JOURNAL OF ECONOMIC THEORY **76**, 362–411 (1997)
ARTICLE NO. ET972307

Vacancy Chains and Equilibration in Senior-Level Labor Markets*

Yosef Blum,[†] Alvin E. Roth,[‡] and Uriel G. Rothblum[†]

[†]*Faculty of Industrial Engineering and Management, Technion—Israel Institute of Technology, Haifa 32000, Israel; and* [‡]*Department of Economics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260*

Received May 31, 1995; revised February 27, 1997

In contrast to entry-level professional labor markets, in which cohorts of candidates and positions become available at the same time (e.g., when candidates graduate from school), senior level positions typically become available when an incumbent retires, or a new position is created, and when a senior position is filled a new vacancy is often created elsewhere. We model senior level labor markets as two-sided matching markets in which matchings are destabilized by retirements and new entries, and can return to stability by a decentralized process of offers and acceptances. This generalizes the standard analysis in a way which has points of contact with the sociological literature on vacancy chains. *Journal of Economic Literature* Classification Numbers: D00, C78, J63. © 1997 Academic Press

1. INTRODUCTION

Labor markets for professionals function differently at the entry level and at the senior level. Candidates for entry level positions typically become available to begin work in cohorts which attain their professional qualifications at the same time, e.g., as they graduate from medical, law, or business school, or university. The behavior of firms seeking to hire at the entry level is typically adapted to this fact, so that entry level positions also become available in cohorts. So entry level markets are characterized by

* This work was partially supported by NSF Grant SES-9121968. We also are grateful to have received the then unpublished paper by Marilda Sotomayor which motivated the approach taken here and for very constructive suggestions of an anonymous referee (see footnote 14). We also acknowledge helpful discussions with Masaki Aoyagi and Phil Reny about what we have called Realization Independent Equilibria.

0022-0531/97 $25.00

the simultaneous availability of vacant positions and candidates who are in need of a position[1].

However senior level positions, which become available when an incumbent vacates a position (e.g., by retirement) or when a new position is created, are often filled with candidates who are themselves incumbents in other, similar positions. A firm with a vacancy may hire an incumbent from another firm, which then must fill its own newly created vacancy, creating a chain of vacancies that propagate from firm to firm, ending only when a position remains unfilled, or when a position is filled by a candidate who is not presently an incumbent in a comparable position.

For this reason, entry level markets return to equilibrium following the entry of a new cohort differently than senior level markets return to equilibrium after the creation of new vacancies.

Substantial progress has been made in studying how equilibrium is reached in entry level markets, using two-sided matching models of the kind first formally studied by Gale and Shapley [12]. They showed that a simple *deferred acceptance* algorithm, in which firms proposed to workers who could hold at most one unrejected offer at any time, would always produce a *stable* matching between firms and workers, in which no firm and worker who were not matched to each other would mutually prefer to be matched to one another. A considerable theoretical literature on two-sided matching has since emerged [32]. These models have in turn made possible the empirical study of the evolution of market institutions found in many entry level labor markets.[2] Indeed some markets have even developed centralized market clearing institutions which produce stable matchings through variations on the deferred acceptance algorithm, while the evolution of particular centralized or decentralized institutions in many other markets turns out to be at least partially related to how well they succeed in producing stable matches [26, 28–30, 35, 36] for discussions of entry level labor markets in medicine, psychology, law, and business, in the United States, Great Britain, Canada, and Japan).[3]

[1] This does not mean that the recruitment and hiring of entry level candidates may not become diffused or advanced in time. Just such timing phenomena, and their consequences for market organization, are explored in Roth and Xing [35].

[2] Crawford [8] and Crawford and Knoer [9] make clear why such models are particularly well suited to labor markets. Matching models have also been applied to non-labor-market matching processes such as sorority rush (Mongell and Roth, [23]) and marriage (Bergstrom and Bagnoli, [6]; Pollak, [24]). Finally, the matching literature and the present paper are part of a general growth in the application of game theoretic tools to labor economics, of which some admirable recent examples are Gibbons and Katz [14], Gibbons and Murphy [15], and Holmstrom and Milgrom [18].

[3] Interestingly, the appearance of deferred acceptance algorithms in labor markets preceded their discussion in the theoretical literature by at least 10 years—see Roth [26] for the demonstration that a medical labor market institution dating from the early 1950's is equivalent to the deferred acceptance algorithm.

It seems likely that centralized market institutions will turn out to be rarer in senior labor markets than in entry level markets, and that those centralized institutions that do exist will function differently.[4]

This paper extends the theory of two-sided matching models in a way we hope will provide a framework which can be applied to senior level (and decentralized) labor markets. We will show how a market may regain stability after a stable matching is disrupted, and what this implies about the market process, and about the strategic problems facing agents on both sides of the market. This will also allow us to make contact with research on senior-level labor markets which has grown up in the sociology literature, following the work of White [39], who studied vacancy chains as Markov processes, and applied the Markov model to the market for pastors of Protestant churches.

The applied work on vacancy chains in the sociology literature focuses on labor markets in which each firm employs exactly one senior worker, in order that vacancies can be unambiguously identified. Its emphasis is on grouping the positions into "strata," estimating the probability that a vacancy from one stratum will be filled with an incumbent from each other stratum, and comparing the observed and predicted length of chains initiated by vacancies in each stratum. For example, White found that vacancy chains which were initiated with the retirement of a pastor of a large church were longer than those initiated by vacancies in small churches, as positions in large churches were often filled by incumbents in other large or medium size churches, while vacancies in small churches were often filled by the promotion of an assistant pastor (possibly from another church). Similarly, in the market for head football coaches of American college teams [37], the longest vacancy chains are those which start with a vacancy for an NCAA Division IA team, as these are most likely to be filled with someone who is already a head coach at another college, while less prestigious vacancies are more often filled with the promotion of an assistant coach.[5]

One question which has not been addressed in the sociology literature on vacancy chains is "What is the source of the randomness which is being

[4] Granovetter [16, pp. 21–22] discusses the centralized market institutions for Conservative rabbis in the United States. In Granovetter [17] he remarks about this market: "This situation is different from those described by Roth and Xing (1994) in two important ways: (1) The placement procedure applies to all vacancies, none of which are explicitly defined as 'entry-level', and thus the distribution over time of vacancies is not sharply spiked, as it would be for new medical interns, but spread out in some stochastic way; and (2), closely related to (1), rabbis at all career stages, not just entry level, may be interested in new positions."

[5] One of the more unusual markets to have been studied from this point of view is the late 19th century market for superintendents of lunatic asylums [1]. The data on initial vacancies were culled in part from the appropriately named *American Journal of Insanity*.

estimated?" Because the analyst does not observe the preferences that employers have for individuals and individuals have for jobs, these appear random. (Indeed, no individuals of any sort appear in these sociological models.) But another possible source of randomness lies in the details of the market: e.g., in a year when several senior positions become vacant, which incumbents fill which vacancies might depend on the order in which firms make offers. We will show that in fact such unobserved market details need not contribute *any* randomness to the outcome. Under the re-equilibration processes we study, and given the (strict) preferences of the agents, the stable matching which is reached when the process is allowed to go to completion is independent of the ordering of offers.[6]

To explore the mathematical relationship between vacancy chain models and two sided matching models, we will consider one of the simplest generalizations of the original *marriage* model introduced by Gale and Shapley [12]. In this model there are two finite sets of agents, called *firms* and *workers*.[7] Each agent has a preference relation over the members of the opposite set, and over the possibility of remaining unmatched. An outcome of the market is a matching of firms and workers (which allows the possibility that some firms and workers may remain unmatched).[8] We will take the rules of the market to be the usual ones, that a firm may offer employment to any worker it wishes, who is free to accept or reject the offer. That is, a firm and worker may be matched to one another if both agree, but every agent is free to remain unmatched. Under these rules, an outcome of the market, i.e., a matching, is *stable* if no agent is matched to one whom he finds unacceptable (i.e. to whom he finds it preferable to be unmatched), and if there are no *blocking pairs*, consisting of a firm and a worker who each prefer the other to their outcome under the given matching.

[6] Of course a random element would be introduced if agents did not have strict preferences and made choices randomly when they were indifferent between two alternatives. A random element can also be introduced if the re-equilibration process is prematurely terminated. (For example, if the process is terminated in so short a time that only one offer can be made, it will obviously be important which firm gets to make that offer.) Roth and Xing [36] study a related sort of premature termination, which arises in a way which suggests it may be a common phenomenon. In such markets, the present results provide a benchmark against which the consequences of premature termination can be measured.

[7] It is called the marriage model because matching is one to one, and for this reason the agents are sometimes referred to as men and women, instead of firms and workers. Gale and Shapley's original model did not allow agents to remain unmatched, as we do here.

[8] To keep matters simple we consider the salary associated with each position to be a fixed part of the job description, rather than something to be negotiated separately between each firm and worker. This is of course more descriptive of some markets (e.g., civil service positions and parsonages) than of others (e.g., head football coaches). But two-sided matching models of this sort are well equipped to handle negotiable wages as well as fixed wages (cf. Roth and Sotomayor, [32]), albeit at some additional complexity and notational burden.

Gale and Shapley [12] showed that the deferred acceptance algorithm transforms the empty matching, in which all agents are unmatched, into a stable matching. In the deferred acceptance algorithm, each firm initially makes an offer to its most preferred worker, and following any rejection makes an offer to its most preferred worker who has not yet rejected it. Each worker considers each offer in light of the offers already received, and "holds" the best of these, while rejecting the others (and always rejects any unacceptable offers). Thus, the deferred acceptance algorithm can be viewed as the path of play of a certain kind of game, provided that the firms and workers play in a way which is consistent with their true preferences. It turns out that this straightforward kind of play is not always equilibrium play in such a game, but that, nevertheless, every equilibrium in undominated strategies produces an outcome that is stable not merely with respect to the revealed preferences, but also to the true preferences [32, Theorems 4.4 and 4.16].

Consider now a market which has achieved a stable matching, and which is then disrupted by the retirement of some workers or the entry of new (and unmatched) firms. The matching which results will in general no longer be stable (with respect to the agents now in the market), because there will be some unmatched firms, i.e., firms with vacant positions, which would like to hire workers. If there is a worker who would prefer to be hired by such a firm (rather than remain in her current position, or lack of one), then the firm and worker will form a blocking pair. But the previous stability of the matching, before it was disrupted, will insure that any blocking pairs will involve unmatched firms. We will call such matchings *firm-quasi-stable*, and this paper will be devoted to studying how markets can regain stability after experiencing disruptions which result in firm-quasi-stable matchings.[9]

We will consider markets in which, when a firm has a vacancy, it seeks to fill it by offering the vacant position to the worker it prefers most, who considers the offer in light of the position she already has, and the other offers she receives. When a worker already has a position (or offers) which she prefers, she will reject this new offer, but otherwise will tentatively accept it (pending the possible arrival of even better offers).

