On Random Matching Markets: Properties and Equilibria

Joana Pais[*]

UECE - Research Unit on Complexity in Economics

ISEG/UTL

Rua Miguel Lupi, 20

1249-078 Lisboa, Portugal

October 2005

Abstract

We consider centralized matching markets in which, starting from an arbitrary matching, firms are successively chosen in a random fashion and offer their positions to the workers they prefer the most. We propose an algorithm that generalizes some well-known algorithms and explore some of its properties. In particular, different executions of the algorithm may lead to different output matchings. We then study incentives in the revelation game induced by the algorithm. We prove that ordinal equilibria always exist. Furthermore, every matching that results from an equilibrium play of the game is stable for a particular preference profile. Namely, if an ordinal equilibrium exists in which firms reveal their true preferences, only matchings that are stable for the true preferences can be obtained.

JEL Classification: C78; J44

Keywords: Matching Markets; Stability; Random Mechanism

1 Introduction

Simple models of two-sided matching have proved to be very useful in understanding the organization and evolution of many markets, namely labor markets, as well as other economic environments. The term "two-sided" refers to the fact that agents belong to one of two disjoint sets and can never interchange roles. Thus, we may have, for instance, firms and workers, hospitals and interns, colleges and students, men and women. Each agent has preferences over the other side of the market and the prospect of being unmatched and the matching problem reduces to assigning the members of these two sets to one another. When each agent may be matched with at most one agent of the opposite set we speak of a "marriage model." This tractable model gives a lot of insight on many phenomena observed in real markets as documented in the large body of literature devoted to it.[1]

Stable matchings are those that we may expect to observe in practice: if the market outcome is unstable, there is an agent or a pair of agents (henceforth, a firm and a worker) with an incentive to circumvent the matching. Under a stable matching every agent prefers his partner to being alone and, moreover, no pair of agents, consisting of a firm and a worker, who are not matched to each other would rather prefer to be so matched. In a seminal paper, Gale and Shapley (1962) demonstrated that at least one stable matching exists for every marriage market. Their proof of existence of stable matchings consists of a procedure, the "deferred-acceptance" algorithm which, for every stated preferences, transforms the empty matching (in which all agents are unmatched) into a stable matching.

In this paper we consider an extension to Gale and Shapley's algorithm or, to be precise, to the version proposed by McVitie and Wilson (1970). We start from an arbitrary matching and the algorithm proceeds by creating, at each step, a provisional matching. Hence, at each moment in time, a firm is randomly chosen and the best worker on its list of preferences is considered. If this worker is already holding a firm he prefers,

---

[1]For an excellent survey on the matching problem, see Roth and Sotomayor (1990).

the matching goes unchanged and this particular worker is removed from the firm's list. Otherwise, they are (temporarily) matched, pending the possible draw of even better firms willing to match this worker. McVitie and Wilson's algorithm is an instance of the one we are proposing, when the initial matching is the empty matching. Moreover, it also encompasses the algorithm proposed by Blum, Roth, and Rothblum (1997) to explore the vacancy chain problem when the input matching is firm-quasi-stable (i.e., a matching whose stability was disrupted by the emergence of a new position or the retirement of a worker).

We then analyze incentives in a centralized market where agents submit ordered lists of preferences on prospective partners to a clearinghouse, which then produces a matching by processing these lists according to the algorithm we propose. The random order in which firms are selected when the algorithm is run introduces some uncertainty in the output reached. It may happen that, starting with the same input matching, different executions of the algorithm yield different outcomes for the same preference profile. Since agents' preferences are merely ordinal in nature, we use a concept of equilibrium based on first-order stochastic dominance. This guarantees that in equilibrium each agent plays his best response to the others' strategies for every utility representation of the preferences.[2] We prove the existence of equilibria and show that *some* stability is preserved in every equilibrium. Following the literature, we then focus on equilibria in which one side of the market, in particular the firms' side, tells the truth and provide a partial characterization of such equilibria. Contrary to Gale and Shapley, possibly not every stable matchings can be supported at equilibrium, since the initial matching constrains the set of achievable matchings, but we will show that some stable matchings can be reached with probability one. Furthermore, we prove that, even though workers may not play straightforwardly, stability with respect to the true preferences holds for *any* matching that results from a play of equilibrium strategies in which firms reveal their true preferences.

We proceed as follows. In Section 2 we present the simple marriage model and

---

[2]This concept has been used in the context of matching markets with incomplete information in Roth and Rothblum (1999), Ehlers (2003, 2004), and Ehlers and Massó (2003).

introduce notation. We formally describe the algorithm in Section 3, showing that it captures other algorithms. In addition, some of its features are explored. In Section 4 we turn our attention to a different class of questions, related to individual decision making. The matching process is modeled as a game and its equilibria are characterized. Some concluding remarks follow in Section 5.

2　The Marriage Model

Consider two finite and disjoint sets $F = \{f_1, ..., f_n\}$ and $W = \{w_1, ..., w_p\}$, where F is the set of firms and W is the set of workers. Let $V = F \cup W$. Sometimes we refer to a generic agent by v and we use f and w to represent a generic firm and worker, respectively. Each agent has a strict, complete, and transitive preference relation over the agents on the other side of the market and remaining unmatched. The preferences of a firm f, for example, can be represented by $P(f) = w_3, w_1, f, w_2, ..., w_4$, indicating that f's first choice is to be matched to w_3, its second choice is w_1 and it prefers remaining unmatched to being assigned to any other worker. Sometimes it is sufficient to describe only f's ranking of workers it prefers to remaining unmatched, so that the above preferences can be abbreviated as $P(f) = w_3, w_1$. Let $P = (P(f_1), ..., P(f_n), P(w_1), ..., P(w_p))$ denote the profile of all agents' preferences; we sometimes write it as $P = (P(v), P_{-v})$ where P_{-v} is the set of preferences of all agents other than v. Further, we may use P_U, where $U \subseteq V$, to denote the profile of preferences $(P(v))_{v \in U}$. We write $v'P(v)v''$ when v' is preferred to v'' under preferences $P(v)$ and we say that v *prefers* v' to v''. We write $v'R(v)v''$, when v likes v' *at least as well as* v'' (it may be the case that v' and v'' are the same agent). Formally, a *marriage market* is a triple (F, W, P). Let $A(P(f))$ denote the set of workers that are *acceptable* to firm f, i.e., $A(P(f)) = \{w \in W : wP(f)f\}$; $A(P(w))$ is defined analogously. A pair $(f, w) \in F \times W$ is acceptable if f and w are acceptable to each other.

　　An outcome for a marriage market, a *matching*, is a function $\mu : V \longrightarrow V$ satisfying the following: *(i)* for each f in F and for each w in W, $\mu(f) = w$ if and only if $\mu(w) = f$;

(ii) if $\mu(f) \neq f$ then $\mu(f) \in W$; *(iii)* if $\mu(w) \neq w$ then $\mu(w) \in F$. If $\mu(v) = v$, then v is *unmatched* under μ, while if $\mu(w) = f$, we say that f and w are *matched* to one another. A description of a matching is given by $\mu = \{(f_1, w_2), (f_2, w_3)\}$, indicating that f_1 is matched to w_2, f_2 is matched to w_3 and the remaining agents in the market are unmatched. A matching μ is *individually rational* if each agent is acceptable to its partner, i.e., $\mu(v)R(v)v$, for all $v \in V$. We denote the set of all individually rational matchings by $IR(P)$. Two agents f and w form a *blocking pair* for μ if they prefer each other to the agents they are actually assigned to under μ, i.e., $fP(w)\mu(w)$ and $wP(f)\mu(f)$. A matching μ is *stable* if it is individually rational and it is not blocked by any pair of agents. A matching μ is *firm-quasi-stable* if it is individually rational and if every blocking pair for μ contains an unmatched firm. We denote the set of all stable matchings by $S(P)$ and the set of all firm-quasi-stable matchings by $QS(P)$. The set $S(P)$ forms a lattice (see Roth and Sotomayor (1990) for a formal statement of this result, attributed to John Conway), with the extreme elements being the firm-optimal stable matching μ_F and the worker-optimal stable matching μ_W. There exists no stable matching μ that matches any firm f to a partner that it prefers to $\mu_F(f)$. Analogously, μ_W is optimal for workers. Finally, we define a firm f and a worker w to be *achievable* for each other if f and w are paired at some stable matching.

