust than the workers-proposing DA when incentives of both sides of the market are important.

APPENDIX.

A Proof of Theorem 1

Let $\tilde{P}$ be a common belief.

($\Leftarrow$) Suppose that for all profiles in the support of $\tilde{P}$, truth-telling is a monotonic NE in DA_F under complete information. Let P be such that $\Pr\{\tilde{P} = P\} > 0$. Then, for all $w \in W$ and all $P'_w \in \mathcal{P}_w$, $DA_F[P](w)R_w DA_F[P'_w, P_{-w}](w)$ and for all $f \in F$ and all $P'_f \in \mathcal{P}_f$, $DA_F[P](f)R^*_f DA_F[P'_f, P_{-f}](f)$ for all $P^*_f \in mresp(P_f)$. Hence, for all $w \in W$ and all $P_w \in \mathcal{P}_w$ such that $\Pr\{\tilde{P}_w = P_w\} > 0$, we have that for all $P'_w \in \mathcal{P}_w$,

$$DA_F[P_w, \tilde{P}_{-w}|_{P_w}](w)P^{sd}_w DA_F[P'_w, \tilde{P}_{-w}|_{P_w})](w),$$

and for all $f \in F$ and all $P_f \in \mathcal{P}_f$ such that $\Pr\{\tilde{P}_f = P_f\} > 0$, we have that for all $P'_f \in \mathcal{P}_f$,

$$DA_F[P_f, \tilde{P}_{-f}|_{P_f}](f)P^{sd}_f DA_F[P'_f, \tilde{P}_{-f}|_{P_f})](f).$$

That is, P is a monotonic OBNE in DA_F under $\tilde{P}$, the desired conclusion.

($\Rightarrow$) Let DA_F be monotonic OBIC under incomplete information $\tilde{P}$. Let P be such that $\Pr\{\tilde{P} = P\} > 0$. By Theorem 2, $|C(P)| = 1$. To obtain a contradiction suppose that P is not a monotonic NE in DA_F under complete information P. We first show in Lemma 1 below that then some firm with quota of at least two has a profitable deviation such that (i) the firm is only matched to acceptable workers, (ii) the firm is matched to q_f workers and (iii) for some monotonic responsive extension the firm strictly prefers the assigned workers to the set received under truth-telling.

Lemma 1. *Let P be such that $|C(P)| = 1$. If P is not a monotonic NE in DA_F under complete information P, then there exists a firm $f \in F$ such that $q_f \geq 2$ and for some $\hat{P}_f$ we have*

(i) $DA_F[\hat{P}_f, P_{-f}](f) \subseteq A(P_f)$,

(ii) $|DA_F[\hat{P}_f, P_{-f}](f)| = q_f = |DA_F[P](f)|$, *and*

(iii) $DA_F[\hat{P}_f, P_{-f}(f) P^*_f DA_F[P](f)$ *for some* $P^*_f \in mresp(P_f)$.

Proof of Lemma 1. Let $P \in \mathcal{P}$ be such that $|C(P)| = 1$. Then, using a similar argument to the one used in the sufficiency proof of Ehlers and Massó (2007, Theorem 1), it can be seen that no worker has a profitable deviation from P in DA_F. The same argument applies to any firm with quota one. Thus, if P is not a monotonic NE in DA_F under complete information P, then for some $f \in F$ with $q_f \geq 2$ and $\hat{P}_f$, we have

$$DA_F[\hat{P}_f, P_{-f}](f) P^*_f DA_F[P](f) \tag{4}$$

for some $P^*_f \in mresp(P_f)$, which is (iii) of the lemma. Because P^*_f is responsive, $DA_F[\hat{P}_f, P_{-f}](f) \cap A(P_f) R^*_f DA_F[\hat{P}_f, P_{-f}](f)$. Thus, by (4) and monotonicity of P^*_f,

$$|DA_F[\hat{P}_f, P_{-f}](f) \cap A(P_f)| \geq |DA_F[P](f)|. \tag{5}$$

But then we may suppose w.l.o.g. that $DA_F[\hat{P}_f, P_{-f}](f) \subseteq A(P_f)$. To see that, assume $DA_F[\hat{P}_f, P_{-f}](f) \nsubseteq A(P_f)$. Then, choose any preference $P'_f \in \mathcal{P}_f$ such that $A(P'_f) = DA_F[\hat{P}_f, P_{-f}](f) \cap A(P_f)$. Now consider the matching market where the set of workers $\hat{W} = DA_F[\hat{P}_f, P_{-f}](f) \backslash A(P_f)$ is not present with profile $(P'_f, P_{-\hat{W} \cup \{f\}})$. Then, it is easy to see that $DA_F[P'_f, P_{-\hat{W} \cup \{f\}}](f) = A(P'_f)$. Now letting workers in $\hat{W}$ reenter the market, firms should weakly gain meaning that $DA_F[P'_f, P_{-f}](f) = A(P'_f)$.[18] Since $DA_F[P'_f, P_{-f}](f) P^*_f DA_F[P](f)$, f has a profitable deviation from P in DA_F where f is only matched to acceptable workers. Hence, we can replace in (4) $\hat{P}_f$ by P'_f. Thus, $DA_F[\hat{P}_f, P_{-f}](f) \subseteq A(P_f)$ holds. But the inclusion is (i) of the lemma.