Processes of this sort have previously been examined in the matching literature. Given an acceptable matching μ' with blocking pair (f, w) we say that a matching μ is *obtained from* μ' *by satisfying the blocking pair*

[9] The idea of firm-quasi-stability was first introduced by Sotomayor [38] when presenting a new, non-constructive proof for the existence of a stable matching. The effects of adding new entrants have been considered in earlier work on matching. See in particular Kelso and Crawford [19], Gale and Sotomayor [13], Mo [22], Roth and Sotomayor ([32], Chap. 2), and Bennett [5]—the intersection between those papers and the present one will be greatest in Section 5.

(f, w) if f and w are matched to each other in μ, their mates (if any) in μ' are unmatched in μ, and the status of all other individuals remains unchanged. Knuth [20] demonstrated that starting with an arbitrary matching and iteratively satisfying blocking pairs need not lead to a stable matching. Roth and Vande Vate [33], however, showed that it is always *possible* to reach a stable matching from any arbitrary matching by satisfying a sequence of blocking pairs. Abeledo and Rothblum [3] identified a broad class of methods which accomplish this task.

Such a process leads to vacancy chains, since as one firm succeeds in filling its vacancy it may cause another firm to have one. We will see that when, starting from a firm-quasi-stable matching, the process we consider is allowed to run to completion, it terminates at a stable matching. Furthermore, the stable matching which results is completely determined by the preferences of the agents, together with the particular firm-quasi-stable matching at which the process starts (which is in turn determined by the initial stable matching and by which workers have retired).

The process of re-equilibration we consider coincides with the deferred acceptance algorithm when we start from the matching at which all positions are vacant, and here we show that it allows us to start from any firm-quasi-stable matching. We will see that in the decentralized game in which agents may behave strategically (e.g., in which firms may choose to give offers to other than their most preferred workers, or workers may hold other than their most preferred offer) it is nevertheless the case that a natural class of equilibria produce outcomes that are stable with respect to the true preferences. Thus our analysis will suggest that the process by which senior level labor markets can regain stability following the creation of vacancies may have much in common with the way entry level markets achieve stable matchings with new cohorts of entry level workers. However we will also identify some important differences, both between entry and senior level matching and between centralized and decentralized matching at any level.

This paper is organized as follows. Section 2 introduces the formal matching model, and reviews some results on stable matchings. Section 3 explores applications of the deferred acceptance algorithm to arbitrary matchings. Section 4 introduces *firm-quasi-stable* matchings, and shows that these are precisely the matchings which result from stable matchings when vacancies appear due either to the retirement of incumbents or to the creation of new positions. It also shows that, starting with any firm-quasi-stable matching, the deferred acceptance algorithm defines a process by which stability can be achieved. Section 5 explores the relationships among the stable matching in an initial market, the firm-quasi-stable matching which results from retirements and new entries, and the stable matching which is achieved in the new market. Section 6 begins the study of the strategic environment facing the firms and workers, by defining a game which models the decentralized process by which firms offer positions and workers accept or reject them. In this game, straightforward (preference revealing) behavior by the agents mimics the behavior of the restabilization procedure studied in Section 5, and the results of this section therefore provide the benchmarks against which agents' incentives in the decentralized game can be assessed. Section 7 studies the equilibria of the decentralized game, and shows that while workers will generally have incentives not to reveal their true preferences, there are robust equilibrium outcomes of the game which are stable with respect to the true preferences. The final section concludes.

This paper can almost—but not quite—be read as two complementary but separate papers. Sections 2, 4, and 5 study the structure of the set of stable matchings, which in the model we consider equals the core of the game, and the related set of quasi-stable matchings. (Section 3 provides technical properties of the Deferred Acceptance Algorithm which is the engine that drives the process we explore.) Sections 6 and 7 study the strategic and equilibrium behavior in decentralized matching games which begin with quasi-stable matchings. Not too many years ago, the sections studying stable matchings and the sections studying strategic behavior would have fallen squarely into the traditions labeled "cooperative" and "noncooperative" game theory, respectively, with the implicit understanding that the cooperative theory was relevant to games in which binding agreements could be made, and the noncooperative theory applied to games in which all agreements had to be self enforcing. One thing this paper illustrates is why that distinction is not as useful as it was once thought to be—the two kinds of theory do not address different kinds of games here, but answer different questions about the same game. More generally, one clear lesson from all the empirical work on matching institutions is that market institutions whose strategic equilibria do not implement stable outcomes fail (and are abandoned) much more often than those which do. Thus the two kinds of analysis complement each other in much more than a technical sense.

2. STABLE MATCHINGS IN THE MARRIAGE MODEL

Formally, a *marriage market* is a triplet (F, W, P) where F and W are disjoint finite sets of *firms* and *workers*, respectively, and P is a function that maps each $f \in F$ into a strict preference relation P_f over the set $\{f\} \cup W$ and each $w \in W$ into a strict preference relation P_w over the set $\{w\} \cup F$. We refer to $V \equiv F \cup W$ as the set of *agents*. We write $v' >_v v''$ when v' is preferred to v'' under the preference relation P_v and in this case we say that v *prefers* v' to v''. The relations $<_v$, $\geqslant_v$, $\leqslant_v$ are derived in the standard way. Given a subset $U \subseteq V$ we write $\min_v U$ and $\max_v U$ to denote the v-least-preferred mate and v-most-preferred mate, respectively, in U. We sometimes add superscripts and write, for example, $v' >_v^P v''$ rather than $v' >_v v''$ to emphasize dependence on particular preferences.

A *matching* is a subset of $F \times W$ (i.e., a subset of pairs) such that any agent v appears in at most one of the pairs. A matching μ is identified with a one-to-one correspondence $\mu: V \to V$ where $\mu(f) = w$ and $\mu(w) = f$ if $(f, w) \in \mu$, and $\mu(v) = v$ if no pair in μ contains v. Given a matching μ, we call $\mu(v)$ the *outcome for* v *under* μ. If $\mu(v) \neq v$ we say that v is *matched under* μ, and call $\mu(v)$ the *mate of* v *under* μ. We say that v is *unmatched under* μ if $\mu(v) = v$. Given two matchings μ and μ' and a set of agents $V' \subseteq V$ we denote $\mu \geqslant_{V'} \mu'$ if for each $v \in V'$, $\mu(v) \geqslant_v \mu'(v)$, and we denote $\mu >_{V'} \mu'$ if $\mu \geqslant_{V'} \mu'$ and $\mu(v) \neq \mu'(v)$ for some $v \in V'$.[10]

For $v \in V$ we define the *acceptable set* of v *under* P to be the set of agents v would rather be matched with than remain unmatched, i.e., $A_v(P) \equiv \{v' \in V: v' >_v v\}$. Of course, if $v \in F$ then $A_v(P) \subseteq W$ and if $v \in W$ then $A_v(P) \subseteq F$. We say v is *acceptable to* v' if $v \in A_{v'}(P)$. A pair $(f, w) \in F \times W$ is called *acceptable* if f and w are acceptable to each other. The set of all acceptable pairs under P is denoted by $A(F, W, P)$, or, briefly, $A(P)$. Generally we represent the preferences of the agents by lists for each agent v, with the members of $A_v(P)$ listed in decreasing order of P_v. A matching is called *acceptable* if it is a subset of $A(P)$.

A *blocking pair* for a matching μ is a pair $(f, w) \in F \times W$ such that $w >_f \mu(f)$ and $f >_w \mu(w)$. Of course, if μ is acceptable then each blocking pair for μ is in $A(P)$. A matching is called *stable* if it is acceptable and has no blocking pairs. The idea is that if μ is a matching which admits a blocking pair (f, w), then it is unstable, since f and w would prefer to be matched to each other, and the rules allow them to arrange this.

We denote the set of all stable matchings by $S(F, W, P)$, or briefly $S(P)$. We say that v is *achievable for* v' if $\mu(v') = v$ for some $\mu \in S(P)$, i.e., if v and v' can be matched at a stable matching. Gale and Shapley [12] proved that a stable matching must exist, and there is now considerable empirical evidence that it is extremely difficult to enforce market outcomes which are *not* stable [29, 30]. In the model we consider here, the set of stable matchings equals the core of the game.

When we study the properties of firm-quasi-stable matchings, it will be useful to recall the following properties of stable matchings. For any two

[10] It is standard in the matching literature to define the matching to be the one-to-one correspondence μ, rather than the set of matched pairs. For a fixed set of agents, these two definitions are equivalent. Our definition will facilitate the discussion of matchings when the set of agents is variable.

matchings μ and μ' we define $\mu \vee_F \mu'$ and $\mu \wedge_F \mu'$ to be the two correspondences that map every $f \in F$ to its more preferred outcome and less preferred outcome, respectively, of $\mu(f)$ and $\mu'(f)$. We define $\mu \vee_W \mu'$ and $\mu \wedge_W \mu'$ to be the two correspondences that map every $w \in W$ to her more preferred outcome and less preferred outcome, respectively, of $\mu(w)$ and $\mu'(w)$. For an arbitrary pair of matchings those four correspondences might fail to be matchings. But for a pair of stable matchings, Knuth [20] credits Conway with the following theorem [32, pp. 36–39].

THEOREM 2.1. *Let $\mu, \mu' \in S(P)$. Then each of the four correspondences $\mu \vee_F \mu'$, $\mu \wedge_F \mu'$, $\mu \vee_W \mu'$ and $\mu \wedge_W \mu'$ defines a matching and each of these matchings is stable. Moreover, $\mu \vee_F \mu' = \mu \wedge_W \mu'$ and $\mu \vee_W \mu' = \mu \wedge_F \mu'$.*

Theorem 2.1 asserts that the set of stable matchings forms a lattice under the partial order $\leqslant_F$ with lattice operators $\vee_F$ and $\wedge_F$ being defined coordinate-wise by firms' preferences. Further, this lattice is the dual of the lattice defined on $S(P)$ with the partial order $\leqslant_W$ and lattice operators $\vee_W$ and $\wedge_W$. It implies the existence of a stable matching which is optimal for all firms and the existence of a stable matching which is optimal for all workers. These two stable matchings are called *firm-optimal* and *worker-optimal*, respectively, and denoted by $\mu_F(P)$ and $\mu_W(P)$, respectively. Moreover, each of those two matchings is the worst for all agents of the opposite set. Thus we can make welfare comparisons among the stable matchings.