3 The Algorithm

In this section, we provide an informal description of Gale and Shapley's algorithm, as well as of the one proposed by Blum, Roth, and Rothblum (1997). Subsequently, we present the generalized deferred-acceptance algorithm and explore some of its properties.

Gale and Shapley (1962) showed that a stable matching exists for every marriage market. Their proof is in fact an algorithm for finding such a matching. Starting from a situation where no agent is matched, in the "deferred-acceptance" algorithm (DA-algorithm), firms propose to workers who can hold at most one unrejected offer at any time. At any step of the algorithm, every rejected firm proposes to its next choice, as

long as there are acceptable workers on its list to whom it has not made an offer yet. The algorithm stops after the step in which every rejected firm has proposed to all of its acceptable workers.

McVitie and Wilson (1970) proposed a different version of this algorithm, which turned out to be a key piece in obtaining the full set of stable matchings. The difference with respect to the DA-algorithm is that at each step of this algorithm only one randomly chosen firm makes an offer. Nevertheless, the output matching of McVitie and Wilson's algorithm is independent of the order in which firms are selected to propose and it coincides with the output produced by the DA-algorithm. Furthermore, it is the firm-optimal stable matching μ_F. (Alternatively, if in any of the two algorithms described the workers proposed, μ_W would be obtained.)

These algorithms were used to study entry-level markets, characterized by the availability of cohorts of vacant positions and, simultaneously, of candidates in need of a position. Blum, Roth, and Rothblum (1997) developed a deferred-acceptance algorithm to model senior level labor markets, where positions become available when an incumbent worker retires or when a new firm comes into the market. This leads to vacancy chains, since as one firm succeeds in filling its vacancy it may cause another firm to have one. The algorithm starts with an arbitrary matching, selects a firm whose position is vacant and lets it approach its most preferred workers in order of preference. At each step a blocking pair is satisfied, but only when the firm's position is vacant and the offer is acceptable. This process is iterated until there is no firm eligible to propose. It is shown that all executions of this algorithm with the same input terminate after a finite number of steps and yield the same output matching. Moreover, when the input matching is firm-quasi-stable, the algorithm terminates at a stable matching.

3.1   Definition of the DA$^{\mu^I}$-algorithm

In what follows, we describe a modified version of McVitie and Wilson's algorithm to be applied to *any* input matching. It differs from the algorithm proposed by Blum, Roth,

and Rothblum (1997) in the fact that not only firms with vacancies can make proposals. Indeed, *any* firm can be greedy and invite the most preferred workers on its list of preferences. Thus, starting with an arbitrary matching μ^I, at each step, a randomly selected firm, say f, approaches the first worker on its preference list to whom it has not made an offer yet, say w. If the worker rejects, no change occurs. If the worker accepts, a new matching is formed where f and w are matched and their previous partners—if any—remain unmatched. This process is repeated until no firm is willing or able to make a new offer (either its proposal was accepted and is held by some worker or the firm has already proposed to all the acceptable workers on its list). Formally:

Definition 1 *The Generalized Deferred-Acceptance Algorithm (DA^{μ^I}-algorithm):*

Input: a matching μ^I; a preference profile P.

Initialization.

1. (a) For all $f \in F$, $A_f^0 = A(P(f)) \cup \{f\}$;

    (b) $\mu^0 = \mu^I$; $i := 1$;

2. If, for all $f \in F$, $\mu^{i-1}(f) = \max_{P(f)} A_f^{i-1}$, then stop with μ^{i-1}.

3. Else, take any firm f such that:

    (a) either $\max_{P(f)} A_f^{i-1} = f$ and $\mu^{i-1}(f) \neq f$, leading to $\mu^i = \mu^{i-1} \setminus \{(f, \mu^{i-1}(f))\}$;

    (b) or $\max_{P(f)} A_f^{i-1} = w$ and $\mu^{i-1}(f) \neq w$, in which case:

        i. if $\mu^{i-1}(w) P(w) f$, then $\mu^i = \mu^{i-1}$ and $A_f^i = A_f^{i-1} \setminus \{w\}$, $A_{f'}^i = A_{f'}^{i-1}$, for all $f' \neq f$;

        ii. else:

            A. if $\mu^{i-1}(f) = f$ and $\mu^{i-1}(w) = w$, then $\mu^i = \mu^{i-1} \cup \{(f, w)\}$ and $A_{f'}^i = A_{f'}^{i-1}$, for all $f' \in F$;

            B. if $\mu^{i-1}(f) \neq f$ and $\mu^{i-1}(w) = w$, then $\mu^i = (\mu^{i-1} \cup \{(f, w)\}) \setminus \{(f, \mu^{i-1}(f))\}$ and $A_{f'}^i = A_{f'}^{i-1}$, for all $f' \in F$;

C. if $\mu^{i-1}(f) = f$ and $\mu^{i-1}(w) \neq w$, then $\mu^i = (\mu^{i-1} \cup \{(f, w)\}) \setminus \{(\mu^{i-1}(w), w)\}$
   and $A^i_{\mu^{i-1}(w)} = A^{i-1}_{\mu^{i-1}(w)} \setminus \{w\}$, $A^i_{f'} = A^{i-1}_{f'}$, for all $f' \neq \mu^{i-1}(w)$;

D. if $\mu^{i-1}(f) \neq f$ and $\mu^{i-1}(w) \neq w$, then $\mu^i = (\mu^{i-1} \cup \{(f, w)\}) \setminus \{(f, \mu^{i-1}(f)),$
   $(\mu^{i-1}(w), w)\}$ and $A^i_{\mu^{i-1}(w)} = A^{i-1}_{\mu^{i-1}(w)} \setminus \{w\}$, $A^i_{f'} = A^{i-1}_{f'}$, for all $f' \neq$
   $\mu^{i-1}(w)$;

4. $i := i + 1$; go to 2.

3.2   Properties of the DA$^{\mu^I}$-algorithm

In the DA$^{\mu^I}$-algorithm no firm proposes to the same worker twice: if a firm, say f, is rejected by some worker w at step i, he will not be part of A^{i+1}_f and hence, permanently removed from its list of workers to be proposed. This feature guarantees that cycling is avoided, ensuring that every execution of the algorithm with an arbitrary input matching terminates after a finite number of iterations. Still, as the following example shows, for a given input matching and a preference profile, the output matching need not be unique.

Example 1 *The outcome depends on the selection of the order by which firms propose.*

Let (F, W, P) be a marriage market with P such that

$$P(w_1) = f_2, \ f_1 \quad P(f_1) = w_1, \ w_2$$
$$P(w_2) = f_1, \ f_2 \quad P(f_2) = w_2, \ w_1.$$

Consider the DA$^{\mu^I}$-algorithm applied to P, with $\mu^I = \{(f_1, \ w_2)\}$.

Start by considering the case in which f_1 is the first to make an offer. According to the algorithm (step 3(b)iiB), f_1 proposes to w_1 and w_1 accepts this proposal, as he is initially unmatched and f_1 is an acceptable firm. Then, f_2's opportunity comes and it proposes to its most preferred worker, w_2, who is currently unmatched (step 3(b)iiA). As both firms are matched to the workers they proposed to, the algorithm stops (step 2). The firm-optimal matching $\mu_F = \{(f_1, \ w_1), (f_2, \ w_2)\}$ is obtained.