Next we show that $|DA_F[\hat{P}_f, P_{-f}](f)| = |DA_F[P](f)|$. To obtain a contradiction, assume otherwise. By (5) and (i), we have $|DA_F[\hat{P}_f, P_{-f}](f)| > |DA_F[P](f)|$. Then at $DA_F[P]$ firm f has some positions unfilled and the ranking of P_f over $A(P_f)$ is irrelevant for the stability of $DA_F[P]$ under P. Let $k = |DA_F[\hat{P}_f, P_{-f}](f)|$. In particular, $DA_F[P]$ is stable under P if firm f's quota is reduced from q_f to k. W.l.o.g. we may suppose that $A(\hat{P}_f) = DA_F[\hat{P}_f, P_{-f}](f)$. Now again in any matching which is stable under $(\hat{P}_f, P_{-f})$ firm f is matched to $DA_F[\hat{P}_f, P_{-f}](f)$. Again firm f's quota may be reduced to k. But now consider P'_f such that $A(P'_f) = A(P_f)$ and $DA_F[\hat{P}_f, P_{-f}](f)$ are the first k most

[18] See Theorem 2.25 in Roth and Sotomayor (1990).

preferred workers under P'_f. But then both $DA_F[\hat{P}_f, P_{-f}]$ and $DA_F[P]$ must be stable under (P'_f, P_{-f}), which is a contradiction because $|DA_F[\hat{P}_f, P_{-f}](f)| \neq |DA_F[P](f)|$ and any firm is matched to the same number of workers at all stable matchings under the profile (P'_f, P_{-f}).

We have shown $|DA_F[\hat{P}_f, P_{-f}](f)| = |DA_F[P](f)|$. Furthermore, if $|DA_F[P](f)| < q_f$, then, by (i), both matchings $DA_F[P]$ and $DA_F[\hat{P}_f, P_{-f}]$ are stable under P. By (4), $DA_F[\hat{P}_f, P_{-f}](f) \neq DA_F[P](f)$ which means that $|C(P)| \neq 1$, a contradiction. Hence, $|DA_F[P](f)| = q_f = |DA_F[\hat{P}_f, P_{-f}](f)|$ (which is (ii)). □

We now proceed with the proof of Theorem 1. Let $f \in F$, with $q_f \geq 2$, and $\hat{P}_f$ be the firm and its preferences identified in Lemma 1, for which (i), (ii) and (iii) hold. Letting $\hat{\mu} = DA_F[\hat{P}_f, P_{-f}]$ and $\mu = DA_F[P]$, order the workers in $\hat{\mu}(f)$ and $\mu(f)$ according to P_f: let $\hat{\mu}(f) = \{\hat{w}_1, \ldots, \hat{w}_{q_f}\}$ and $\mu(f) = \{w_1, \ldots, w_{q_f}\}$. Furthermore, because of (i)-(iii) and we are using DA_F, we may assume w.l.o.g. that

(a) $A(\hat{P}_f) = B(w_{q_f}, P_f) \cup B(\hat{w}_{q_f}, P_f)$,

(b) $\hat{P}_f|\hat{\mu}(f) = P_f|\hat{\mu}(f)$,

(c) $\hat{P}_f|B(\hat{w}_{q_f}, \hat{P}_f) = P_f|B(\hat{w}_{q_f}, \hat{P}_f)$,

(d) $\hat{P}_f|A(\hat{P}_f)\backslash B(\hat{w}_{q_f}, \hat{P}_f) = P_f|A(\hat{P}_f)\backslash B(\hat{w}_{q_f}, \hat{P}_f)$, and

(e) $\hat{w}\hat{P}_f w$ for all $\hat{w} \in B(\hat{w}_{q_f}, \hat{P}_f)$ and all $w \in A(\hat{P}_f)\backslash B(\hat{w}_{q_f}, \hat{P}_f)$.

Note that since $\hat{\mu}(f) \subseteq B(\hat{w}_{q_f}, \hat{P}_f)$, (c) implies (b), which in turn implies $\hat{\mu}(f) \subseteq B(\hat{w}_{q_f}, P_f)$.

We distinguish two cases: $\hat{w}_{q_f} P_f w_{q_f}$ and $w_{q_f} R_f \hat{w}_{q_f}$.

Case 1: $\hat{w}_{q_f} P_f w_{q_f}$.

Let $\bar{W}^*$ consist of the q_f workers which are P_f-least preferred in $B(w_{q_f}, P_f)\backslash\{w_{q_f}\}$. Then choose the monotonic responsive extension $\bar{P}^*_f$ of P_f such that $W'\bar{R}^*_f\bar{W}^*$ iff W' contains q_f workers in $B(w_{q_f}, P_f)\backslash\{w_{q_f}\}$. Let P''_f be such that $A(P''_f) = B(w_{q_f}, P_f)\backslash\{w_{q_f}\}$ and $P''_f|A(P''_f) = \hat{P}_f|A(P''_f)$. Then $\hat{\mu} \in C(P''_f, P_{-f})$ and by construction, $DA_F[P''_f, P_{-f}](f)$ contains q_f workers in $A(P''_f)$, but $DA_F[P](f)$ contains at most $q_f - 1$ workers in $A(P''_f)$. Hence, by definition of $\bar{W}^*$ and since $w_{q_f} \notin A(P''_f)$,

$$DA_F[P''_f, P_{-f}](f)\bar{R}^*_f\bar{W}^*\bar{P}^*_f DA_F[P_f, P_{-f}](f). \tag{6}$$

To obtain a contradiction with the fact that DA_F is monotonic OBIC under incomplete information $\tilde{P}$, we consider any profile P'_{-f} such that $\Pr\{\tilde{P} = (P_f, P'_{-f})\} > 0$ and restrict the attention to the upper contour set of $\bar{P}^*_f$ at $\bar{W}^*$. We want to show that if $DA_F[P_f, P'_{-f}](f)\bar{R}^*_f\bar{W}^*$, then $DA_F[P''_f, P'_{-f}](f)\bar{R}^*_f\bar{W}^*$. By the definitions of $\bar{W}^*$ and $\bar{P}^*_f$, $DA_F[P_f, P'_{-f}](f)\bar{R}^*_f\bar{W}^*$ implies that $DA_F[P_f, P'_{-f}](f)$ contains q_f workers in $A(P''_f)$. If