However the following theorem says that these welfare comparisons only concern matched agents, because any agent who is unmatched at some stable matching (i.e., any firm whose position remains vacant, or any worker who remains unemployed) is unmatched at *every* stable matching.[11]

THEOREM 2.2. *There exists a subset $V_1 \subseteq V$ such that for each $v \in V$ and $\mu \in S(P)$, $\mu(v) \neq v$ if and only if $v \in V_1$.*

3. DEFERRED ACCEPTANCE WITH ARBITRARY INPUT

In this section we study executions of the Deferred Acceptance (DA) Algorithm with arbitrary input matchings. We show that each such execution is finite and that all executions with the same input yield a common output matching; we further characterize the output in terms of the input matching and the players' preferences.

Given an unstable matching μ' with a blocking pair (f, w) we say that the matching μ *is obtained from μ' by satisfying the blocking pair (f, w)* if

[11] This result was demonstrated by Mcvitie and Wilson [21] for the case in which all agents are mutually acceptable; see Roth and Sotomayor [32, p. 42] for the present model.

f and w are matched to each other in μ, their mates (if any) in μ' are unmatched in μ, and the outcomes for all other agents remain unchanged, i.e.,

$$\mu(v) = \begin{cases} w & \text{if} \quad v = f \\ f & \text{if} \quad v = w \\ v & \text{if} \quad v = \mu'(f) \quad \text{and} \quad v \neq f \\ v & \text{if} \quad v = \mu'(w) \quad \text{and} \quad v \neq w \\ \mu'(v) & \text{otherwise.} \end{cases}$$

Of course, if μ' is a matching, μ defined above is a matching as well. Satisfying blocking pairs seems a natural procedure for "correcting instability," but as noted earlier, if blocking pairs are chosen arbitrarily the process may cycle and thus fail to terminate at a stable matching.

Given a matching μ', a pair (f, w) is a *firm-maximal blocking pair for* μ' if (f, w) is a blocking pair for μ' and w is the f-most-preferred worker among those with whom f forms a blocking pair for μ'. Of course, if a firm is a part of a blocking pair, it is also part of a firm-maximal blocking pair. Thus, it is sensible that when adopting a scheme for "correcting instability″ by satisfying blocking pairs, attention be restricted to satisfying only firm-maximal blocking pairs, but Abeledo and Rothblum [3] demonstrated through an example that starting with an acceptable matching and iteratively satisfying even firm-maximal blocking pairs does not necessarily produce a converging procedure. The DA Algorithm which is described next iteratively satisfies firm-maximal blocking pairs, but it does so only when the firm's position is vacant.[12]

THE DEFERRED ACCEPTANCE ALGORITHM

Input

A matching μ'.

Initialization

0. a. For all $f \in F$: $A_f^0 := A_f(P) \setminus \{\mu'(f)\}$.
   b. $\mu^0 := \mu'$, $i := 1$.

Main Iteration

1. If there is no $f \in F$ such that $\mu^{i-1}(f) = f$ and $A_f^{i-1} \neq \emptyset$, stop with output μ^{i-1}.
2. Let f be a firm such that $\mu^{i-1}(f) = f$ and $A_f^{i-1} \neq \emptyset$, and set $w \equiv \max_f A_f^{i-1}$.

[12] The algorithm we describe is not symmetric in the treatment of firms and workers, and by exchanging the roles of firms and workers we get an alternative algorithm to which we refer as the "worker version" of the DA Algorithm.

3. a. If $\mu^{i-1}(w) >_w f$, then $\mu^i := \mu^{i-1}$.
   b. Else,
      if $\mu^{i-1}(w) = w$, then $\mu^i := \mu^{i-1} \cup \{(f, w)\}$.
      Else, $\mu^i := \mu^{i-1} \cup \{(f, w)\} \backslash \{(\mu^{i-1}(w), w)\}$.
4. $A_f^i := A_f^{i-1} \backslash \{w\}$ and for all $f' \neq f$, $A_{f'}^i := A_{f'}^{i-1}$.
5. $i := i+1$, Go to 1.

An execution of the DA Algorithm starts with an arbitrary matching, selects a firm (randomly or otherwise) whose position is vacant, and lets it approach its most preferred workers (in order of preference) checking whether they form a blocking pair. If they do, this will be a firm-maximal blocking pair, and when this blocking pair is satisfied a new matching is formed. This process is then iterated until there is no firm with a vacant position that is part of a blocking pair.

Gale and Shapley [12] introduced the deferred acceptance algorithm with empty input.[13] They showed that the DA Algorithm converges to a stable matching when the input is the empty matching. The next theorem extends the convergence of the DA algorithm when applied to arbitrary initial matchings. It is established by showing that executions of the DA Algorithm with given input μ' can be viewed as executions of the algorithm with empty input when the given preferences P are replaced by preferences $P^{\mu'}$ derived from P in the following way:[14]

1. For each μ'-matched firm f, $\mu'(f)$ (whether acceptable or not to f) is moved to the top position of f's preference list, and

2. for a μ'-matched worker w, w is moved (up or down) to just below $\mu'(w)$ (so that $\mu'(w)$ becomes her least preferred acceptable mate).

3. The preferences of μ'-unmatched agents remain unchanged.

Of course, for the empty matching $\varnothing$ we have $P^{\varnothing} = P$. A matching μ is called $P^{\mu'}$-*acceptable* or $P^{\mu'}$-*stable* if it is, respectively, acceptable or stable with respect to the preferences $P^{\mu'}$. The set of all $P^{\mu'}$-stable matchings will be denoted $S(P^{\mu'})$.

THEOREM 3.1. *Let μ' be a matching. Then every execution of the DA Algorithm with input μ' terminates after a finite number of iterations with a*

[13] In fact, the above formulation corresponds to the McVitie–Wilson [21] version of the deferred acceptance algorithm where at each step at most one pair is satisfied. A "block-pivot" version that corresponds to the original deferred acceptance algorithm of Gale and Shapley [12] may also be used. In this version some firm-maximal blocking pairs are satisfied simultaneously at each step.

[14] We are indebted to an anonymous referee for suggesting the use of the modified preferences to shorten our earlier proof of the forthcoming Theorem 3.1.

common output which is the $P^{\mu'}$-worker-worst and $P^{\mu'}$-firm-optimal outcome in $S(P^{\mu'})$.

Proof. We already observed that when $\mu' = \varnothing$ and the preferences are arbitrary, the DA Algorithm reduces to the McVitie–Wilson [21] variant of the Gale–Shapley Algorithm and the conclusions of the theorem are well known [32, Theorem 2.8, pp. 27–28]. We next show that the iterations of any one execution of the DA Algorithm with input μ' and preferences P coincide with the steps of an execution of the algorithm with empty initial matching and preferences $P^{\mu'}$ after $|\mu'|$ "preliminary steps" in which the μ'-matched firms make offers (consecutively) to their μ'-mates, respectively, and (as these firms are $P^{\mu'}$-acceptable to their μ'-mates) their offers are accepted in turn.

When the $|\mu'|$ "preliminary steps" of the DA Algorithm with empty input and preferences $P^{\mu'}$ are completed, we have $A_f^{|\mu'|} = A_f(P^{\mu'}) \backslash \{\mu'(f)\} = A_f(P) \backslash \{\mu'(f)\}$ for all $f \in F$ which are the values of $A_f^0(P)$ at initiation of the DA Algorithm with input μ' and preferences P. Also, the preferences $P_f^{\mu'}$ and P_f coincide over $A_f(P^{\mu'}) \backslash \{\mu'(f)\}$ for each, $f \in F$, and the preferences $P_w^{\mu'}$ and P_w coincide over $A_w(P^{\mu'}) \backslash \{w\}$ for each $w \in W$. Thus, after the $|\mu'|$ preliminary steps, an execution of the DA Algorithm with empty input and preferences $P^{\mu'}$ can proceed exactly by the main iterations of the specified execution of the DA Algorithm with input μ' and preferences P. The conclusions of the theorem then follow from the known conclusions of the theorem with empty input when the preferences are $P^{\mu'}$. ∎

Theorem 3.1 assures that, if the straightforward process of offers and acceptances and rejections modeled by the DA Algorithm captures the market dynamics, then the random order in which offers are made does not add any extra randomness to the final outcomes of the vacancy chains which result. (We will return to this question in Section 7, when we consider strategic play.) This conclusion allows us to define an operator DA on the set of matchings that maps a matching μ' to the common output of the executions of the DA Algorithm with input μ', namely, to the worker-worst matching in $S(P^{\mu'})$. When needed we write DA^P rather than DA to emphasize dependence of the output matching on particular preferences.

Theorem 3.1 specializes to the classic result of Gale and Shapley [12] asserting that starting from the empty matching, deferred acceptance yields $\mu_F(P)$, namely the firm-optimal stable matching. (This result was, in fact, used in the proof of Theorem 3.1, but a direct proof of Theorem 3.1 is available.)

Corollary 3.2. $\mathrm{DA}(\varnothing) = \mu_F(P)$

Proof. The conclusion is immediate from Theorem 3.1 and the fact that $P^{\varnothing} = P$. ∎

Theorem 3.1 expresses the outcome of the DA Algorithm with given input μ' in terms of the preferences $P^{\mu'}$ over the set of $P^{\mu'}$-stable matchings. We next develop representations in terms of the original preferences P. For that purpose we need some further definitions that relate to, but do not coincide with, $P^{\mu'}$-acceptability and $P^{\mu'}$-stability.

Let μ' be an arbitrary matching. A matching μ is μ'-*acceptable* if $\mu \geqslant_W \mu'$ and if for every $f \in F$ with $\mu(f) <_f f$ we have that $\mu(f) = \mu'(f)$. A matching μ is μ'-*stable* if it is μ'-acceptable and for every blocking pair (f, w) for μ, $\mu(f) = \mu'(f) \in W$. Denote by $S(\mu', P)$ the set of μ'-stable matchings. We observe that for $\mu' = \varnothing$, μ'-acceptability and μ'-stability coincide, respectively, with (regular) acceptability and stability. In general μ'-stability does not imply acceptability or stability. For example, any matching μ' is μ'-acceptable, further, if all firms are matched, it is μ'-stable; but, obviously, such a matching μ' need not be stable nor acceptable.

The following lemma summarizes the potential instabilities in a μ'-stable matching. In particular, it shows that any instability in a μ'-stable matching μ is also present in μ'.