Nevertheless, if the first randomly chosen firm is f_2, its proposal to w_2 is refused, as this worker is still matched to f_1 (step 3(b)i). Then, we can either have f_2 proposing again or f_1, both to w_1. If f_2 proposes first, w_1 accepts (step 3(b)iiA); next, it must be f_1's turn to propose to w_1, who rejects this offer (step 3(b)i), and finally to w_2, who accepts it. On the other hand, if f_1 proposes w_1 first, he accepts (step 3(b)iiB); however, he exchanges it for f_2, when this firm is given the opportunity to move (step 3(b)iiC). Thus, according to this order of proposals, f_1 is also assigned to w_2. In both cases, the worker-optimal matching $\mu_W = \{(f_1, \ w_2), (f_2, \ w_1)\}$ is reached as the outcome of the DA^{μ^I}-algorithm. $\diamond$

This example shows that different executions of the DA^{μ^I}-algorithm with the same input matching may yield different output matchings. In what follows we will be more precise in describing this uncertainty and introduce some notation.

We consider lotteries over sequences of firms, where each sequence corresponds to an order in which firms are given the opportunity to make an offer. The randomization over the set of firms is not simple: only firms whose preference lists have not been exhausted and that are not matched to their best elements are contemplated. Therefore, given a sequence, we start from the last firm that has been considered and take the next firm in the sequence that fulfills these requirements. In between, every ineligible firm (i.e., a firm that is currently matched to the best worker on its list of preferences or whose list of workers is already empty) is discarded. The game ends when every firm in the remainder of the sequence is ineligible to propose. In order to ensure that, once started, every execution of the algorithm is run to completion, we will allow for infinite sequences, where each firm appears an infinite number of times. The sample space over which lotteries are considered is denoted by Σ.

Although a random element appears each time a firm is chosen, all the uncertainty is fully translated into a probability distribution over the set of matchings. For each input matching and for each profile of preferences, a lottery over matchings is obtained. Hence, fix a probability distribution on Σ and take an initial matching μ^I, a preference profile P,

and an arbitrary worker w. We will let $\widetilde{DA}^{\mu^I}[P]$ denote the probability distribution over the set of matchings induced by the DA^{μ^I}-algorithm and $\widetilde{DA}^{\mu^I}[P](w)$ be the distribution that $\widetilde{DA}^{\mu^I}[P]$ induces over $F \cup \{w\}$. The expression $\Pr\{\widetilde{DA}^{\mu^I}[P] = \mu\}$ represents the probability that μ is the output of the DA^{μ^I}-algorithm with preferences P. Observe that this probability rests on the probability distribution on Σ, but all results hold regardless of this lottery. Finally, for all $w \in W$, $v \in F \cup \{w\}$, the subset of all possible orders leading to an output matching where w is assigned to v is denoted by $\Sigma_{v,w}$.

In the particular case that the input matching is the empty matching, $\emptyset$, a degenerate probability distribution over the set of matchings is obtained. In fact, it turns out that, when $\mu^I = \emptyset$, the DA^{μ^I}-algorithm specializes to McVitie and Wilson's algorithm and the firm-optimal stable matching is obtained with probability one. For illustration, consider the matching market in Example 1 and assume the algorithm starts with the empty matching. If f_1 is the first firm to propose, it invites w_1 and w_1 accepts this proposal. Then, f_2 follows and proposes to w_2, who also accepts. If we reverse the order of events and f_2 is the first to move, w_2 accepts its proposal, given that he is currently unmatched; f_1 invites the best worker on its list, w_1, who also accepts. Thus, we always reach μ_F for every order of proposals.

Proposition 1 *For any matching market* (F, W, P), $\Pr\{\widetilde{DA}^{\emptyset}[P] = \mu_F\} = 1$.

Proof. First, we will show that no worker rejects a proposal from its partner at μ_F in any execution of the algorithm. By contradiction, assume that there exists an order of proposals under which at least one worker rejects its partner at μ_F. Suppose that w is the first worker to reject $\mu_F(w)$. Let $f = \mu_F(w) \in F$. This implies w obtained a proposal from a firm he strictly prefers, say $\widehat{f}$. So, $\widehat{f}P(w)f$; given that μ_F is stable, we must have $\mu_F(\widehat{f})P(\widehat{f})w$. Then, before inviting w, $\widehat{f}$ must have proposed to $\mu_F(\widehat{f})$ and $\mu_F(\widehat{f})$ must have rejected its proposal, contradicting the fact that w was the first worker to reject his partner at μ_F.

It follows that, in the output matching, for every order in which firms propose, every firm must be assigned to a worker at least as good as its mate at μ_F. Suppose that there

exists an output matching μ and a firm, say f', matched to some $w' \in W$ under μ, such that $w'P(f')\mu_F(f')$. This implies that μ is not stable by definition of μ_F. Naturally, no firm ever proposes to a worker that it finds unacceptable; on the other hand, a worker never accepts a proposal from an unacceptable firm. Together with the fact that every agent is unmatched in the initial matching, this implies that μ is individually rational. Thus, if μ is not stable there must exist a pair that blocks μ, say f'' and w''. Since $w''P(f'')\mu(f'')$, f'' must have proposed to w'' and w'' must have rejected this proposal. But this means w'' received a better offer, from a firm he strictly prefers to f''. Then, $\mu(w'')P(w'')f''$, contradicting the fact that f'' and w'' block μ. As a consequence, no firm can be matched to a worker it strictly prefers to its partner at μ_F. Therefore, for every order of proposals, μ_F is the matching that is reached as the outcome of the $DA^{\emptyset}$-algorithm. ∎

Another case worth describing is when the input matching is a firm-quasi-stable matching, as defined by Sotomayor (1996) and Blum, Roth, and Rothblum (1997). In Proposition 2 we show that when the initial matching is firm-quasi-stable, the same stable output matching is obtained, independently of the order in which firms propose.

Remark 1 turns out to be crucial in what follows.

Remark 1 *The DA^{μ^I}-algorithm implies that once a firm proposes to a worker and he accepts, this firm cannot fire him nor exchange him for another worker. In fact, when the proposal is made, the firm reveals that this particular worker is the best among all who have not rejected it. If the worker accepts, the only occasion under which the firm can make a proposal again is when the worker it holds accepts an offer from a different firm.*

Proposition 2 *Let (F, W, P) be a matching market. For all $\mu^I \in QS(P)$, there is some $\mu \in S(P)$ such that $\Pr\{\widetilde{DA}^{\mu^I}[P] = \mu\} = 1$.*

Proof. Take $\mu^I \in QS(P)$. For every order of proposals, the first firm to have its offer accepted must be unmatched at μ^I. In fact, by definition of firm-quasi-stability, if

11

(f, w) blocks μ^I, f must be unmatched at μ^I. Assume hence that f proposes to w and that this proposal is the first to be accepted. It follows that, after this acceptance, w is strictly better off and every other worker is holding its initial partner. The rest of the proof now follows using an induction argument.

Suppose that up to step i in the algorithm only firms with vacancies have had their proposals accepted. Let μ^i be the matching at the beginning of step $i + 1$. Assume that all workers are weakly better off at μ^i than at the initial matching μ^I. We will show, by way of contradiction, that the next firm to be accepted by some worker must be unmatched. So assume that f is matched to w at μ^i, it proposes to w' and this proposal is accepted. Thus, at μ^{i+1}, f and w' are matched to each other and their former partners are unmatched. By Remark 1, if f is matched to w at μ^i and it is willing to propose to another worker, it must be the case that $\mu^I(f) = w$. Now, by assumption, $\mu^i(w')R(w')\mu^I(w')$. Since $fP(w')\mu^i(w')$, we have $fP(w')\mu^I(w')$. Further, $w'P(f)\mu^I(f) = w$. Thus, (f, w') form a blocking pair to μ^I and $\mu^I(f) \neq f$, contradicting the fact that $\mu^I \in QS(P)$.