$DA_F[P''_f, P'_{-f}](f)$ contains at most $q_f - 1$ workers in $A(P''_f)$, then the ranking does not matter and the matching $DA_F[P''_f, P'_{-f}]$ is also stable under (P'''_f, P'_{-f}), where P'''_f is obtained by letting $A(P'''_f) = A(P''_f)$ and $P'''_f|A(P'''_f) = P_f|A(P'''_f)$. Furthermore, $DA_F[P_f, P'_{-f}]$ is stable under (P'''_f, P'_{-f}) and $\left|DA_F[P_f, P'_{-f}]\right| = q_f$, by definition of $\bar{W}^*$. But then f would be matched to different numbers of workers under different matchings belonging to $C(P'''_f, P'_{-f})$, a contradiction. Thus, $DA_F[P''_f, P'_{-f}](f)$ contains q_f workers in $A(P''_f)$. Now by our choice of $\bar{W}^*$ and $\bar{P}^*_f$, we obtain for any profile P'_{-f} in the support of $\tilde{P}_{-f}|_{P_f}$, $DA_F[P_f, P'_{-f}](f)\bar{R}^*_f\bar{W}^*$ implies $DA_F[P''_f, P'_{-f}](f)\bar{R}^*_f\bar{W}^*$. Then, together with (6) and $\Pr\{\tilde{P} = (P_f, P_{-f})\} > 0$, this implication means that DA_F is not monotonic OBIC under $\tilde{P}$ because not $DA_F[P_f, \tilde{P}_{-f}|_{P_f}](f)P^{sd}_{f1}DA_F[P''_f, \tilde{P}_{-f}|_{P_f}](f)$ since for the monotonic responsive extension $\bar{P}^*_f$ of P_f,

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

Case 2: $w_{q_f} R_f \hat{w}_{q_f}$.

Let k be the first index such that $\hat{w}_k P_f w_k$. By (iii) in Lemma 1, k exists. Note that $k < q_f$ and $\hat{w}_k P_f w_k R_f w_{q_f}$. Let P''_f be such that $A(P''_f) = B(\hat{w}_{q_f}, P_f)$, $B(\hat{w}_k, P_f) \cup \hat{\mu}(f) P''_f A(P''_f)\backslash(B(\hat{w}_k, P_f) \cup \hat{\mu}(f))$,[19] $P''_f|B(\hat{w}_k, P_f) \cup \hat{\mu}(f) = P_f|B(\hat{w}_k, P_f) \cup \hat{\mu}(f)$, and $P''_f|A(P''_f)\backslash(B(\hat{w}_k, P_f) \cup \hat{\mu}(f)) = P_f|A(P''_f)\backslash(B(\hat{w}_k, P_f) \cup \hat{\mu}(f))$. We illustrate P''_f below:

$$\begin{array}{c} P''_f \\ \hline P_f|B(\hat{w}_k, P_f) \\ P_f|\hat{\mu}(f)\backslash B(\hat{w}_k, P_f) \\ P_f|A(P''_f)\backslash(B(\hat{w}_k, P_f) \cup \hat{\mu}(f)) \end{array},$$

i.e. P''_f ranks first all elements in $B(\hat{w}_k, P_f)$ according to P_f, then all elements in $\hat{\mu}(f)\backslash B(\hat{w}_k, P_f)$ according to P_f, and then the remaining P''_f-acceptable workers according to P_f. Because $\hat{w}_1, \ldots, \hat{w}_k$ belong to $B(\hat{w}_k, P_f)$, we have $\hat{\mu}(f)\backslash B(\hat{w}_k, P_f) = \{\hat{w}_{k+1}, \ldots, \hat{w}_{q_f}\}$ (and this set consists of exactly $q_f - k$ workers).

If f is matched to fewer than q_f workers in any matching belonging to $C(P''_f, P_{-f})$, then the differences between ranking P''_f and P_f do not matter and f would be matched to fewer than q_f workers in $DA_F[P](f)$, the unique matching in $C(P)$, a contradiction. Thus, f is matched to q_f workers in any matching belonging to $C(P''_f, P_{-f})$.

If $\hat{\mu} \in C(P''_f, P_{-f})$, then by construction and (4), $DA_F[P''_f, P_{-f}](f)R^*_f\hat{\mu}(f)P^*_f\mu(f)$ implying

$$DA_F[P''_f, P_{-f}](f)P^*_f\mu(f). \tag{7}$$

Suppose that $\hat{\mu} \notin C(P''_f, P_{-f})$. Then, some $(w, \hat{f})$ blocks $\hat{\mu}$ under (P''_f, P_{-f}). By the definition of P''_f, $\hat{f} = f$ and $w \in B(\hat{w}_k, P_f)$. We show that $DA_F[P''_f, P_{-f}](f)$ contains at least k workers in $B(\hat{w}_k, P_f)$. We show this with the following two steps:

Step 1: Let w^1 be the P_f-highest ranked worker such that (w^1, f) forms a blocking pair of $\hat{\mu}$ under (P''_f, P_{-f}). Let $\hat{\mu}^0 = \hat{\mu}$. Then match w^1 to f and consider $\hat{\mu}^1$ such that (i)

[19]Here we use the convention SP''_fT iff sP''_ft for all $s \in S$ and all $t \in T$.