Lemma 3.3. *Let μ' be a matching and let $\mu \in S(\mu', P)$. Also, let $f \in F$ and $w \in W$. Then*:

1. *If $\mu(f) <_f f$, then $\mu'(f) <_f f$.*
2. *If $\mu(w) <_w w$ then $\mu'(w) <_w w$.*
3. *If (f, w) is a blocking pair for μ then (f, w) is a blocking pair for μ'.*

Proof. By the μ'-stability of μ, if $\mu(f) <_f f$ then $\mu(f) = \mu'(f) <_f f$; also, if $\mu(w) <_w w$ then $\mu'(w) \leqslant_w \mu(w) <_w w$. Next, assume that (f, w) is a blocking pair for μ, that is, $w >_f \mu(f)$ and $f >_w \mu(w)$. Then by the μ'-stability of μ, $\mu(f) = \mu'(f)$. Hence, $w >_f \mu(f) = \mu'(f)$. Further, by the μ'-acceptability of μ, $\mu(w) \geqslant_w \mu'(w)$; hence, $f >_w \mu(w) \geqslant_w \mu'(w)$ and we have that (f, w) is a blocking pair for μ'. ∎

In order to characterize the output of the DA Algorithm in terms of the original preferences we need the following two lemmas. The first relates $P^{\mu'}$-stability and μ'-stability and the second records properties of the output of the DA Algorithm in terms of the original preferences.

Lemma 3.4. *Let μ' be a matching and let μ be a $P^{\mu'}$-stable matching that satisfies $\mu \geqslant_W \mu'$. Then μ is μ'-stable.*

Proof. Since μ is $P^{\mu'}$-acceptable, if $\mu(f) <_f f$ for $f \in F$, then $P_f^{\mu'}$ is obtained from P_f by "moving' $\mu(f)$, that is, $\mu(f) = \mu'(f) \in W$. So μ is μ'-acceptable. To see that μ is μ'-stable assume that (f, w) is a blocking pair of μ, that is, $w >_f \mu(f)$ and $f >_w \mu(w)$. Since μ is $P^{\mu'}$-stable, (at least) one of the two above inequalities does not hold under $P^{\mu'}$; so, either $P_f^{\mu'}$ is obtained from P_f by moving $\mu(f)$ or $w = \mu'(f)$, or $P_w^{\mu'}$ is obtained from P_w by moving $\mu(w) = w$ ($f \in F$ is not moved in the construction of $P_w^{\mu'}$). If $P_f^{\mu'}$ is obtained from P_f by moving $\mu(f)$, then $\mu(f) = \mu'(f) \in W$. If $P_f^{\mu'}$ is obtained from P_f by moving w then $w = \mu'(f)$ and so $\mu'(w) = f >_w \mu(w)$, contradicting $\mu \geqslant_W \mu'$. Finally, if $P_w^{\mu'}$ is obtained from P_w by moving $\mu(w)$ then we must have $\mu(w) = w \leqslant_w \mu'(w)$ contradicting the assumption $\mu \geqslant_W \mu'$. So, $\mu(f) = \mu'(f) \in W$, completing the proof that μ is μ'-stable. ∎

LEMMA 3.5. *Let μ' be a matching. Then*:

1. $\mathrm{DA}(\mu') \geqslant_W \mu'$.
2. $\mathrm{DA}(\mu')$ *is μ'-stable.*
3. *if $\bar{\mu}$ is a μ'-stable matching satisfying $\bar{\mu} \geqslant_W \mu'$, then $\bar{\mu} \geqslant_W DA(\mu')$.*

Proof. (1) Let $\mu^0 = \mu', \mu^1, ..., \mu^k$ be the sequence of distinct matchings generated by any one execution of the DA Algorithm with input μ', and let f^i and w^i be the firm and worker selected in step 2 of the main iteration in which μ^i is generated. Then $\mu^i(w^i) = f^i >_{w^i} \mu^{i-1}(w^i)$ and $\mu^i(w) = \mu^{i-1}(w)$ for each $w \in W \backslash \{w^i\}$; hence, $\mu^i \geqslant_W \mu^{i-1}$. By iterating this inequality we conclude that $\mu' = \mu^0 \leqslant_W \mu^1 \leqslant_W ... \leqslant_W \mu^k = \mathrm{DA}(\mu')$.

(2) By Theorem 3.1, $\mathrm{DA}(\mu')$ is $P^{\mu'}$-stable, and by part (1), $\mathrm{DA}(\mu') \geqslant_W \mu'$. Hence, by Lemma 3.4, $\mathrm{DA}(\mu')$ is μ'-stable.

(3) Suppose $\bar{\mu}$ is a μ'-stable matching satisfying $\bar{\mu} \geqslant_W \mu'$. As in the proof of (1) assume that $\mu^0 = \mu', \mu^1, ..., \mu^k$ are generated during an execution of the DA Algorithm with input μ', and we will show, by induction, that $\bar{\mu} \geqslant_W \mu^i$ for each $i = 0, 1, ..., k$. For $i = 0$, the conclusion follows from the assumption $\bar{\mu} \geqslant_W \mu'$. Assume that for some $i - 1 \in \{0, 1, ..., k-1\}$, $\bar{\mu} \geqslant_W \mu^{i-1}$ and consider μ^i. Let (f^i, w^i) be the firm-maximal blocking pair for μ^{i-1} which is satisfied when μ^i is created. As $\bar{\mu}(w') \geqslant_{w'} \mu^{i-1}(w') = \mu^i(w')$ for $w' \in W \backslash \{w^i\}$, it suffices to show that $\bar{\mu}(w^i) \geqslant_{w^i} \mu^i(w^i) = f^i$. We next assume that $\bar{\mu}(w^i) <_{w^i} f^i$ and we will establish a contradiction.

As $\bar{\mu}$ is μ'-stable and $\bar{\mu} \geqslant_W \mu'$ we have that either $\bar{\mu}(f^i) = \mu'(f^i) \in W$ or (f^i, w^i) is not a blocking pair for $\bar{\mu}$. Now, if $w' \equiv \bar{\mu}(f^i) = \mu'(f^i) \in W$, we get from $\mu^{i-1} \geqslant_W \mu'$ (established in the proof of (1)) and $\bar{\mu} \geqslant_W \mu^{i-1}$ (our inductive assumption) that $f^i = \bar{\mu}(w') \geqslant_{w'} \mu^{i-1}(w') \geqslant_{w'} \mu'(w') = f^i$. Thus $\mu^{i-1}(w') = f^i$, in contradiction to the fact that $\mu^{i-1}(f^i) = f^i$. We conclude that (f^i, w^i) is not a blocking pair for $\bar{\mu}$. In view of the assumption

$\bar{\mu}(w^i) <_{w^i} f^i$ it follows that $\bar{\mu}(f^i) >_{f^i} w^i$. As (f^i, w^i) is a blocking pair for μ^{i-1} we also have that $f^i = \mu^{i-1}(f^i) <_{f^i} w^i <_{f^i} \bar{\mu}(f^i)$. Consequently, $w' \equiv \bar{\mu}(f^i) \in W$, and further, as (f^i, w^i) is a firm-maximal blocking pair for μ^{i-1} we also conclude that (f^i, w') is not a blocking pair for μ^{i-1}. Thus $\mu^{i-1}(w') >_{w'} f^i = \bar{\mu}(w')$, contradicting the assumption that $\mu^{i-1} \leqslant_W \bar{\mu}$. ▮

THEOREM 3.6. *Let μ' be a matching. Then $DA(\mu')$ is the worker-worst matching in $S(\mu', P)$.*

Proof. The conclusion is immediate from the three parts of Lemma 3.5. ▮

Theorem 3.6 does not assert that the set $S(\mu', P)$ is a lattice, nor that, in general, the common output of executions of the DA Algorithm is firm-optimal in $S(\mu', P)$. We will see in the next section (Theorem 4.3) that stronger results can be obtained when μ' is a firm quasi-stable matching.

The (common) output of executions of the DA Algorithm with arbitrary input μ' need not be stable. For example, for every matching μ' under which all firms are matched $\text{DA}(\mu') = \mu'$, but such matchings need not be stable. However, the following lemma shows that any firm involved in instability of the output matching is never "active" in any execution of the DA Algorithm with input μ'.

LEMMA 3.7. *Let μ' be a matching, let $\mu \equiv DA(\mu')$ and let $f \in F$ where either $\mu(f) <_f f$ or (f, w) is a blocking pair for μ for some $w \in W$. Then f maintains the same mate in W throughout all steps of all executions of the DA Algorithm.*

Proof. Let $w' \equiv \mu(f)$. Now, if $\mu(f) <_f f$, the μ'-acceptability of μ assures that $\mu'(f) = \mu(f) = w' \in W$. Also, if f belongs to some blocking pair for μ, the μ'-stability of μ assures that $w' = \mu(f) = \mu'(f) \in W$. In either case, $f = \mu(w') = \mu'(w')$ and using the fact that a worker can only improve her outcome along executions of the DA Algorithm (see the proof of part (1) of Lemma 3.5) we conclude that f is matched to $w' \in W$ throughout all steps of all executions of the DA Algorithm with input μ'. ▮

4. FIRM-QUASI-STABLE MATCHINGS

In this section we examine a class of matchings we call firm-quasi-stable and show that they arise when stable matchings are disrupted by the creation of new positions and/or by the retirement of existing workers. We then show that the output of the DA Algorithm on inputs that are in this class are always stable and we provide several characterizations of such outputs.

A matching μ is called *firm-quasi-stable* if it is acceptable and has no blocking pair that contains a matched firm, i.e., if (f, w) is a blocking pair, then the firm f must be unmatched. Let $Q(P)$ be the set of firm-quasi-stable matchings. Of course, the empty matching is firm-quasi-stable and the set of stable matchings $S(P)$ is a subset of $Q(P)$. Firm-quasi-stability was first introduced by Sotomayor [38] as part of a new existence proof.[15] We next consider how the stable matchings of a marriage market are related to the firm-quasi-stable matchings of another marriage market obtained from the first by the retirement of some workers and/or the creation of some new (and vacant) positions.

Let (F, W, P) and (F', W', P') be arbitrary marriage markets. We say the market (F, W, P) is *consistent with* (F', W', P') if the natural restrictions of P and P' to the set $(F \cap F') \cup (W \cap W')$ coincide, i.e., if for $F^* \equiv F \cap F'$ and $W^* \equiv W \cap W'$ the following conditions hold:

1. $A(F, W, P) \cap (F^* \times W^*) = A(F', W', P') \cap (F^* \times W^*)$,
2. For each $f \in F^*$ and $w, w' \in W^*$, $w >_f^P w'$ if and only if $w >_f^{P'} w'$,
3. For each $w \in W^*$ and $f, f' \in F^*$, $f >_w^P f'$ if and only if $f >_w^{P'} f'$,

where we remind the reader that $A(F, W, P)$ is the set of acceptable pairs for the market (F, W, P).