The algorithm starts with an unmatched firm having its proposal accepted and we have proved that it must continue to be so. It follows that the DA$^{\mu^I}$-algorithm reduces to Blum, Roth, and Rothblum's algorithm when μ^I is firm-quasi-stable and all of its results are replicated. Thus, given a matching market (F, W, P) and an input matching $\mu^I \in QS(P)$, the same stable matching will be obtained in any execution of the algorithm. ∎

Starting with a firm-quasi-stable matching, the DA$^{\mu^I}$-algorithm replicates Blum, Roth, and Rothblum's algorithm and a stable matching is obtained with probability one. In the general case, however, we have shown that in a market (F, W, P), given μ^I, different outcomes may be reached depending on the order in which firms propose. Furthermore, as the following example shows, unstable matchings may be obtained with positive probability.

Example 2 *An output matching may not be stable.*

Let (F, W, P) with $F = \{f_1, f_2\}$, $W = \{w_1, w_2\}$ and preferences such that

$$P(w_1) = f_2, \ f_1 \qquad P(f_1) = w_1$$
$$P(w_2) = f_2 \qquad P(f_2) = w_2, \ w_1.$$

Let the initial matching be $\mu^I = \{(f_2, w_1)\}$ and suppose f_1 is the first firm to make a proposal. Then, f_1 invites w_1, the only worker on its list of preferences and w_1 rejects this proposal, given that he is still holding f_2 (step 3(b)i). When f_2 is given its turn to move, it proposes to w_2. Since he is alone and f_2 is the only acceptable firm, w_2 accepts this offer (step 3(b)iiB) and the matching $\mu = \{(f_2, w_2)\}$ is obtained. It is easy to see that f_1 and w_1 block μ. $\diamond$

An execution of the DA^{μ^I}-algorithm with arbitrary input matching need not be stable. Further, any worker involved in instability of the output matching μ must have been matched under the input matching.[3] And, if some firm is part of a blocking pair for μ, it must have been rejected by the worker with whom it forms a blocking pair for μ along the execution of the algorithm.[4]

In the following results we describe some further characteristics of the output of the DA^{μ^I}-algorithm as a function of the initial matching μ^I. First, it is shown that if a worker ends up strictly worse off in the output matching, then there must be at least one worker that strictly improves his match. The only instance under which this can be violated is when the input matching is not individually rational.

Proposition 3 *Let (F, W, P) be a marriage market and $\mu^I \in IR(P)$. Let $\mu \neq \mu^I$ be*

---

[3] The instability of μ may be due to lack of individual rationality for some worker or to the existence of some blocking pair. In both cases, it is necessary that the worker involved is matched to a firm at μ^I; in particular, if μ is not individually rational for some worker, then μ^I cannot be individually rational either.

[4] In fact, the only instance under which a blocking pair may arise is when at some point a worker rejects a proposal from an acceptable firm, say f, because he is still holding the initial partner, ranked higher in his list of preferences. In this case, it may happen that the worker ends up being assigned to a firm he considers worse than f and, as a consequence, he will block the output matching together with f.

such that, for all $w \in W$, $\mu^I(w)R(w)\mu(w)$. Then, $\Pr\{\widetilde{DA}^{\mu^I}[P] = \mu\} = 0$.

Proof. By contradiction, let us suppose that, given an individually rational μ^I, a matching μ such that $\mu \neq \mu^I$ and $\mu^I(w)R(w)\mu(w)$ for all $w \in W$ is reached under some execution of the algorithm. This means that every worker weakly prefers the initial matching μ^I and that there exists at least one worker that strictly prefers it.

No unmatched worker would accept to fill a position in an unacceptable firm. Therefore, a worker who is strictly worse off in the output matching μ must have started matched. Moreover, he must have been fired by his initial partner. So, assume w_1 is the first worker to be fired by $\mu^I(w_1)$. This implies that either $\mu^I(w_1)$ fired w_1 to be alone or it proposed to another worker, say w_2, and he accepted. In the former case the individual rationality of μ^I is contradicted. In the latter case, since by assumption w_2 is still holding $\mu^I(w_2)$, we must have $\mu^I(w_1)P(w_2)\mu^I(w_2)$. By Remark 1, w_2 will never end up worse off in the output matching, contradicting the definition of μ. ∎

A slightly weaker result holds for the firms. An output matching where *every* firm is matched to a worker ranked lower than its initial partner in its preference list cannot be reached with positive probability. Example 3 shows that the requirement of having every firm strictly worse off in the output matching is necessary. Subsequently, we state the result.

Example 3

Let (F, W, P) be a matching market where P is given by:

$$P(w_1) = f_1, \ f_2 \qquad P(f_1) = w_2, \ w_1$$
$$P(w_2) = f_2, \ f_3, \ f_1 \quad P(f_2) = w_1, \ w_2$$
$$P(f_3) = w_2,$$

and let the input matching be $\mu^I = \{(f_1, \ w_2), (f_2, \ w_1)\}$. Every execution of the algorithm leads to the matching $\mu = \{(f_1, \ w_1), (f_2, \ w_2)\}$. In fact, for every order in which firms propose, when f_3 is given the opportunity to act, it makes a successful offer

to w_2, who is still holding f_1 at that point. Later, f_1 is forced to propose to w_1 and f_2 ends up matched to w_2. Hence, $\mu \neq \mu^I$ such that $\mu^I(f)R(f)\mu(f)$ for every $f \in F$ is reached with probability one. ◇

Proposition 4 *Let (F, W, P) be a marriage market, and let μ^I be an arbitrary input matching. Let μ be such that $\mu^I(f)P(f)\mu(f)$, for all $f \in F$. Then, $\Pr\{\widetilde{DA}^{\mu^I}[P] = \mu\} = 0$.*

Proof. Notice that if some firm is not matched at μ^I, then the result trivially holds, since no firm will ever propose to an unacceptable worker. So, let us assume every firm in F is matched under μ^I. The argument now follows by contradiction. Let μ be such that $\mu^I(f)P(f)\mu(f)$, for all $f \in F$ and assume that there is an execution that leads to μ.

Claim 1 The set of unmatched workers is the same under both μ^I and μ.

Proof. Notice that every worker who is initially assigned to a firm cannot end up alone in the output matching μ. Assume not and, without loss of generality, let us say w such that $\mu^I(w) \in F$ is unmatched under μ. This implies that $\mu^I(w)$ fired w. In addition, it follows from Remark 1, that no firm, including $\mu^I(w)$, proposed to w later on. But if this is so, $\mu^I(w)$ must end up matched to a worker ranked higher than w in its list of preferences. This contradicts the fact that $\mu^I(f)P(f)\mu(f)$, for all $f \in F$.

Claim 2 Every firm is matched under μ.

Proof. Immediate from Claim 1 and the fact that every firm starts matched.

Claim 3 An initially unmatched worker accepts no proposals along the execution.

Proof. This follows from Remark 1 and Claim 1.

Consider the last step at which a proposal is made by a firm f and accepted by a worker w. (Note that if no proposal is accepted along the execution, then $\mu = \mu^I$, contradicting the definition of μ.) At the last step of the algorithm, w must be unmatched when he accepts f's proposal. Otherwise, the firm held by w would be unmatched under

15

μ, which contradicts Claim 2.