$\hat{\mu}^1(w^1) = f$ and (ii) $\hat{\mu}^1(w) = \hat{\mu}(w)$ for all $w \neq w^1$. In other words, f does not reject any worker and we allow f to have capacity $q_f + 1$ (*i.e.*, $\hat{\mu}^1(f) = \hat{\mu}^0(f) \cup \{w^1\}$). If $\hat{\mu}^1$ does not contain any blocking pair, set $\bar{\mu} = \hat{\mu}^1$ and $\bar{q}_f = q_f + 1$ and then go to Step 2. Otherwise, $\hat{\mu}^1$ contains a blocking pair, say (w^2, f^2). Again, let w^2 be the P_{f^2}-highest ranked worker such that (w^2, f^2) forms a blocking pair of $\hat{\mu}^1$. But then it must be that either (a) $f^2 = f$ (which would mean that $w^2 \notin \hat{\mu}^1(f)$) or (b) $f^2 \neq f$ and $w^2 \in \hat{\mu}^1(f)$ or (c) $f^2 \neq f$ and $w^2 \notin \hat{\mu}^1(f)$. Note that if $f^2 \neq f$, then we must have $f^2 = \hat{\mu}(w^1)$. Now consider $\hat{\mu}^2$ such that (i) $\hat{\mu}^2(w^2) = f^2$ and (ii) $\hat{\mu}^2(w) = \hat{\mu}^1(w)$ for all $w \neq w^2$. For (b) $f^2 \neq f$ and $w^2 \in \hat{\mu}^1(f)$, then $\hat{\mu}^2$ is stable under (P''_f, P_{-f}) and $\hat{\mu}^2(f)$ contains at least k workers in $B(\hat{w}_k, P_f)$ since $\hat{\mu}^2(f) = (\hat{\mu}^0(f) \cup \{w^1\}) \backslash \{w^2\}$ and $\hat{\mu}^0(f)$ contains at least k workers in $B(\hat{w}_k, P_f)$. By construction, the same is true for $DA_F[P''_f, P_{-f}](f)$ since for worker w, the member of the original blocking pair, $w^1 R_f w$ and $w \in B(\hat{w}_k, P_f)$ imply that $w^1 \in B(\hat{w}_k, P_f)$. Otherwise, if (a) $f^2 = f$ and $w^2 \notin \hat{\mu}^1(f)$, then $\hat{\mu}^2(f) = \hat{\mu}^0(f) \cup \{w^1, w^2\}$, we now allow f to have capacity $q_f + 2$, and so on. For (c) $f^2 \neq f$ and $w^2 \notin \hat{\mu}^1(f)$, we continue the above vacancy chain dynamics in the same manner until we go to Step 2 or we again reduce the capacity of f to q_f or the worker belonging to the blocking pair was previously assigned to f. In other words, we never reject a worker and only vacant positions are filled (where we allow "overbooking" for f). At the same time, we reduce f's capacity by one whenever one of the workers assigned to f leaves. Note that this process terminates as the workers' preference always weakly improves at each iteration (because f rejects no worker). If at some point exactly q_f workers are matched to f, then we must have found a stable matching for (P''_f, P_{-f}). Let $\bar{\mu}$ denote the resulting matching, and say $|\bar{\mu}(f)| = \bar{q}_f$. Then $\bar{\mu}$ is stable under (P''_f, P_{-f}) where f has $\bar{q}_f \geq q_f$ positions (instead of q_f). By construction, $\bar{\mu}(f)$ contains at least k workers in $B(\hat{w}_k, P_f)$.

Step 2: Let $\bar{q}_f$ and $\bar{\mu}$ be the outcomes of Step 1. Let $\bar{\mu}(f) = \{\bar{w}_1, \ldots, \bar{w}_{\bar{q}_f}\}$ where the workers are listed according to the ranking in P''_f. Now consider the matching market where $\bar{W} = \{\bar{w}_{q_f+1}, \ldots, \bar{w}_{\bar{q}_f}\}$ is not present. Let $\bar{\mu}'$ be defined by $\bar{\mu}'(f) = \{\bar{w}_1, \ldots, \bar{w}_{q_f}\}$ and $\bar{\mu}'(f') = \bar{\mu}(f')$ for all $f' \neq f$. Then $\bar{\mu}' \in C(P''_f, P_{-\bar{W} \cup \{f\}})$. Since $\bar{\mu}'$ contains at least k workers in $B(\hat{w}_k, P_f)$ and by definition of P''_f, $DA_F[P''_f, P_{-\bar{W} \cup \{f\}}](f)$ contains at least k workers in $B(\hat{w}_k, P_f)$. Bringing back in $\{\bar{w}_{q_f+1}, \ldots, \bar{w}_{\bar{q}_f}\}$, all firms must weakly benefit, i.e. $DA_F[P''_f, P_{-f}](f)$ must contain at least k workers in $B(\hat{w}_k, P_f)$, the desired conclusion.