We say the market (F', W', P') *leads to* (F, W, P), written $(F', W', P') \to (F, W, P)$, if $F' \subseteq F$, $W' \supseteq W$ and (F', W', P') is consistent with (F, W, P). To interpret this definition note that if $(F', W', P') \to (F, W, P)$ we may view (F, W, P) as obtained from (F', W', P') by the creation of new positions and retirement of existing workers. In particular, the set of new firms is $F \backslash F'$ and the set of retired workers is $W' \backslash W$. The next theorem shows that a matching is firm-quasi-stable if and only if it can arise from the disruption of a stable matching in this way.

THEOREM 4.1. *Let μ' be a matching for (F, W, P). Then $\mu' \in Q(F, W, P)$ if and only if there exist a marriage market (F', W', P') and a matching μ'' for (F', W', P') such that $(F', W', P') \to (F, W, P)$, $\mu'' \in S(F', W', P')$ and $\mu' = \mu'' \cap (F \times W)$.*

Proof. Assume $(F', W', P') \to (F, W, P)$, $\mu'' \in S(F', W', P')$ and $\mu' = \mu'' \cap (F \times W)$. Since $\mu'' \subseteq A(F', W', P')$ and since $A(F', W', P') \cap (F \times W) \subseteq A(F, W, P)$, then $\mu' = \mu'' \cap (F \times W) \subseteq A(F, W, P)$ which means that μ' is an acceptable matching under (F, W, P). Next, if $f \in F' \subseteq F$ and

[15] In a similar way we may define a worker-quasi-stable matching to be an acceptable matching in which each blocking pair contains an unmatched worker and a quasi-stable matching to be a matching in which each blocking pair contains at least one unmatched agent. The concept of worker-quasi-stability is symmetric to firm-quasi-stability and each result in this paper has a dual one, corresponding to worker-quasi-stability.

$\mu'(f) = \mu''(f)$, then by the stability of μ'' under (F', W', P') it follows that no pair consisting of f and a worker $w \in W \subseteq W'$ is a blocking pair for μ' (here we use the fact that for each $w \in W$, $\mu'(w) = \mu''(w)$; hence the pair (f, w) which is not a blocking pair for μ'' is not a blocking pair for μ' as well). Further, if either $f \in F \backslash F'$ or $f \in F'$ but $\mu'(f) \neq \mu''(f)$ then, $\mu'(f) = f$. We proved that no pair that contains a matched firm blocks μ' under (F, W, P); hence, $\mu' \in Q(F, W, P)$.

Assume now that $\mu' \in Q(P)$. Let $F' \equiv \{f \in F: \mu'(f) \neq f\}$, let $W' \equiv W$ and let P' be the natural restriction of P to $F' \cup W'$. Then $(F', W', P') \to (F, W, P)$. Further, since only unmatched firms and pairs which include them were excluded, μ' is acceptable under (F', W', P'). As $A(F', W', P') \subseteq A(F, W, P)$ and P' is the restriction of P to $F' \cup W'$, it also follows that only pairs that block μ' under (F, W, P) might block it under (F', W', P'). But each blocking pair for μ' under (F, W, P) contains an unmatched firm and all unmatched firms are not in F'. Hence there is no blocking pair for μ' under (F', W', P') and in particular, $\mu' \in S(F', W', P')$. Now, $\mu'' \equiv \mu'$ satisfies the requirements of the theorem. ▮

Note that the market (F', W', P') defined in the "only if" part of the proof is not unique. In fact, we have to exclude from F only firms that belong to some blocking pairs for the given matching. We can also include in W' any set of additional workers and in this case, P' and μ'' should be extended appropriately. The selection in the proof is thus sufficient but may be relaxed.

Theorem 4.1 proves that the firm-quasi-stable matchings of a given market are restrictions of the stable matchings of consistent markets with fewer positions and more workers. In particular, it shows that stable matchings become firm-quasi-stable after the creation of new positions and/or the retirement of workers.[16]

By Theorem 4.1, results about firm-quasi-stable matchings can be formulated as results about stable matchings of markets which lead to (F, W, P). Lemma A.2 and Theorem A.3 in the Appendix are examples of such formulations. These two results appear in [13] in terms of the creation of new positions.[17]

[16] Dual arguments show that stable matchings become worker-quasi-stable after the removal of existing positions and/or the entrance of new workers to the market. Further, stable matchings become quasi-stable when all these phenomena occur, that is, creation of new positions, deletion of existing positions, entrance of new workers, and retirement of existing workers.

[17] These results are stated in Gale and Sotomayor [13] for the case of "extending the preferences" of the firms which is a concept that generalizes the creation of new positions. In fact, firm-quasi-stability captures "extending the preferences" as well and hence such extensions can be regarded as another source of firm-quasi-stability.

We next consider the execution of the DA Algorithm on firm-quasi-stable matchings. We start by characterizing stability with respect to such matchings.

LEMMA 4.2. *Let* $\mu' \in Q(P)$. *Then* $S(\mu', P) = \{\mu \in S(P)\colon \mu \geqslant_W \mu'\}$.

Proof. The inclusion $\{\mu \in S(P)\colon \mu \geqslant_W \mu'\} \subseteq S(\mu', P)$ is trivial (and holds even if $\mu' \notin Q(P)$). To see the reverse inclusion assume that μ is μ'-stable. The acceptability of μ' and parts (1) and (2) of Lemma 3.3 assure that μ is acceptable, and the μ'-acceptability of μ assures that $\mu \geqslant_W \mu'$. Finally, if (f, w) is a blocking pair of μ, the μ'-stability of μ assures that $\mu(f) = \mu'(f) \in W$ and part (3) of Lemma 3.3 assures that (f, w) is a blocking pair of μ', in contradiction to the quasi-stability of μ'. So μ has no blocking pairs and the proof that $\mu \in S(P)$ and $\mu \geqslant_W \mu'$ is complete. ∎

Given an acceptable matching μ', define

$$\underline{S}_W^{\mu'}(P) \equiv \{\mu \in S(P)\colon \mu \geqslant_W \mu'\},$$

that is, $\underline{S}_W^{\mu'}(P)$ is the set of stable matchings that the workers weakly prefer to μ'. Lemma 4.2 shows that, if μ' is firm-quasi-stable, $S(\mu', P) = \underline{S}_W^{\mu'}(P)$. Further, in the Appendix (Theorem A.6) we prove that when μ' is a firm-quasi-stable matching, $\underline{S}_W^{\mu'}(P)$ is a nonempty sub-lattice of $S(P)$, with lattice operators $\vee_W$ and $\wedge_W$. Thus, $\underline{S}_W^{\mu'}(P)$ contains the matching $\wedge_W \underline{S}_W^{\mu'}(P)$, which is the worker-worst stable matching in $\underline{S}_W^{\mu'}(P)$. Further, $\vee_F$ coincides with $\wedge_W$ on $S(P)$; so $\wedge_W \underline{S}_W^{\mu'}(P) = \vee_F \underline{S}_W^{\mu'}(P) \in S(P)$ and this matching is also the firm-optimal matching in $\underline{S}_W^{\mu'}(P)$. We next show that this stable matching is the output of every execution of the DA Algorithm with input μ'.

THEOREM 4.3. *Let* $\mu' \in Q(P)$. *Then* $DA(\mu') = \vee_F \underline{S}_W^{\mu'}(P) = \wedge_W \underline{S}_W^{\mu'}(P)$.

Proof. The result is immediate from Theorem 3.6 and Lemma 4.2 and the paragraph preceding this theorem. ∎

Theorem 4.3 implies that the DA operator maps firm-quasi-stable matchings into stable matchings. The next theorem gives a simple representation of the output of the application of this operator on firm-quasi-stable matchings in terms of the input matching and the firm-optimal stable matching $\mu_F(P)$.

THEOREM 4.4. *Let* $\mu' \in Q(P)$. *Then* $DA(\mu') = \mu' \vee_W \mu_F(P)$.

Proof. Let $\mu \equiv \mu' \vee_W \mu_F(P)$. Since $\mu \geqslant_W \mu'$ and since by Theorem A.3 $\mu \in S(P)$, we have that $\mu \in \underline{S}_W^{\mu'}(P)$. Next, for $\mu'' \in \underline{S}_W^{\mu'}(P)$ we have that $\mu'' \in S(P)$ and $\mu'' \geqslant_W \mu'$; as the former condition implies that $\mu'' \geqslant_W \mu_F(P)$, we conclude that $\mu'' \geqslant_W \mu' \vee_W \mu_F(P) = \mu$. So, μ is the worker-worst matching in $\underline{S}_W^{\mu'}(P)$; that is $\mu = \wedge_W \underline{S}_W^{\mu'}(P)$. By Theorem 4.3, $\text{DA}(\mu') = \wedge_W \underline{S}_W^{\mu'}(P) = \mu = \mu' \vee_W \mu_F(P)$. ∎

COROLLARY 4.5. *If μ'_1 and μ'_2 are firm-quasi-stable matchings with $\mu'_1 \geqslant_W \mu'_2$ then $DA(\mu'_1) \geqslant_W DA(\mu'_2)$.*

Corollary 4.5 shows that the DA operator is $\geqslant_W$-monotone on firm-quasi-stable matchings. Although the operator need not be $\geqslant_F$-monotone, in Section 5 we will establish a related result that applies to the firms.

Theorem 4.4 shows that if $\mu' \in Q(P)$, then for each $w \in W$, $[\text{DA}(\mu')](w) = \max_w\{\mu'(w), [\mu_F(P)](w)\}$. Thus we have a closed-form expression for the outcome of a worker under the output of the DA Algorithm as the maximum of her initial assignment and her worst achievable outcome. The next corollary provides another closed-form representation of the output outcome that applies to both the workers and the firms.

COROLLARY 4.6. *Let $\mu' \in Q(P)$ and $v \in F \cup W$. Then*

$$[DA(\mu')](v) = \begin{cases} \mu'(v) & \text{if } \mu'(v) \text{ is achievable for } v \\ [\mu_F(P)](v) & \text{if } \mu'(v) \text{ is not achievable for } v. \end{cases}$$

Proof. Let $\mu \equiv \text{DA}(\mu')$ and assume first $v = w \in W$. If $\mu'(w)$ is achievable for w, then, because $\mu_F(P)$ is the worker-worst stable matching, $\mu'(w) \geqslant_w [\mu_F(P)](w)$; hence $[\mu' \vee_W \mu_F(P)](w) = \mu'(w)$ and by Theorem 4.4, $\mu(w) = [\mu' \vee_W \mu_F(P)](w) = \mu'(w)$. If, alternatively, $\mu'(w)$ is not achievable for w, then by Corollary A.5, $\mu'(w) <_w [\mu_F(P)](w)$. Hence $[\mu' \vee_W \mu_F(P)](w) = [\mu_F(P)](w)$ and by Theorem 4.4, $\mu(w) = [\mu' \vee_W \mu_F(P)](w) = [\mu_F(P)](w)$.