By Claim 3, w must be matched under μ^I, let us say $\widehat{f} = \mu^I(w)$. Firm f is not w's initial partner, or else $\mu^I(f) = \mu(f)$, contradicting the definition of μ. By Claim 2 and given that we are considering the last step of the algorithm, $\widehat{f}$ is matched at this stage. Given that every firm is worse off under the output matching, it must be the case that $\widehat{f}$ is matched to a worker ranked lower than w in $P(\widehat{f})$. As a consequence, $\widehat{f}$ must have proposed to w and this proposal was rejected. By Remark 1, this implies that w is matched to a firm preferred to $\widehat{f}$ at this last step of the algorithm and we get another contradiction: w was not alone when he accepted f's proposal. ∎

4   The Game

We have so far informally described an algorithm in terms of the actions of the agents—proposals by the firms, and acceptances and rejections by the workers. Consider now a mechanism where agents face the single decision of submitting lists of preferences over prospective partners to a central clearinghouse, which uses this information to arrange a matching of workers to firms by means of the generalized deferred-acceptance algorithm. Clearly, in the game induced by this mechanism, agents may behave strategically: firms may choose not to reveal how they rank the workers in the market, or it may be sensible for workers to put forward other than their true ordering of positions. Therefore, we will now turn to a different class of questions, investigating how we may expect individuals to behave. In this section we discuss the strategic environment facing the agents in the revelation game induced by the DA^{μ^I}-algorithm.

Since we are dealing with a centralized market, the strategy space of a player in the game is confined to the set of all possible preference lists over the other side of the market. Hence, strategies will be represented by the corresponding preference profile—Q, for instance—while true preferences will always be denoted by P.

To address strategic questions we need to develop ideas about what constitutes a

"best decision" to be taken by an agent. With this purpose in mind, take two probability distributions over the set of matchings, $\widetilde{\mu}$ and $\widetilde{\mu}'$. Without loss of generality, consider $w \in W$ (what follows also holds for a representative firm, with the obvious modifications); $\widetilde{\mu}(w)$ and $\widetilde{\mu}'(w)$ denote the distributions induced over w's set of assignments by $\widetilde{\mu}$ and $\widetilde{\mu}'$, respectively. We say that $\widetilde{\mu}(w)$ first order stochastically $P(w)$-dominates $\widetilde{\mu}'(w)$ if $\Pr\{\widetilde{\mu}(w)R(w)v\} \geq \Pr\{\widetilde{\mu}'(w)R(w)v\}$, for all $v \in F \cup \{w\}$. Thus, for all $v \in F \cup \{w\}$, the probability of w being assigned to v or to a strictly preferred agent is higher under $\widetilde{\mu}(w)$ than under $\widetilde{\mu}'(w)$. Now, consider the problem that player w would face if the strategy choices Q_{-w} of the other players were known. In this case, any strategy $Q(w)$ by w would determine the probability distribution induced by the mechanism over the set of matchings. Therefore, a particular strategy choice $Q(w)$ is preferred if the induced probability distribution over the set of matchings stochastically dominates the one induced by any other alternative strategy.

Definition 2 *Given Q_{-w} and the preferences $P(w)$, we say that a strategy $Q(w)$ stochastically $P(w)$-dominates another strategy $\widehat{Q}(w)$ if, for all $v \in F \cup \{w\}$, $\Pr\{\widetilde{DA}^{\mu^I}[Q(w), Q_{-w}] (w)R(w)v\} \geq \Pr\{\widetilde{DA}^{\mu^I}[\widehat{Q}(w), Q_{-w}](w) \ R(w) \ v\}$. In a similar way, given Q_{-f} and the preferences $P(f)$, we define stochastic $P(f)$-dominance.*

In a problem like the one described here, each agent must make a decision without knowing the strategies of the others. It may happen that an arbitrary agent v has a strategy that is a best response to every profile of strategies that the other players may choose. In this case, we say v has a dominant strategy.

Definition 3 *Given an initial matching μ^I and the preferences $P(v)$, a dominant strategy for $v \in V$ is a strategy $Q(v)$ that, for every Q_{-v}, stochastically $P(v)$-dominates every alternative strategy $\widehat{Q}(v)$.*

In Example 1, we have shown that the outcome of the generalized deferred-acceptance algorithm may depend on the random order in which firms' lists are considered. Thus, the study of Nash equilibria in the game induced by the mechanism we have described would

require us to consider not merely agents' preferences over riskless outcomes, but also over lotteries. Since agents' preferences are ordinal and no natural utility representation of these orderings exists, we will adopt the following equilibrium notion.

Definition 4 *Given an initial matching μ^I and a profile of preferences P, the profile of strategies Q is an* ordinal Nash equilibrium *(ON equilibrium) if, for each player v in V, $Q(v)$ stochastically $P(v)$-dominates every alternative strategy $\widehat{Q}(v)$, given Q_{-v}.*

It is clear that we will be concerned in finding a profile of strategies Q with the property that once they are adopted by the agents, no one can profit by unilaterally changing his strategy; further, this is true for all possible utility representations of agents' preferences. This means that by using a strategy other than $Q(v)$, for any v' (an agent with whom it may end up matched), v will not be able to strictly increase the probability of obtaining v' and all agents ranked higher than v' in $P(v)$.

4.1 Strategic questions

In the revelation game induced by Gale and Shapley's DA-algorithm, straightforward behavior is not in every agent's best interest. This means that some agent may have an incentive to misrepresent its preferences. Given that the DA$^{\mu^I}$-algorithm replicates Gale and Shapley's when the initial matching is the empty matching, truth telling may not be an ordinal Nash equilibrium in the revelation game induced by the DA$^{\mu^I}$-algorithm.

Nevertheless, acting according to the true preferences is a dominant strategy for firms in Gale and Shapley's environment (Dubins and Freedman (1981) and Roth (1982)). So, in what firms are concerned, there is a clear sense in which honesty is the best policy under the DA$^{\mu^I}$-algorithm in the particular case that μ^I is the empty matching. Moreover, if μ^I is firm-quasi-stable, firms' true preferences remain a dominant strategy (Blum, Roth, and Rothblum (1997)). Unfortunately, as shown in the example below, truth is not a dominant strategy for firms when an arbitrary input matching is considered. Clearly, a firm will not benefit from using a truncation of its true preference list (i.e., a

strategy that, besides ranking the workers in the same way as the true preference relation, each of its acceptable workers is under the true preferences both acceptable and preferred to any worker which is unacceptable in the truncation strategy). Other manipulations, however, like ranking as acceptable an unacceptable worker, may be beneficial.

Example 4 *Revealing the true preferences is not a dominant strategy for all firms.*

Let (F, W, P) be a matching market with P given by:

$$P(w_1) = f_2 \qquad P(f_1) = w_2$$
$$P(w_2) = f_3, \ f_1 \qquad P(f_2) = w_1$$
$$P(w_3) = f_3 \qquad P(f_3) = w_3, \ w_2.$$

Let $\mu^I = \{(f_3, w_2)\}$. Let $Q(f_1) = w_1, w_2$ be an alternative strategy for f_1. Assume that every agent except for f_1 submits the true preferences. By using either $P(f_1)$ or $Q(f_1)$, f_1 may end up matched to w_2 or unmatched. Consider every sequence for which f_1 is unmatched under the output matching when using $Q(f_1)$, i.e., every sequence where f_1's second draw happens to be before f_3 is considered for the first time. Clearly, in these sequences, the first time f_1 appears is also before f_3, so that f_1 also ends up unmatched by using $P(f_1)$. However, consider, for instance, the sequence that starts with f_1, immediately followed by f_3. In this case, f_1 ends up matched to w_2 only if it acts according to $Q(f_1)$. Otherwise, by using $P(f_1)$, the first time f_1 is drawn and its willingness to match w_2 is taken into account, w_2 is still holding f_3. Since w_2 prefers f_3 to f_1, this worker keeps f_3 and f_1 ends up unmatched. It follows that f_1 profits by deviating from its true preferences. ◇

4.2 Ordinal Nash equilibria

We have observed that faithfully transmitting the true preferences is not necessarily an ordinal Nash equilibrium. Therefore, we need to ask whether ordinal Nash equilibria always exist in the revelation game induced by the DA^{μ^I}-algorithm. Proposition 5 will

show that they do: when μ^I is individually rational, every element of a non-empty subset of $IR(P)$ can be sustained in equilibrium with probability one.