Now let $\bar{W}^*$ consist of the k workers which are P_f-ranked least in $B(\hat{w}_k, P_f)$ and the $q_f - k$ workers P_f-ranked least in $B(\hat{w}_{q_f}, P_f)$. Obviously, $\hat{w}_k, \hat{w}_{q_f} \in \bar{W}^*$. Let $\bar{P}^*_f$ be a monotonic responsive extension of P_f such that $W' \bar{R}^*_f \bar{W}^*$ iff $W' \subseteq B(\hat{w}_{q_f}, P_f)$, $|W'| = q_f$ and W' contains at least k workers in $B(\hat{w}_k, P_f)$. Now we have $DA_F[P''_f, P_{-f}](f) \bar{R}^*_f \bar{W}^*$. By the definition of k, $\hat{w}_k P_f w_k$. Hence, $DA_F[P_f, P_{-f}](f)$ does not contain at least k workers in $B(\hat{w}_k, P_f)$, which implies that

$$DA_F[P''_f, P_{-f}](f) \bar{R}^*_f \bar{W}^* \bar{P}^*_f DA_F[P_f, P_{-f}](f).$$

Summarizing, we have already showed in the case $w_{q_f} R_f \hat{w}_{q_f}$ that there exist $(P_f, P_{-f}) \in supp(\tilde{P})$, $P''_f \in \mathcal{P}_f$, $\bar{W}^* \subseteq W$ and $P^*_f \in mresp(P_f)$ such that $DA_F[P''_f, P_{-f}](f) \in B(\bar{W}^*, \bar{P}^*_f)$ and $DA_F[P_f, P_{-f}](f) \notin B(\bar{W}^*, \bar{P}^*_f)$ simultaneously hold.

Now consider $(P_f, P'_{-f}) \in supp(\tilde{P})$. Let $\mu' = DA_F[P_f, P'_{-f}]$ and $\mu'' = DA_F[P''_f, P'_{-f}]$. Suppose that $\mu'(f)\bar{R}^*_f\bar{W}^*$. By construction, $\mu'(f)$ contains at least k workers in $B(\hat{w}_k, P_f)$, $|\mu'(f)| = q_f$ and $\mu'(f) \subseteq B(\hat{w}_{q_f}, P_f)$. We need to show $\mu''(f)\bar{R}^*_f\bar{W}^*$.

By construction, $\mu''(f) \subseteq B(\hat{w}_{q_f}, P_f) = A(P''_f)$. Furthermore, if $|\mu''(f)| < q_f$, then the ranking of P''_f does not matter and we would have that $\mu'' \in C(P_f, P'_{-f})$. By Theorem 2, $C(P_f, P'_{-f}) = \{\mu'\}$ which is contradiction because $|\mu''(f)| < q_f = |\mu'(f)|$. Thus, $|\mu''(f)| = q_f$.

It remains to be shown that $\mu''(f)$ contains at least k workers in $B(\hat{w}_k, P_f)$. Now if $\mu' \in C(P''_f, P'_{-f})$, then, by definition of P''_f, $B(\hat{w}_{q_f}, P''_f) = B(\hat{w}_{q_f}, P_f)$ and the fact that $\mu'(f)$ contains at least k workers in $B(\hat{w}_k, P_f)$, $\mu''(f)$ must contain at least k workers in $B(\hat{w}_k, P_f)$, the desired conclusion. Suppose that $\mu' \notin C(P''_f, P'_{-f})$. Note that we cannot have $\mu'(f) \subseteq \hat{\mu}(f) \cup B(\hat{w}_k, P_f)$ as otherwise $\mu' \in C(P_f, P'_{-f})$, $P''_f|\hat{\mu}(f) \cup B(\hat{w}_k, P_f) = P_f|\hat{\mu}(f) \cup B(\hat{w}_k, P_f)$, and $\hat{\mu}(f) \cup B(\hat{w}_k, P_f) P''_f A(P''_f)\backslash(\hat{\mu}(f) \cup B(\hat{w}_k, P_f))$ would imply $\mu' \in C(P''_f, P_{-f})$, a contradiction.[20]

To obtain a contradiction, assume $\mu''(f)$ does not contain at least k workers in $B(\hat{w}_k, P_f)$. In particular, this means that $\mu'(f) \neq \mu''(f)$. Then, since $|\hat{\mu}(f)\backslash B(\hat{w}_k, P_f)| = q_f - k$, $\mu''(f) \not\subseteq B(\hat{w}_k, P_f) \cup \hat{\mu}(f)$ as otherwise f is assigned k workers in $B(\hat{w}_k, P_f)$. Let $\underline{w}''$ denote the P''_f-least preferred worker in $\mu''(f)$. By the above we have $\underline{w}'' \notin \hat{\mu}(f) \cup B(\hat{w}_k, P_f)$.

Case 2.1: $\underline{w}''$ is the P_f-least preferred worker in $\mu''(f)$.

Because $\Pr\{\tilde{P} = (P_f, P'_{-f})\} > 0$, Theorem 2 and $\mu' = DA_F[P_f, P'_{-f}]$ imply that $C(P_f, P'_{-f}) = \{\mu'\}$. Hence, $\mu'' \notin C(P_f, P'_{-f})$. Then some pair $(\tilde{w}, f)$ blocks μ'' under (P_f, P'_{-f}). Because $\underline{w}''$ is the P_f-least worker in $\mu''(f)$, we have $\tilde{w} P_f \underline{w}''$. If $\tilde{w} \in B(\hat{w}_k, P_f) \cup \hat{\mu}(f)$, then by construction and $\underline{w}'' \notin \hat{\mu}(f) \cup B(\hat{w}_k, P_f)$ we have $\tilde{w} P''_f \underline{w}''$ and $(\tilde{w}, f)$ blocks μ'' under (P''_f, P'_{-f}), a contradiction to $\mu'' \in C(P''_f, P'_{-f})$. If $\tilde{w} \notin B(\hat{w}_k, P_f) \cup \hat{\mu}(f)$, then by construction and $\underline{w}'' \notin \hat{\mu}(f) \cup B(\hat{w}_k, P_f)$ we have $\tilde{w} P''_f \underline{w}''$ and $(\tilde{w}, f)$ blocks μ'' under (P''_f, P'_{-f}), a contradiction to $\mu'' \in C(P''_f, P'_{-f})$. Thus, Case 2.1 cannot occur.