Next, assume that $v = f \in F$. If $\mu(f) = [\text{DA}(\mu')](f) = f$ then by Theorem 2.2 and the stability of μ, f is unmatched under all stable matchings and thus the only achievable outcome for f is $\mu(f) = [\mu_F(P)](f) = f$ which coincides with the two asserted alternatives for $[\text{DA}(\mu')](v)$. So, assume that $\mu(f) \in W$. Now, if $\mu'(f)$ is achievable for f then Theorem 2.2 and the fact that $\mu(f) \in W$ imply that $w \equiv \mu'(f) \in W$. As w is achievable for f, $f = \mu'(w)$ is achievable for w; hence, by the first part of the proof, $\mu(w) = \mu'(w) = f$ and therefore $\mu(f) = w = \mu'(f)$. Finally, if $\mu'(f)$ is not achievable for f then $w \equiv \mu(f) \neq \mu'(f)$ (where $w \in W$). Then $f = \mu(w) \neq \mu'(w)$ and by the first part of the proof it follows that $f = \mu(w) = [\mu_F(P)](w)$. In particular, $\mu(f) = w = [\mu_F(P)](f)$. ∎

Theorem 4.4 and Corollary 4.6 show that when the input is a firm-quasi-stable matching, the outcome of an agent under the DA Algorithm is determined by its/her initial outcome and is independent of the initial outcome of the other agents. In particular, if the assignments of an agent under two firm-quasi-stable matchings coincide, then the assignments of that agent coincide under the corresponding outputs of the DA Algorithm. Further, as the output of the DA Algorithm under the empty matching is the firm-optimal stable matching, we conclude that a firm that is unmatched under a (firm-quasi-stable) input matching is assigned under the corresponding output matching to its optimal achievable outcome and a worker who is unmatched under the input matching ends up with her worst achievable outcome.

Abeledo and Rothblum [3] demonstrated that, with the empty matching as input, the DA Algorithm has a natural interpretation as an execution of the dual-simplex method for finding an extreme point of a corresponding polyhedron. Their observation can be extended to arbitrary firm-quasi-stable input matchings. The proof of this generalization follows from the arguments of Abeledo and Rothblum with minor modifications.

5. RESTABILIZATION AFTER RETIREMENTS AND NEW ENTRIES

In Section 4 we showed that firm-quasi-stable matchings arise from stable matchings following the creation of new jobs and/or the retirement of workers, e.g., in senior-level labor markets. As the DA Algorithm moves from firm-quasi-stable matchings to stable ones, its re-equilibration process connects the stable matching in the original (pre-job creation and retirement) market with the stable matching achieved in the new market. In the current section we use results of Section 4, about the relations of the firm-quasi-stable input matching for the DA Algorithm and the corresponding output, to explore the connection between the original stable matching and the one obtained after the re-equilibration in the new market. As the set of players is not invariant in the development described in the current section, we index acceptable sets and sets of stable matchings both by preferences and by the sets of firms and workers, e.g., $A(F, W, P)$ and $S(F, W, P)$.

Throughout this section we assume that (F', W', P') is a marriage market such that $(F', W', P') \rightarrow (F, W, P)$ where we remind the readers that the relation $\rightarrow$ is introduced in Section 4 with the interpretation that (F, W, P) is obtained from (F', W', P') through creation of new jobs and/or retirement of workers. Also, recall that if μ' is the stable matching for the original market (F', W', P'), then $\mu' \cap (F \times W)$ is a firm-quasi-stable matching in the new market (F, W, P). The DA Algorithm defines a *(re-)Equilibration Operator*, which transforms a stable matching μ' of the original market, which becomes unstable when new jobs are created and/or workers retire, into a stable matching in the new market

$$E^{F, W, P}_{F', W', P'} : S(F', W', P') \mapsto S(F, W, P)$$

where for each $\mu' \in S(F', W', P')$,

$$E^{F, W, P}_{F', W', P'}(\mu') \equiv \mathrm{DA}(\mu' \cap [F \times W]).$$

The first result of this section shows that this re-equilibration is independent of the details of timing. It does not affect the outcome if positions start to be filled before all retirements and new entries have occurred, or after. In particular, requilibration via intermediate markets coincides with the outcome of direct equilibration.

THEOREM 5.1. *Let (F'', W'', P'') be a marriage market such that $(F', W', P') \to (F'', W'', P'') \to (F, W, P)$ and let $\mu' \in S(F', W', P')$. Then*

$$E^{F, W, P}_{F'', W'', P''}[E^{F'', W'', P''}_{F', W', P'}(\mu')] = E^{F, W, P}_{F', W', P'}(\mu').$$

Proof. Assume $\mu^0 = \mu' \cap (F'' \times W''), ..., \mu^l$ is a sequence of (distinct) matchings generated by an execution of the $E^{F'', W'', P''}_{F', W', P'}$ Algorithm with input μ' and set $\mu'' \equiv \mu^l \in S(F'', W'', P'')$. Then for each $i = 0, ..., l-1$, μ^{i+1} is obtained from μ^i by satisfying a firm-maximal blocking pair in (F'', W'', P''). Now, for $i = 0, ..., l$ set $\tilde{\mu}^i \equiv \mu^i \cap (F \times W)$. Let $0 \leqslant i < l$ and assume μ^{i+1} is obtained from μ^i by satisfying the firm-maximal blocking pair (f^i, w^i) under (F'', W'', P''). If $w^i \in W'' \backslash W$ then $\tilde{\mu}^i = \tilde{\mu}^{i+1}$. Alternatively, $w \in W$. As $\tilde{\mu}^i(f^i) = \mu^i(f^i) = f^i$ and for each $w' \in W \subseteq W''$, $\tilde{\mu}^i(w') = \mu^i(w')$, the consistency of (F, W, P) and (F'', W'', P''), implies that the pair (f^i, w^i) which is a firm-maximal blocking pair for μ^i under (F'', W'', P'') is also a firm-maximal blocking pair for $\tilde{\mu}^i = \mu^i \cap (F \times W)$ under (F, W, P). It follows that by excluding duplicate matchings from the sequence $\tilde{\mu}^0, ..., \tilde{\mu}^l$ we get a (sub)sequence of matchings for (F, W, P) such that each is obtained from its predecessor by satisfying a firm-maximal blocking pair for that matching under (F, W, P).

We showed that, with possibly dropping intermediary matchings, $\tilde{\mu}^0 = \mu' \cap (F \times W), ..., \tilde{\mu}^l = \mu'' \cap (F \times W)$ is a sequence of matchings where each is obtained from its predecessor by satisfying a firm-maximal blocking pair under (F, W, P). By continuing this partial execution of the DA Algorithm from $\tilde{\mu}^l$ to completion (in (F, W, P)) we see that $\mathrm{DA}(\tilde{\mu}^0) = \mathrm{DA}(\tilde{\mu}^l)$. Hence,

$$\begin{aligned} E^{F, W, P}_{F', W', P'}(\mu') &= \mathrm{DA}(\tilde{\mu}^0) = \mathrm{DA}(\tilde{\mu}^l) = \mathrm{DA}\{[E^{F'', W'', P''}_{F', W', P'}(\mu')] \cap (F \times W)\} \\ &= E^{F, W, P}_{F'', W'', P''}[E^{F'', W'', P''}_{F', W', P'}(\mu')]. \quad \blacksquare \end{aligned}$$

Theorem 3.1 implies that when the equilibration mechanism of a market is the DA Algorithm, the output matching is insensitive to the order in which the firms seek to fill their vacant positions. In a similar spirit, Theorem 5.1 shows that “intermediate” stages have no effect on the final outcome of the market, that is, the output is also insensitive to whether new positions are filled immediately, or only after a number of new positions have accumulated (e.g., seasonally).

An implication of Theorem 5.1 is that global firm-optimality (and worker-pessimality) is preserved by the re-equilibration operator; thus, if a market reaches firm-optimality at any stage, it will maintain firm-optimality at all later stages.[18] Henceforth, let μ_F and $\mu'_{F'}$ be the firm-optimal stable matchings for the markets (F, W, P) and (F', W', P'), respectively.

THEOREM 5.2.

$$E^{F, W, P}_{F', W', P'}(\mu'_{F'}) = \mu_F.$$

Proof. Let (F^*, W', P^*) be the marriage market such that $F^* = \emptyset$ and P^* is the empty preference profile. Then $(F^*, W', P^*) \to (F', W', P') \to (F, W, P)$ and $\emptyset \in S(F^*, W', P^*)$. By two applications of Corollary 3.2 and Theorem 5.1 (where (F^*, W', P^*) and (F', W', P') play the roles of (F', W', P') and (F'', W'', P''), respectively), we have that

$$\mu_F = E^{F, W, P}_{F^*, W', P^*}(\emptyset) = E^{F, W, P}_{F', W', P'}[E^{F', W', P'}_{F^*, W', P^*}(\emptyset)] = E^{F, W, P}_{F', W', P'}(\mu'_{F'}). \quad \blacksquare$$

Throughout the remainder of this section we consider the re-equilibration process only from (F', W', P') to (F, W, P). Hence, we drop the superscripts and the subscripts from the equilibration operator and simply use the notation E rather than $E^{F, W, P}_{F', W', P'}$.

By Corollary 4.6 and the paragraph following it, Theorem 5.2 can be refined by considering individual agents, as follows.

COROLLARY 5.3. Let $\mu' \in S(F', W', P')$ and $v \in F' \cup W$. If $\mu'(v) = \mu'_{F'}(v)$, then, $[E(\mu')](v) = \mu_F(v)$.

Proof. Let $\tilde{\mu}' \equiv \mu' \cap (F \times W)$ and $\tilde{\mu}'_{F'} \equiv \mu'_{F'} \cap (F \times W)$. As $\tilde{\mu}'(v) = \tilde{\mu}'_{F'}(v)$, Theorem 5.2 and the paragraph following Corollary 4.6 show that

$$[E(\mu')](v) = [\mathrm{DA}(\tilde{\mu}')](v) = [\mathrm{DA}(\tilde{\mu}'_{F'})](v) = [E(\mu'_{F'})](v) = \mu_F(v). \quad \blacksquare$$

We saw in Lemma 3.5 that the output of the DA Algorithm with given input μ' is weakly preferred by all workers μ'. We next show that when the DA algorithm is used to restabilize the market, the entry of new firms and

[18] This result is implicit in several of the earlier papers which consider the effect of new entrants on the market–see, e.g., Kelso and Crawford [19].

the retirement of workers cannot be good for any of the original firms, and cannot be bad for any of the workers who have not retired.[19]

LEMMA 5.4. *Let* $\mu' \in S(F', W', P')$ *and* $\mu \equiv E(\mu')$. *Then* $\mu \geqslant^P_W \mu'$ *and* $\mu \leqslant^{P'}_{F'} \mu'$.