Definition 5 *Let μ^I be an arbitrary matching. We say that μ is individually rational with respect to μ^I if $\mu \in IR(P)$ and if, for all $f \in F$, $w' = \mu^I(f)P(f)\mu(f)$, implies $\mu(w') \neq w'$.*

We will denote by $IR^{\mu^I}(P)$ the set of all individually rational matchings with respect to μ^I. This set is always non-empty since it includes $S(P)$, the set of stable matchings (Pais (2004)).

Proposition 5 *Let μ^I be an individually rational matching for (F, W, P) and let $\mu \in IR^{\mu^I}(P)$. Then, there exists an ordinal Nash equilibrium Q in the revelation game induced by the DA^{μ^I}-algorithm that leads to μ. Furthermore, $\Pr\{\widetilde{DA}^{\mu^I}[Q] = \mu\} = 1$.*

Proof. Define $Q(v) = \mu(v)$, for all $v \in V$. It is clear that every play of the game with the profile Q will lead to the output matching μ. Thus, $\Pr\{\widetilde{DA}^{\mu^I}[Q] = \mu\} = 1$.

Let us show that for every firm f, $Q(f)$ is a best reply to Q_{-f}. First, as long as $\mu(f) \neq \mu^I(f)$, f never holds its initial match under μ. Indeed, it is clear that if $\mu^I(f)P(f)\mu(f)$, then $\mu^I(f)$ receives and accepts another firm's proposal (and in the case that $\mu(f)P(f)\mu^I(f)$, $\mu^I(f)$ is not a temptation). Hence, when $\mu(f) \in W$, given that the only worker willing to accept f's proposal is $\mu(f)$, the only choice f can actually make is between being assigned to this worker or staying alone. From individual rationality we have $\mu(f)P(f)f$ which implies that f will not be able to profit from deviating from $Q(f)$. Obviously, for f such that $\mu(f) = f$, no worker accepts f's proposal and it can do no better than staying alone.

Finally, for any w, $Q(w)$ is a best reply to Q_{-w}. In fact, given firms' strategies, w gets at most one proposal and, considering μ is individually rational, the best he can do is to accept it. This completes the proof. ∎

Although the strategies used can be seen as an amazing act of coordination, they serve the purpose of finding a sufficient condition for ordinal Nash equilibrium outcomes. In what necessary conditions for equilibrium are concerned, it is obvious that every output matching reached with positive probability in equilibrium must be individually rational with respect to true preferences. Furthermore, in the result that follows, we will show that *some* stability is preserved in every ordinal Nash equilibrium.

Theorem 1 *Let μ^I be an individually rational input matching for $(F, W, (Q_F, P_W))$. Assume that the strategy profile Q is an ordinal Nash equilibrium in the revelation game induced by the DA^{μ^I}-algorithm. Then, the probability distribution obtained over the set of matchings is such that every element in its support is a member of $S(Q_F, P_W)$.*

Proof. Suppose that $\{\mu_1, ..., \mu_k\}$ is the support of the distribution induced by the DA^{μ^I}-algorithm over the set of matchings. Assume that for some $i \in \{1, ..., k\}$, $\mu_i \notin S(Q_F, P_W)$. We will prove that Q is not an ON equilibrium.

To start, notice that for every firm f it must be the case that its assignment, $\mu_i(f)$, is individually rational with respect to $Q(f)$, as this is the strategy firm f is using. On the other hand, individual rationality with respect to P must hold for every worker. Assume that this is not the case and that there exists a worker, say w, such that $wP(w)\mu_i(w)$. Individual rationality of the matching μ^I implies $\mu_i(w) \neq \mu^I(w)$. Hence, w must have, at some point, accepted $\mu_i(w)$'s proposal. This means that under $Q(w)$ we have $\mu_i(w)Q(w)w$. Now take an alternative strategy $\widetilde{Q}(w)$ in which all firms are considered unacceptable, meaning that no offer is accepted by w. By using $\widetilde{Q}(w)$, w may end up unmatched or matched to his original firm $\mu^I(w)$, but he is never assigned to a firm considered unacceptable under $P(w)$. Thus, the following holds:

$$1 = \Pr\{\widetilde{DA}^{\mu^I}[\widetilde{Q}(w), Q_{-w}](w)R(w)w\} > Pr\{\widetilde{DA}^{\mu^I}[Q](w)R(w)w\}$$

and $Q(w)$ is not a best reply to Q_{-w}.

We have proved that μ_i is individually rational. Thus, there must exist a blocking pair for μ_i when the preference profile (Q_F, P_W) is considered. Let us say (f, w) blocks μ_i,

i.e., $fP(w)\mu_i(w)$ and $wQ(f)\mu_i(f)$. This implies that f proposed to and was rejected by w in the course of every execution leading to μ_i. By Remark 1, either $\mu_i(w)Q(w)f$ (case (i)) or, if not, w must have rejected f while he was still holding $\mu^I(w)$ and $\mu^I(w)Q(w)f$ (case (ii)).

(i) Assume $\mu_i(w)Q(w)f$. We will prove that $Q(w)$ is not a best reply to Q_{-w}. Define $\widetilde{Q}(w)$ that preserves the same ordering as in $Q(w)$, except that f holds the first position under $\widetilde{Q}(w)$. Formally, for all $v, \widehat{v} \in (F\backslash\{f\})\cup\{w\}$, $[v\widetilde{Q}(w)\widehat{v} \iff vQ(w)\widehat{v}]$ and $f\widetilde{Q}(w)v$.

Let us prove that the probability of being assigned to f is strictly higher under $\widetilde{Q}(w)$ than under $Q(w)$. We know that in a path leading to μ_i, firm f must have proposed to w. If, instead of using $Q(w)$, w deviates and acts according to $\widetilde{Q}(w)$, w holds f until the algorithm stops. Thus, every order that originally lead to μ_i results in an output matching where f and w are together. If, under $Q(w)$, $\Sigma_{f,w} = \emptyset$, so that f and w are never matched under the original strategy profile, then the probability of having f and w matched is strictly increased when w deviates. Otherwise, for $\Sigma_{f,w} \neq \emptyset$, by moving f up in the ranking of w's preferences, f is still assigned to w for every element of $\Sigma_{f,w}$. Indeed, under any such order of offers, f proposes to w, whether w is using $Q(w)$ or $\widetilde{Q}(w)$, and in both cases w accepts this offer. Hence, the probability of having f and w matched is also strictly increased when w uses $\widetilde{Q}(w)$.