Case 2.2: $\underline{w}''$ is not the P_f-least preferred worker in $\mu''(f)$.

Let $\hat{w}$ denote the P_f-least preferred worker in $\mu''(f)$. By $\hat{w} \neq \underline{w}''$, and the fact that by definition of P''_f, $P''_f|W\backslash(B(\hat{w}_k, P_f) \cup \hat{\mu}(f))$ coincides with $P_f|W\backslash(B(\hat{w}_k, P_f) \cup \hat{\mu}(f))$, we have $\hat{w} \in B(\hat{w}_k, P_f) \cup \hat{\mu}(f)$. Furthermore, by $\underline{w}'' \notin B(\hat{w}_k, P_f) \cup \hat{\mu}(f)$ and $\underline{w}'' P_f \hat{w}$, we cannot have $\hat{w} \in B(\hat{w}_k, P_f)$. Then by construction we must have $\hat{w} \in \hat{\mu}(f)$. But now we may just exchange the positions of $\hat{w}$ and $\underline{w}''$ and consider $P''^{\hat{w}\leftrightarrow\underline{w}''}_f$. We have $\mu'' \in C(P''^{\hat{w}\leftrightarrow\underline{w}''}_f, P'_{-f})$. In $\mu''(f)$ the P_f-least preferred worker and the $P''^{\hat{w}\leftrightarrow\underline{w}''}_f$-least preferred worker coincide and is $\hat{w}$. But now we are in Case 2.1 and this is a contradiction (where $\hat{\mu}(f)$ is replaced by $\hat{\mu}(f)^{\hat{w}\leftrightarrow\underline{w}''} = (\hat{\mu}(f)\backslash\{\hat{w}\}) \cup \{\underline{w}''\}$).

Thus, $DA_F[P''_f, P'_{-f}](f)$ contains q_f workers in $A(P''_f)$ and at least k workers in $B(\hat{w}_k, P_f)$. Now by our choice of $\bar{W}^*$ and $\bar{P}^*_f$ we obtain that, for any profile P'_{-f} in the support of $\tilde{P}_{-f}|_{P_f}$, $DA_F[P_f, P'_{-f}](f)\bar{R}^*_f\bar{W}^*$ implies $DA_F[P''_f, P'_{-f}](f)\bar{R}^*_f\bar{W}^*$.

[20]Here we use again the convention SP''_fT iff sP''_ft for all $s \in S$ and all $t \in T$.

Since $DA_F[P''_f, P_{-f}](f)\bar{R}^*_f\bar{W}^*\bar{P}^*_f DA_F[P_f, P_{-f}](f)$ and $P \in supp(\tilde{P})$, we have

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

Hence, not $DA_F[P_f, \tilde{P}_{-f}|_{P_f}](f)P^{sd}_f DA_F[P''_f, \tilde{P}_{-f}|_{P_f}](f)$ and DA_F is not monotonic OBIC under $\tilde{P}$, a contradiction. □

B Proof of Theorem 2

Let φ be a stable mechanism and $\tilde{P}$ be a common belief. To obtain a contradiction, suppose that there is some $P \in \mathcal{P}$ such that both $\Pr\{\tilde{P} = P\} > 0$ and $|C(P)| \geq 2$ hold. By stability of φ, $\varphi[P] \in C(P)$. Let $\mu \in C(P)\backslash\{\varphi[P]\}$ be arbitrary. If there is some $w \in W$ such that $\mu(w)P_w\varphi[P](w)$, then similarly to Ehlers and Massó (2007, Theorem 1) we can show that φ is not monotonic OBIC under $\tilde{P}$. Thus, for all $\mu \in C(P)\backslash\{\varphi[P]\}$,

$$\varphi[P](w)R_w\mu(w) \tag{8}$$

for all $w \in W$. Hence, and since $DA_W[P]$ is the worker-optimal stable matching,

$$\varphi[P] = DA_W[P]. \tag{9}$$

Observe that (9) needs to hold for all profiles belonging to the support of $\tilde{P}$.

Fix a matching $\mu \in C(P)\backslash\{\varphi[P]\}$. Since $\mu \neq \varphi[P]$ and $\varphi[P] = DA_W[P]$, there exists some $f \in F$ such that $\mu(f) \neq \varphi[P](f)$. Then, by Roth and Sotomayor (1990, Theorems 5.12 and 5.13), $|\mu(f)| = |\varphi[P](f)| = q_f$. Note that $DA_W[P]$ matches each firm f with f's worst set of partners−according to any monotonic responsive extension of P_f−, among all partners that f is matched to across all stable matchings. Thus, by (9),

$$\mu(f)P^*_f\varphi[P](f) \tag{10}$$

for all monotonic responsive extensions P^*_f of P_f.

Let $\bar{w} \in \mu(f)$ be such that $wR_f\bar{w}$ for all $w \in \mu(f)$. Let $P'_f \in \mathcal{P}_f$ be such that (i) $A(P'_f) = B(\bar{w}, P_f)$ and (ii) $P'_f|A(P'_f) = P_f|A(P'_f)$. By construction of P'_f, $\mu \in C(P'_f, P_{-f})$. Hence, either $\varphi[P'_f, P_{-f}](f) = \mu(f)$, in which case, by (10),

$$\varphi[P'_f, P_{-f}](f)P^*_f\varphi[P](f)$$

for all monotonic responsive extensions P^*_f of P_f, or else $\varphi[P'_f, P_{-f}](f) \neq \mu(f)$. We want to show that $\varphi[P'_f, P_{-f}](f)P^*_f\mu(f)$ for all monotonic responsive extensions P^*_f of P_f. Suppose otherwise; since $\varphi[P'_f, P_{-f}](f) \neq \mu(f)$, this means that $\mu(f)\hat{P}^*_f\varphi[P'_f, P_{-f}](f)$ for some monotonic responsive extension $\hat{P}^*_f$ of P_f. Then, by Roth and Sotomayor (1989, Theorem 4), for all $w \in \mu(f)$ and all $w' \in \varphi[P'_f, P_{-f}](f)\backslash\mu(f)$,