Proof. Let $\tilde{\mu}' \equiv \mu' \cap (F \times W)$. By Theorem 4.1, $\tilde{\mu}' \in Q(P)$ and by Lemma 3.5, $\mu = E(\mu') = \mathrm{DA}(\tilde{\mu}') \geqslant^P_W \tilde{\mu}'$. As $F' \subseteq F$, we have that $\mu'(w) = \tilde{\mu}'(w)$, for each $w \in W \subseteq W'$; hence $\mu \geqslant^P_W \mu'$ as claimed. To prove the second inequality, assume by way of contradiction that $\mu(f) >^{P'}_f \mu'(f) \geqslant^{P'}_f f$ for some $f \in F'$. Then $w \equiv \mu(f) \in W$ and by the first conclusion $f = \mu(w) >^P_w \mu'(w)$ (there is no equality since $w = \mu(f) \neq \mu'(f)$). Further, since (F, W, P) and (F', W', P') are consistent, the inequality $f >^P_w \mu'(w)$ implies that $f >^{P'}_w \mu'(w)$. Hence, the pair $(f, w) \in F' \times W'$ is a blocking pair for μ' under (F', W', P'), in contradiction to its asserted stability. ∎

We next characterize the output matching of the re-equilibration process in terms of the "original" stable matching of the market (F', W', P'). We first consider the output outcomes of the new firms and show that each such firm ends up with its optimal achievable outcome.

LEMMA 5.5. *Let* $\mu' \in S(F', W', P')$ *and* $f \in F \backslash F'$. *Then* $[E(\mu')](f) = \mu_F(f)$.

Proof. Let $\tilde{\mu}' \equiv \mu' \cap (F \times W)$. Then $\tilde{\mu}'(f) = f$ and by Theorem 4.1, $\tilde{\mu}' \in Q(F, W, P)$; hence by Corollary 3.2 and the paragraph following Corollary 4.6,

$$[E(\mu')](f) = [\mathrm{DA}(\tilde{\mu}')](f) = [\mathrm{DA}(\varnothing)](f) = \mu_F(f). \quad \blacksquare$$

The next theorem gives a closed-form representation of the output outcome for each of the "original" agents. It is parallel to the representation in Theorem 4.4.

THEOREM 5.6. *Let* $\mu' \in S(F', W', P')$ *and* $v \in F' \cup W$. *Then*

$$[E(\mu')](v) = \begin{cases} \max^P_w \{\mu'(w), \mu_F(w)\} & \text{if } \; v = w \in W \\ \min^{P'}_f \{\mu'(f), \mu_F(f)\} & \text{if } \; v = f \in F'. \end{cases}$$

Proof. Let $\mu \equiv E(\mu')$ and set $\tilde{\mu}' \equiv \mu' \cap (F \times W)$. Then $\mu = \mathrm{DA}(\tilde{\mu}')$. As $\tilde{\mu}'(w) = \mu'(w)$ for each $w \in W$, Theorem 4.4 implies that

$$\mu(w) = [\mathrm{DA}(\tilde{\mu}')](w) = \max^P_w \{\tilde{\mu}'(w), \mu_F(w)\} = \max^P_w \{\mu'(w), \mu_F(w)\}.$$

[19] For a related result in the assignment market, see Mo [22]; for related results for entry level marriage models, see Section 2.5 in Roth and Sotomayor [32].

Let $f \in F'$. If $\mu(f) = \mu'(f)$ then by the firm-optimality of μ_F and the achievability of $\mu'(f)$ for f, $\mu_F(f) \geqslant_f^{P'} \mu'(f)$ and as (F, W, P) and (F', W', P') are consistent, this inequality holds under P' as well. In particular, $\mu(f) = \mu'(f) = \min_f^{P'}\{\mu'(f), \mu_F(f)\}$. Assume now that $\mu(f) \neq \mu'(f)$. Then by Lemma 5.4, $\mu(f) <_f^{P'} \mu'(f)$. Consider two cases: if $\tilde{\mu}'(f) = \mu'(f)$, then $\mu(f) \neq \mu'(f) = \tilde{\mu}'(f)$; hence by the representation in Corollary 4.6, $\mu(f) = \mu_F(f)$. Thus, $\mu_F(f) = \mu(f) <_f^{P'} \mu'(f)$, implying that $\mu(f) = \min_f^{P'}\{\mu'(f), \mu_F(f)\}$. If alternatively $\tilde{\mu}'(f) \neq \mu'(f)$, then $\tilde{\mu}'(f) = f$ and by Corollaries 3.2, 4.6 and the paragraph following Corollary 4.6, $\mu(f) = [\mathrm{DA}(\tilde{\mu}')](f) = [\mathrm{DA}(\varnothing)](f) = \mu_F(f)$. Hence, $\mu_F(f) = \mu(f) <_f^{P'} \mu'(f)$ implying (as before) that $\mu(f) = \min_f^{P'}\{\mu'(f), \mu_F(f)\}$. ∎

As a corollary of the above separable representation we get the monotonicity of the Equilibration Algorithm with respect to the individual preferences of the agents and with respect to the partial orders $\geqslant_W$ and $\geqslant_F$.

COROLLARY 5.7. *Let $\mu_1', \mu_2' \in S(F', W', P')$. Then:*

1. *If $\mu_1'(v) \geqslant_v^{P'} \mu_2'(v)$ for $v \in F' \cup W$ then $[E(\mu_1')](v) \geqslant_v^{P} [E(\mu_2')](v)$.*
2. *If $\mu_1' \geqslant_{F'}^{P'} \mu_2'$ then $E(\mu_1') \geqslant_{F'}^{P} E(\mu_2')$.*
3. *If $\mu_1' \geqslant_W^{P'} \mu_2'$ then $E(\mu_1') \geqslant_W^{P} E(\mu_2')$.*

Proof. The conclusions of the corollary with respect to the workers are immediate from Theorem 5.6 and the fact that for each $w \in W$, $\mu_1'(w) \geqslant_w^{P'} \mu_2'(w)$ if and only if $\mu_1'(w) \geqslant_w^{P} \mu_2'(w)$ (as (F, W, P) and (F', W', P') are consistent and $\mu_1'(w), \mu_2'(w) \in F' \cup \{w\}$). Similarly, the conclusions with respect to the firms are immediate from Theorem 5.6 and the fact that for each $f \in F'$, $\mu_1(f) \geqslant_f^{P} \mu_2(f)$ if and only if $\mu_1(f) \geqslant_f^{P'} \mu_2(f)$ (as (F, W, P) and (F', W', P') are consistent and $\mu_1(f), \mu_2(f) \in W \cup \{f\}$).

6. THE DECENTRALIZED DEFERRED ACCEPTANCE GAME

In this section we consider a game, to be called the *Decentralized Deferred Acceptance* (*DDA*) *game*, in which firms and workers make offers according to the general rules of the deferred acceptance algorithm, but are free to issue offers and make acceptances and rejections as they please, i.e. not necessarily by straightforwardly acting on their true preferences.

The game is given by a market (F, W, P) in which we refer to P as the *true preferences*, and an initial matching μ' known to all the players. Although the game is defined for arbitrary initial matchings, we will generally take μ' to be firm-quasi-stable under the true preferences. That is, the game we consider begins at a firm-quasi-stable matching; equivalently, it begins with a stable matching whose stability is disrupted after new firms have entered the market, and old workers have retired.[20] At any moment in the game, one firm is randomly selected from among those with vacant positions and is allowed to make an offer to any worker who has not previously rejected an offer from that firm.[21] (The random selection of firms can be thought of as reflecting unmodeled institutional features, such as the internal governance structure of firms, which causes some firms to act more quickly than others, the speed of the mail, which causes some offers to arrive earlier than others, or even the telephone switching system, which causes one firm to get through first while others, attempting to call simultaneously, get busy signals.) A worker who receives an offer immediately compares it with any offer (or position) she may be holding, and rejects one (and holds the other). A worker who is holding no position or offer is also free to reject an offer just received. However, a (temporarily or initially) matched worker is allowed to reject her current position only when she receives and accepts an alternative offer.

The process continues as long as there exists at least one firm which has a vacant position (i.e., neither occupied by an incumbent worker nor held by a worker to whom the position has been offered) and which wishes to make an offer. The game ends when no firm with a vacant position wishes to make any further offers, at which point each worker who has received and accepted offers during the game is matched with the offer she is holding at the end, and each worker who has received no offers or rejected all her offers is matched as in the initial matching μ'. During the course of the game each firm learns only if its own position becomes vacant, and if an offer it makes is rejected. Each worker learns only if it receives an offer from a given firm (and the order in which offers are received). Thus no agent in the market learns of the actions of others, except as they directly impinge on it/her.

That is, the game tree begins with a node at which nature chooses a firm at random from among those with a vacant position at the initial matching μ'. The selected firm chooses a worker, who accepts it (in which case she rejects her current position, if any) or rejects it, and in either case the next

[20] No new entries or retirements will occur during the course of the game; i.e., we model the process of making offers as fast relative to the frequency of retirements. This assumption is especially apt in markets in which contracts are seasonal on a common calendar, e.g., university professors and head football coaches.

[21] The restriction that a firm does not make an offer twice to the same worker in a given hiring season can be thought of as reflecting some unmodeled institutional features of the game (e.g., about the hiring committees of firms, or about attributes of the worker signalled by rejections). However we impose this restriction primarily for simplicity, and it does not appear that relaxing it would materially change the conclusions when the initial matching is firm-quasi-stable.

node is again one at which nature chooses a firm from among those (now) eligible to make offers.

Each information set of a firm is identified by whether its incumbent worker (if any) has vacated its position (either by retirement or after receiving another offer), and, if so, by an ordered list of workers to which the firm has already offered the position, and if they have rejected the offer.[22] Consequently a strategy for a firm, which must decide what worker if any to make the next offer to at any information set at which it has a vacant position, can be identified with a preference ordering (not necessarily the firm's true preferences, and different strategy sets will be reached when a firm plays different preferences).

Similarly each information set of a worker can be identified by her initial outcome and an ordered list of firms which have made her an offer, and her responses, together with the current offer. Because firms make offers one at a time, the information sets at which a worker is called on to make a decision have at most two unrejected offers, a new offer and a previously received offer that was held. The worker must decide which offer to reject and which to hold. A natural class of worker strategies is therefore the class of "preference strategies," i.e., strategies which at any information set tell the worker to reject the offer that is lower on some fixed preference ordering (not necessarily the same as the worker's true preferences). However, in contrast to the firms, workers also have strategies which are not preference strategies, but may depend on the history of offers received. For example, the strategy of always holding any new offer and rejecting the old one is feasible, since workers know the order in which offers have been made.