In order to prove $Q(w)$ is not a best reply to Q_{-w}, assume, without loss of generality, that $P(w) = f_1, f_2,..., f_{m-1}, f, f_{m+1},..., w,..., f_n$. Consider a firm f_j, with $j = 1, ...,$ $m-1$, and consider $\Sigma_{f_j,w}$ when $Q(w)$ is used. It cannot be guaranteed that every element in $\Sigma_{f_j,w}$ still gives f_j assigned to w when he deviates and acts according to $\widetilde{Q}(w)$. Clearly, if f_j is ranked below f in $Q(w)$, no change occurs. If f_j is ranked higher than f, for all the orders in $\Sigma_{f_j,w}$ that involved f proposing w at some step of the algorithm, by using $\widetilde{Q}(w)$, w now holds f's proposal until the end. Thus, for every element of $\Sigma_{f_j,w}$, w either ends up matched with f_j or with f. Hence,

$$\Pr\{\widetilde{DA}^{\mu^I}[\widetilde{Q}(w), Q_{-w}](w)R(w)f\} > Pr\{\widetilde{DA}^{\mu^I}[Q](w)R(w)f\},$$

contradicting that Q is an ON equilibrium.

(ii) Now take the case in which $\mu^I(w)Q(w)fQ(w)\mu_i(w)$ (notice $f \neq \mu_i(w)$, otherwise f and w could not block μ_i). Define the deviation, $\widetilde{Q}(w)$, as before. Under $\widetilde{Q}(w)$, w accepts f at any step of the algorithm and hold its offer until the end. Then, it is obvious that the chances of having f matched to w in the final output increase—at least—in the probability of all orders of proposals that originally lead to μ_i.

Again, suppose $P(w) = f_1, f_2,..., f_{m-1}, f, f_{m+1},..., w,..., f_n$. Using the same argument as before, we can guarantee that for any order of proposals that gives w matched to any firm f_j, $j = 1, ..., m - 1$, by acting according to $\widetilde{Q}(w)$, w will either be assigned to f or to f_j. Once more, it is true that $Q(w)$ is not a best reply to Q_{-w} as

$$\Pr\{\widetilde{DA}^{\mu^I}[\widetilde{Q}(w), Q_{-w}](w)R(w)f\} > Pr\{\widetilde{DA}^{\mu^I}[Q](w)R(w)f\}.$$

This completes the proof. ∎

An immediate implication of this result is worth noticing. As proved in McVitie and Wilson (1970) and Roth (1982), in a market (F, W, P) with strict preferences, the set of unmatched agents is the same for all stable matchings. Hence, for any two matchings that arise with positive probability under an ordinal Nash equilibrium, the set of unmatched agents is the same—when agents act strategically, no one can hold chance responsible for ending up unmatched. This provides a further step towards describing ordinal Nash equilibria.

The following result is an important special case of Theorem 1.

Corollary 1 *Let μ^I be an individually rational input matching for (F, W, P). Assume (P_F, Q_W) is an ordinal Nash equilibrium in the revelation game induced by the DA^{μ^I}-algorithm. Then, the probability distribution obtained over the set of matchings is such that every element in its support is a member of $S(P)$.*

Proof. Immediate from Theorem 1 with $Q_F = P_F$. ∎

Remarkably, in any equilibrium in which firms play straightforwardly stability with respect to true preferences is recovered. This result generalizes a known feature of

the game induced by Gale and Shapley's mechanism (Roth (1984)), as well as a result obtained by Blum, Roth, and Rothblum (1997) with a firm-quasi-stable matching as an input. Focusing on truth telling is easily justifiable. In some settings, sophisticated strategic play by one side of the market does not even make sense (e.g., universities select students according to their grades). Also, in an environment where agents do not know how the others will play and given the multiplicity of available strategies, acting according to the true preferences can be seen as an easy resort.

When the initial matching is empty, any stable matching can result from some equilibrium where firms play according to their true preferences (Gale and Sotomayor (1985)). Thus, a group of workers with more than one achievable outcome can reveal preferences to compel any jointly achievable outcome. Moreover, Blum, Roth, and Rothblum (1997) have shown that this result can be generalized to a game that starts at a firm-quasi-stable matching as long as agents must use strategies that are identifiable with preference lists. It is no longer the case that every stable matching can be reached; what happens is that any jointly achievable outcome for the workers that are unmatched at μ^I can result from an equilibrium in which firms use their true preferences. In the next proposition we extend these results.

Definition 6 *Let $\mu \in S(P)$. Let μ^I be an arbitrary matching. We say that μ is stable with respect to μ^I if, for all $f \in F$ such that $\mu^I(f)P(f)\mu(f)$, we can define a non-empty subset of firms $\widehat{F}(f) = \{f_1, f_2, ..., f_r\}$, $r \leq n$, for which the following conditions hold:*

1. $\mu(f_{i+1}) = \mu^I(f_i)$, for all $i = 1, ..., r - 1$, and $\mu(f_1) = \mu^I(f_r)$;

2. $\mu(\mu^I(f)) \in \widehat{F}(f)$;

3. $\mu(f_i)P(f_i)\mu^I(f_i)$, for some $i = 1, ..., r$.

Let $S^{\mu^I}(P)$ be the set of all stable matchings with respect to μ^I. This set may be empty, as the following example shows.

Example 5 *(Example 3 continued)*

In the matching market of Example 3, the only stable matching is $\mu = \{(f_1, w_1), (f_2, w_2)\}$. Comparing μ with the initial matching $\mu^I = \{(f_1, w_2), (f_2, w_1)\}$, it is clear that no firm is strictly better off under μ than under μ^I. Hence, condition 3 is not fulfilled and $S^{\mu^I}(P)$ is empty. $\diamond$

We will show that, when μ^I is individually rational and $S^{\mu^I}(P)$ is non-empty, there is an ordinal Nash equilibrium where firms tell the truth leading to each element of $S^{\mu^I}(P)$. As it will become clear when the equilibrium strategies are described, a lot of coordination is still needed to achieve a particular equilibrium.

Proposition 6 *Let μ^I be an individually rational input matching for (F, W, P). Let $\mu \in S^{\mu^I}(P)$. Then, there exists an ordinal Nash equilibrium (P_F, Q_W) in the revelation game induced by the DA^{μ^I}-algorithm that leads to μ. Moreover, $\Pr\{\widetilde{DA}^{\mu^I}[P_F, Q_W] = \mu\} = 1$.*

Proof. Define $Q(w) = \mu(w)$, for all $w \in W$. Let us start by showing that the profile of strategies (P_F, Q_W) always leads to the matching μ, i.e., $\Pr\{\widetilde{DA}^{\mu^I}[P_F, Q_W] = \mu\} = 1$. If this is not the case, then there exists an order of proposals leading to $\widehat{\mu} \neq \mu$. But this is equivalent to having a firm, say f, whose partner, $\widehat{\mu}(f)$, is different from $\mu(f)$ after some execution of the algorithm. Given the strategies of the workers, we can either have $\widehat{\mu}(f) = f$—when $f \neq \mu(f)$—or $\widehat{\mu}(f) = \mu^I(f)$—if $\mu^I(f) \neq \mu(f)$. To start, assume that $\widehat{\mu}(f) = f$. Since $\mu(f)$ would accept f's proposal and f is acting according to its true preferences, it must be the case that $fP(f)\mu(f)$. But this contradicts the stability of μ. Now suppose that $\widehat{\mu}(f) = \mu^I(f)$, with $\mu^I(f) \neq \mu(f)$. Again, given f's strategy, we must have $\mu^I(f)P(f)\mu(f)$. Besides, $\mu^I(f)$ cannot be matched under μ. Otherwise, he would receive and accept a proposal from its assignment at μ (notice that from the definition of $S^{\mu^I}(P)$ there exists $\widehat{f} \in \widehat{F}(f)$ such that $\mu(\widehat{f})P(\widehat{f})\mu^I(\widehat{f})$, guaranteeing that such a proposal would actually be made). So assume that $\mu^I(f)$ is unmatched at μ. However, we know that $fP(\mu^I(f))\mu^I(f)$ by individual rationality of μ^I. Also, as μ is stable, $\mu^I(f)$ must prefer to be matched to its partner at μ, rather than staying with f, i.e., $\mu(\mu^I(f))P(\mu^I(f))f$. Thus, we have $\mu(\mu^I(f)) \neq \mu^I(f)$ and, once more, we obtain a contradiction.