$$wP'_fw'. \tag{11}$$

Consider any $w' \in \varphi[P'_f, P_{-f}](f)\backslash\mu(f)$. Since $\varphi[P'_f, P_{-f}]$ is individually rational under (P'_f, P_{-f}), the definition of P'_f implies that $w' \in B(\bar{w}, P_f)$. Hence, by $P'_f|B(\bar{w}, P_f) = P_f|B(\bar{w}, P_f)$, $w'R_f\bar{w}$, which either contradicts (11) or else it means that $\varphi[P'_f, P_{-f}](f) \subsetneq \mu(f)$. Then, $\left|\varphi[P'_f, P_{-f}](f)\right| < |\mu(f)|$. But this contradicts Theorem 5.13 in Roth and Sotomayor (1990). Thus, $\varphi[P'_f, P_{-f}](f)P^*_f\mu(f)$ for all monotonic responsive extensions P^*_f of P_f. By (10),

$$\varphi[P'_f, P_{-f}](f)P^*_f\varphi[P](f) \tag{12}$$

for all monotonic responsive extensions P^*_f of P_f.

We have already shown that there exist $P \in \mathcal{P}$ with $\Pr\{\tilde{P} = P\} > 0$, $f \in F$ and $P'_f \in \mathcal{P}_f$ such that (12) holds for all $P^*_f \in mresp(P_f)$. To obtain a contradiction with truth-telling being a monotonic OBNE, we will find a $\bar{P}^*_f \in mresp(P_f)$ such that for all $P'_{-f} \in \mathcal{P}_{-f}$ with $\Pr\{\tilde{P} = (P_f, P'_{-f})\} > 0$ and $\varphi[P_f, P'_{-f}](f) \in B(\mu(f), \bar{P}^*_f)$, we have that $\varphi[P'_f, P'_{-f}](f) \in B(\mu(f), \bar{P}^*_f)$. This would imply then that the inequality

$$\sum_{\{S' \in 2^W | S' \bar{R}^*_f \mu(f)\}} \Pr\{\varphi[P'_f, \tilde{P}_{-f} \mid_{P_f}](f) = S'\} \geq \sum_{\{S' \in 2^W | S' \bar{R}^*_f \mu(f)\}} \Pr\{\varphi[P_f, \tilde{P}_{-f} \mid_{P_f}](f) = S'\} \tag{13}$$

holds. Since $\varphi[P'_f, P_{-f}](f) \in B(\mu(f), \bar{P}^*_f)$, $\varphi[P_f, P_{-f}](f) \notin B(\mu(f), \bar{P}^*_f)$ and $\Pr\{\tilde{P} = P\} > 0$, we could conclude then that the inequality in (13) is strict and hence, $\varphi[P_f, \tilde{P}_{-f} \mid_{P_f}](f)$ does not first order stochastically dominate $\varphi[P'_f, \tilde{P}_{-f} \mid_{P_f}](f)$.

For showing this, let $\bar{P}^*_f$ be the monotonic responsive extension of P_f such that for all $W' \in 2^W$, $W'\bar{P}^*_f\mu(f)$ if and only if (i) $|W'| = q_f$ (note that $|\mu(f)| = q_f$) and (ii) there is a one-to-one mapping $h : W' \to \mu(f)$ such that $w'R_f h(w')$ for all $w' \in W'$ with strict preference holding for some worker in W'. In other words, W' is strictly preferred to $\mu(f)$ under $\bar{P}^*_f$ only if W' is unambiguously strictly preferred to $\mu(f)$ for all monotonic responsive extensions of P_f.

Let $P'_{-f} \in \mathcal{P}_{-f}$ be such that both $\Pr\{\tilde{P} = (P_f, P'_{-f})\} > 0$ and $\varphi[P_f, P'_{-f}](f) \in B(\mu(f), \bar{P}^*_f)$. We want to show that $\varphi[P'_f, P'_{-f}](f) \in B(\mu(f), \bar{P}^*_f)$. Let $\hat{\mu} = \varphi[P_f, P'_{-f}]$ and $\mu' = \varphi[P'_f, P'_{-f}]$ and assume $\hat{\mu} \neq \mu'$ (otherwise, the statement follows trivially). By the definitions of $\bar{P}^*_f$ and P'_f, we have that $\hat{\mu}(f)\bar{R}^*_f\mu(f)$, $\hat{\mu}(f) \subseteq A(P'_f)$ and $|\hat{\mu}(f)| = q_f$. Thus, by $A(P'_f) = B(\bar{w}, P_f)$ and $P'_f|A(P'_f) = P_f|A(P'_f)$, $\hat{\mu} \in C(P'_f, P'_{-f})$. Hence, $\mu', \hat{\mu} \in C(P'_f, P'_{-f})$. We distinguish between two cases.

Case 1: $\mu'(f)\bar{P}^*_f\hat{\mu}(f)$. Since $\hat{\mu}(f) \in B(\mu(f), \bar{P}^*_f)$, we have that $\mu'(f) \in B(\mu(f), \bar{P}^*_f)$.