25

Let us now prove that, for every firm f, $P(f)$ stochastically $P(f)$-dominates every other strategy $Q(f)$. We will consider the most general case, assuming that $\mu^I(f), \mu(f) \in W$ and $\mu^I(f) \neq \mu(f)$ (the proofs for other cases follow easily from this one). Given that the only worker who is willing to accept f is $\mu(f)$, by choosing its strategy appropriately, f can either be alone, hold $\mu(f)$ or, eventually, remain with $\mu^I(f)$ under the output matching. By stability of μ, $\mu(f)P(f)f$. If, additionally, $\mu(f)P(f)\mu^{I\prime}(f)$, firm f can do no better than obtaining $\mu(f)$ and truth telling guarantees $\mu(f)$ is assigned to f with probability one. Otherwise, if $\mu^I(f)P(f)\mu(f)$, f is not able to retain $\mu^I(f)$. In fact, given the definition of $S^{\mu^I}(P)$, $\mu^I(f)$ is matched to some firm under μ and obtains a proposal from this firm. Thus, f cannot do better than being assigned to $\mu(f)$ and $P(f)$ stochastically $P(f)$-dominates every other strategy $Q(f)$.

Now take the case of an arbitrary worker, w. Suppose, by way of contradiction, that $Q(w)$ does not stochastically $P(w)$-dominate a different strategy $\widehat{Q}(w)$. This implies that $\Pr\{\widetilde{DA}^{\mu^I}[P_F, \widehat{Q}(w), Q_{-w}](w)R(w)\mu(w)\} = 1$ and that there exists a firm, say f, such that the following holds: $fP(w)\mu(w)$ and $\Pr\{\widetilde{DA}^{\mu^I}[P_F, \widehat{Q}(w), Q_{-w}](w) = f\} > 0$. But this means that, for some order of proposals, f approaches w before making an offer to $\mu(f)$. In fact, it cannot be the case that f proposes to $\mu(f)$ first and he does not accept it, as $\mu(f)$ is acting according to his original strategy, $Q(\mu(f))$. Thus, f must prefer w to $\mu(f)$. However, in this case (f, w) forms a blocking pair for μ, contradicting the fact that μ is stable. ∎

Proposition 6 showed that there are ordinal Nash equilibria at which firms reveal their true preferences and the output is stable for the true preferences. These equilibria involve misrepresentation by the workers. Further, by misstating their preferences "appropriately," workers can compel the best achievable stable matching. However, as the following example shows, the above proposition does not exhaust all ordinal Nash equilibria.

Example 6 *(Example 3 continued) There may be more ordinal Nash equilibria than those given in Proposition 6.*

Recall that in the matching market in Example 3, when $\mu^I = \{(f_1,\ w_2),\ (f_2,\ w_1)\}$ is considered, every execution of the algorithm with P leads to $\mu = \{(f_1,\ w_1),\ (f_2,\ w_2)\}$. Under μ, workers obtain the best possible positions and firms cannot improve by deviating. No manipulation will enable f_1 and f_2 to keep the workers they hold under μ^I, given the presence of f_3. As a result, P is an ordinal equilibrium, even though $S^{\mu^I}(P)$ is empty. $\diamond$

5   Concluding remarks

In this paper we have tried to extend the theoretical analysis of two-sided matching models, by describing a mechanism that generalizes the original deferred-acceptance algorithm proposed by Gale and Shapley (1962). In fact, we consider matching beginning from arbitrary input matchings instead of just from the empty matching, under which all candidates and positions are available. Furthermore, we have shown that the outlined mechanism encompasses Blum, Roth, and Rothblum's, in the particular case that we start from a firm-quasi-stable matching (a stable matching destabilized by the entry of a firm or the retirement of a worker).

The strategic decisions facing players were also considered, in a revelation game that follows the rules laid out by the algorithm at hand. The uncovered results extend those obtained for the Gale and Shapley's DA-algorithm. It is shown that in general truth revealing behavior is not an equilibrium, but that there may be equilibria at which firms behave straightforwardly. A class of equilibria is described in which this side of the market plays according to the true preferences and, although the workers need not be frank about their preferences, outcomes are stable. Nevertheless, some of the presented equilibria are unlikely to be observed in reality. In fact, the strategies described for the workers require a lot of coordination among them and the multiplicity of equilibria gives no clue to the form that a sensible strategy should have. A perhaps more serious drawback of this analysis concerns truth telling by firms. How plausible is straightforward behavior by firms is a question to be explored. A natural direction to

pursue further research will be into characterizing equilibria in a more precise way, in particular equilibria where firms are not restricted to truth telling. It was shown that a good part of the individually rational matchings can be obtained as a result of an equilibrium play and that every equilibrium output obeys some form of stability.

In closing, when describing the algorithm, we have assumed that only one side of the market—firms, to be precise—can actually make proposals. However, some of the above results can be extended to a mechanism in which, at each step, an arbitrarily chosen agent—firm or worker—is selected to make a proposal. It turns out that, starting from an arbitrary matching, every ordinal Nash equilibrium outcome must be individually rational. Conversely, every individually rational output matching can be obtained with probability one in equilibrium. Finally, in what equilibria where one side of the market tells the truth are concerned, every stable matching that agents belonging to the truthful side of the market weakly prefer to the initial matching can be sustained as the unique outcome of an equilibrium play.

References

BLUM, Y., A. E. ROTH, AND U. G. ROTHBLUM (1997): "Vacancy Chains and Equilibration in Senior-Level Labor Markets," *Journal of Economic Theory*, 76, 362–411.

DUBINS, L. E., AND D. A. FREEDMAN (1981): "Machiavelli and the Gale-Shapley Algorithm," *American Mathematical Monthly*, 88, 485–494.

EHLERS, L. (2003): "In Search of Advice for Physicians in Entry-Level Medical Markets," Mimeo, Université de Montréal.

——— (2004): "In Search of Advice for Participans in Matching Markets which Use the Deferred-Acceptance Algorithm," *Games and Economic Behavior*, 48, 249–270.

EHLERS, L., AND J. MASSÓ (2003): "Incomplete Information and Small Cores in Matching Markets," Mimeo, Universitat Autònoma de Barcelona.

GALE, D., AND L. S. SHAPLEY (1962): "College Admissions and the Stability of Marriage," *American Mathematical Monthly*, 69, 9–15.

GALE, D., AND M. A. O. SOTOMAYOR (1985): "Ms Machiavelli and the Stable Matching Problem," *American Mathematical Monthly*, 92, 261–268.

MCVITIE, D. G., AND L. B. WILSON (1970): "Stable Marriage Assignment for Unequal Sets," *BIT*, 10, 295–309.

PAIS, J. (2004): "Incentives in Decentralized Random Matching Markets," Mimeo, Universitat Autònoma de Barcelona.

ROTH, A. E. (1982): "The Economics of Matching: Stability and Incentives," *Mathematics of Operations Research*, 7, 617–628.

——— (1984): "Misrepresentation and Stability in the Marriage Problem," *Journal of Economic Theory*, 34, 383–387.

ROTH, A. E., AND U. G. ROTHBLUM (1999): "Truncation Strategies in Matching Markets - In Search of Advice for Participants," *Econometrica*, 67, 21–43.

ROTH, A. E., AND M. A. O. SOTOMAYOR (1990): *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Cambridge University Press, Cambridge, England.

SOTOMAYOR, M. A. O. (1996): "A Non-Constructive Elementary Proof of the Existence of Stable Marriages," *Games and Economic Behavior*, 13, 135–137.