Case 2: $\hat{\mu}(f)\bar{P}^*_f\mu'(f)$. But then $\mu'(f) \subseteq A(P'_f) = B(\bar{w}, P_f) \subseteq A(P_f)$ and by $P'_f|A(P'_f) = P_f|A(P_f)$, we have $\mu' \in C(P_f, P'_{-f})$. Note that (P_f, P'_{-f}) is a profile belonging to the support of $\tilde{P}$ and thus, by (9), $\hat{\mu} = \varphi[P_f, P'_{-f}] = DA_W[P_f, P'_{-f}]$. Hence, $\hat{\mu}(w)R_w\mu'(w)$ for all $w \in W$, and again, by the same argument we used for (12), $\mu'(f)\bar{P}^*_f\hat{\mu}(f)$, a contradiction.

Cases 1 and 2 show that if $P'_{-f} \in \mathcal{P}_{-f}$ is such that $\Pr\{\tilde{P} = (P_f, P'_{-f})\} > 0$ and $\varphi[P_f, P'_{-f}](f) \in B(\mu(f), \bar{P}^*_f)$, then $\varphi[P'_f, P'_{-f}](f) \in B(\mu(f), \bar{P}^*_f)$. Since $\Pr\{\tilde{P}_{-f}|_{P_f} =$

$P_{-f}\} > 0$, $\varphi[P](f) \notin B(\mu(f), \bar{P}^*_f)$, and $\varphi[P'_f, P_{-f}](f) \in B(\mu(f), \bar{P}^*_f)$, it follows that

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

which means φ is not monotonic OBIC under $\tilde{P}$, a contradiction. $\square$

References

Abdulkadiroğlu, A., Y.-K. Che, and Y. Yosuda, Resolving conflicting preferences in school choice: the Boston mechanism reconsidered, Amer. Econ. Rev. 101 (2011) 399–410.

Abdulkadiroğlu, A., P. Pathak, and A.E. Roth, Strategy-proofness versus efficiency in matching with indifferences: redefining the NYC high school match, Amer. Econ. Rev. 99 (2009) 1954–1978.

Abdulkadiroğlu, A. and T. Sönmez, School choice: a mechanism design approach, Amer. Econ. Rev. 93 (2003) 729–747.

Alcalde, J., Implementation of stable solutions to marriage problems, J. Econ. Theory 69 (1996) 240–254.

Bergemann, D. and S. Morris, Robust mechanism design, Econometrica 73 (2005), 1771–1813.

Bergemann, D. and S. Morris, An introduction to robust mechanism design, Foundations and Trends in Microeconomics, Vol. 8, No. 3 (2012), 169–230.

Chen, Y. and O. Kesten, Chinese college admissions and school choice reforms: a theoretical analysis, J. Pol. Economy 125 (2017) 99–139.

Chen, Y. and T. Sönmez, School choice: an experimental study, J. Econ. Theory 127 (2006), 202–231.

D'Aspremont, C. and B. Peleg, Ordinal bayesian incentive compatible representation of committees, Soc. Choice Welfare 5 (1988) 261–280.

Dubins, L. and D. Freedman, Machiavelli and the Gale-Shapley algorithm, Amer. Math. Monthly 88 (1981) 485–494.

Ehlers, L., Truncation strategies in matching markets, Math. Oper. Research 33 (2008) 327–335.

Ehlers, L. and J. Massó, Incomplete information and singleton cores in matching markets, J. Econ. Theory 136 (2007) 587–600.

Ehlers, L. and J. Massó, Matching markets under (in)complete information, J. Econ. Theory 157 (2015) 295–317.

Gale, D. and L. Shapley, College admissions and the stability of marriage, Amer. Math. Monthly 69 (1962) 9–15.

Kesten, O., An advice to the organizers of entry-level labor markets in the United Kingdom, Working Paper (2005).

Ma, J., Stable matchings and rematching-proof equilibria in a two-sided matching market, J. Econ. Theory 66 (1995) 352–369.

Ma, J., Stable matchings and the small core in Nash equilibrium in the college admissions problem, Rev. Econ. Design 7 (2002) 117–134.

Majumdar, D. and A. Sen, Ordinally bayesian incentive-compatible voting rules, Econometrica 72 (2004) 523–540.

Roth, A.E., The economics of matching: stability and incentives, Math. Oper. Research 7 (1982) 617–628.

Roth, A.E., The evolution of the labor market for medical interns and residents: a case study in game theory, J. Pol. Economy 92 (1984a) 991–1016.

Roth, A.E., Misrepresentation and stability in the marriage problem, J. Econ. Theory 34 (1984b) 383–387.

Roth, A.E., The college admissions problem is not equivalent to the marriage problem, J. Econ. Theory 36 (1985) 277–288.

Roth, A.E., A natural experiment in the organization of entry level labor markets: regional markets for new physicians and surgeons in the U.K., Amer. Econ. Rev. 81 (1991) 415–440.

Roth, A.E., The economist as engineer: game theory, experimentation, and computation as tools for design economics, Econometrica 70 (2002) 1341–1378.

Roth, A.E. and E. Peranson, The redesign of the matching market for American physicians: some engineering aspects of economic design, Amer. Econ. Rev. 89 (1999) 748–780.

Roth, A.E. and M. Sotomayor, The college admissions problem revisited, Econometrica 57 (1989) 559–570.

Roth, A.E. and M. Sotomayor, Two-sided Matching: A Study in Game-Theoretic Modelling and Analysis, Cambridge University Press, Cambridge, England, 1990.

Shin, S. and S.-C. Suh, A mechanism implementing the stable rule in marriage problems, Econ. Letters 51 (1996) 185–189.

Sönmez, T., Games of manipulation in marriage problems, Games Econ. Behav. 20 (1997) 169–176.

Ünver, U., On the survival of some unstable two-sided matching mechanisms, Int. J. Game Theory 33 (2005) 239–254.