Note, however, that on any play of the game (both in and out of equilibrium), the choices actually made by any worker in the course of the game must always be *consistent* with some preference ordering, because no firm ever makes an offer to the same worker twice. We can therefore speak (despite the fact that not all strategies are preference strategies) of the preferences "revealed" in any play of the game. Of course only partial orderings may be "revealed" in this way; e.g., if a worker holds an offer from firm f while rejecting one from f', and then receives and rejects an offer from f'' while continuing to hold f, this "reveals" an ordering in which

[22] Note that we do not include in a firm's information set any information about the time that has elapsed, e.g., since an offer was made. This keeps firms' strategy space simple and can be thought of as modeling delays as uninformative, i.e., workers do not formally "hold" offers, and the absence of a rejection might be due to communication delays. So the modeling assumption is that firms learn of workers' actions only when their offers are rejected, or at the end of the game, when their offers are accepted. Also, by modeling each firm's (and worker's) information sets in this way, we are assuming that players cannot update their inferences about one another's characteristics in the course of the game (the simplest interpretation is that players know each other's true preferences).

f is preferred to both f' and f'', but gives no information about comparisons between f' and f''. Thus there may be many particular preference orderings consistent with the choices of a given agent.[23] However, since only the parts of such a preference that are actually revealed in the course of play will effect the outcome of play, for many purposes we will be able to treat any preference consistent with the play as representative of all such preferences. In particular, if P' is a profile of preferences consistent with those "revealed" by the choices made in a play of the game with initial matching μ', then the outcome of this play of the game equals the outcome of the corresponding play (i.e., the same selection of the order in which firms make offers) where the agents select the preference strategies P'. Henceforth, the term "revealed preferences" will denote an arbitrary representative of the preference lists consistent with the choices of the agents under the given play.

In comparing strategic choices with true preferences or with other possible choices we will therefore often be comparing different preference profiles. As before, we will denote particular preferences by adding superscripts, e.g. we write $v' >_v^{P'} v''$ to denote that v prefers v' to v'' under P'_v. We omit the superscript when no ambiguity arises. Given a strategy profile σ and $v \in F \cup W$, we sometimes write, $\sigma = (\sigma_{-v}, \sigma_v)$ where σ_v denotes the strategy of v and σ_{-v} denotes the strategy profile of the other agents. This notation allows us to consider alternative strategies (σ_{-v}, σ'_v) that differ from σ only in v's choice. Similarly, for a given set of agents U we sometimes write $(\sigma_{V \setminus U}, \sigma'_U)$ to denote strategy profiles that differ from σ only in the strategy choices of the agents in U. Preference strategies will be denoted by the corresponding preference.

It is natural to compare the decentralized game described above with the *Centralized Deferred Acceptance* (*CDA*) *game* which occurs in several of the entry level labor markets studied in [26, 29, 30, 35]. Although in those markets the game begins from the empty matching, we can also consider the centralized game starting from any firm-quasi-stable matching. The key difference between the centralized and decentralized games is that, in the centralized game, all agents submit a list of preferences to a central market clearing mechanism, which then conducts the deferred acceptance algorithm using the submitted preference lists. To be precise, we say that the market (F, W, P) and initial matching μ' define a pair of corresponding games, the CDA game and the DDA game, which differ only in that the strategies in the CDA game are restricted to the preference strategies of the

[23] And for a player who is using non-preference strategies, different preferences may be revealed by different plays of the game (i.e., different random selections on the order in which firms make offers, for given strategy choices of the other players), as well as in response to different choices of the other agents.

DDA game. Thus in the CDA game, all players have only preference strategies, whereas workers in the DDA game have a larger class of strategies. Nevertheless, as we will show in Lemma 7.1, equilibria (in preference strategies) of the CDA game are also equilibria of the DDA game.

The following result interprets Theorem 3.1 in the context of the Centralized and Decentralized Deferred Acceptance games introduced in this section.

Corollary 6.1. *In the CDA and DDA games with an arbitrary input matching, all plays with a given profile of preference strategies have a common output outcome.*

Proof. Apply Theorem 3.1 on the revealed preferences. ∎

Corollary 6.1 shows that in the centralized game all plays with given preference-strategies and given initial matching end up with a common output matching. The following example shows that this need not be the case in the decentralized game (when the workers play non-preference strategies).

Example 1. A market in which the output matching in the decentralized game depends on particular plays:

Consider a market with two firms and one worker in which the input is the empty matching. Assume that the two firms choose the (preference) strategy of proposing their position to the worker and the worker chooses the (non-preference) strategy of accepting the first offer and rejecting any later offer. Evidently in this game the output depends on the selection of the order by which the firms make their offers. In particular, the firm which makes the first offer is matched with the worker.

However we will see in Section 7 that, even in the decentralized game, there are robust equilibria in which workers as well as firms all employ preference strategies.

7. STRATEGIC QUESTIONS

In this section we discuss the strategic environment facing the agents in the decentralized deferred acceptance game, with true preferences P. While it will be important for the results concerning the existence of equilibria and the stability of equilibrium outcomes that the initial matching μ' is firm-quasi-stable with respect to P, not all of the intermediate results depend on this assumption.

Example 1 showed that when workers use non-preference strategies in the decentralized game, the outcome may depend on the random order in which firms are selected to make offers. The study of equilibria in the decentralized game could therefore require us to consider not merely agents' preferences over riskless outcomes, but also over lotteries. (This is so even for pure strategy equilibria in preference strategies, which Corollary 6.1 showed produce riskless outcomes, because evaluating potential *deviations* from equilibrium could involve non-preference strategies which induce lotteries over outcomes.) That is, Nash equilibrium strategy profiles are those such that no player can gain a higher expected utility by unilaterally changing strategies, and to study these we might have to consider each agent's utility function. However this turns out not to be necessary, because we now show that equilibria in preference strategies possess a property that makes them dependent only on the ordinal preferences and independent of the risk preferences of the firms and workers.

To formally consider how the random elements in the CDA and DDA games interact with the strategic choices of the players, we have to define a sample space over which the lotteries are considered. In particular, a random element enters the game every time a firm is selected to make a new offer. This is not a simple randomization over firms, because only firms which have a vacant position and have not yet been rejected by all of their acceptable workers may be selected, and which firms meet this condition is determined both by the random selection of firms at previous times in the game and by the strategy choices of all the players.

We will consider a sample space consisting of infinite sequences of firms in which every firm appears infinitely many times. Each point in the sample space corresponds to an order at which firms are given the opportunity to make an offer, where matched firms or firms which have already proposed to all their acceptable workers are skipped. That is, a given point (i.e., infinite sequence) in the sample space determines the random selection of firms as follows. Starting from the last firm in the sequence to have made an offer, elements of the sequence are considered and discarded until the first firm is reached which is eligible to make an offer, and that firm makes the next offer. The game ends whenever all firms remaining in the sequence are ineligible to make an offer (i.e., when no firm has both a vacant position and an acceptable worker who has not yet rejected it). The requirement that each firm appears infinitely many times is stronger than needed (sufficiently long finite sequences would do), but assures that all plays of the game terminate if and only if the deferred acceptance procedure is exhausted. The sample space is denoted by $\mathcal{O}$.

A strategy profile σ is a *Realization-Independent* (*RI*) *equilibrium* in the DDA game with initial matching μ' if there is no firm or worker $v \in V$ and (preference or non-preference) strategy σ'_v for v such that v prefers (according to the true preferences P_v) the outcome that results from the play of the game with strategy profile (σ_{-v}, σ'_v) at even one sample point $o \in \mathcal{O}$ to the outcome that results from the play of the game with strategy profile $\sigma = (\sigma_{-v}, \sigma_v)$ with the order of proposals that, again, is induced by o. The unusual feature of this definition, which makes it much stronger than the usual definition of an equilibrium, is that at an RI equilibrium no player can profit from a strategy deviation at *any* point in the sample space. That is, if σ is an RI equilibrium, not only is no deviation profitable in expected utility, but, in addition, no deviation has even a positive probability of being profitable.

Note that if σ is a preference strategy profile, then all plays of the game with strategy profile σ have a common outcome (Corollary 6.1). Hence σ is an RI equilibrium if and only if for every agent v and (preference or non-preference) strategy σ'_v for v, v does not prefer even *one* outcome that can result with positive probability from a play of the game with strategy profile (σ_{-v}, σ'_v) to the (fixed) outcome that results from all plays of the game with strategy profile σ.

The following lemma shows that preference strategy equilibria in the DDA game are all RI equilibria, and also establishes the relation between equilibria of the centralized and decentralized deferred acceptance games.

LEMMA 7.1. *Let P' be a profile of preference strategies and let μ' be an arbitrary matching. Then the following are equivalent*:

1. *P' is an RI equilibrium of the* DDA *game with initial matching μ'.*
2. *P' is an RI equilibrium of the* CDA *game with initial matching μ'.*
3. *P' is a Nash equilibrium of the* DDA *game with initial matching μ'.*
4. *P' is a Nash equilibrium of the* CDA *game with initial matching μ'.*

Proof. In the CDA game all strategies are preference strategies. Since Corollary 6.1 implies that every preference strategy is associated with a single outcome, the notions of Nash equilibria and RI equilibria in the CDA game coincide trivially. Thus $(2) \Leftrightarrow (4)$.

Next, assume that P' is a Nash equilibrium in either the CDA or DDA game with initial matching μ', but that it is not an RI equilibrium in the DDA game. Then there is a worker w and a nonpreference strategy σ_w for w such that w prefers the outcome of *some* play with the strategy profile (P'_{-w}, σ_w) (call this outcome μ) to the common outcome of the plays of the game with the preference strategy profile P'. Let P''_w be any preference relation for w consistent with the choices of w "revealed" in the course of the play of the strategy profile (P'_{-w}, σ_w) which produces μ, and let $P'' = (P'_{-w}, P''_w)$. Then the particular play that leads to μ under the strategies (P'_{-w}, σ_w) is a feasible play under $P'' = (P'_{-w}, P''_w)$, and by Corollary 6.1 *all* plays of P'' lead to w's preferred outcome μ. Hence P' is not a Nash equilibrium in the CDA or DDA games, contrary to our assumption. So, $(3) \Rightarrow (1)$ and $(4) \Rightarrow (1)$

Finally, assume that P' is an RI equilibrium in preference strategies in the DDA game with initial matching μ'. Then trivially, P' is a Nash equilibrium of the DDA game with initial matching μ'. Also, as the strategies for the CDA game are a subset of those of the DDA game, P' is also a Nash equilibrium in the CDA game with initial matching μ'. So, $(1) \Rightarrow (3)$ and $(1) \Rightarrow (4)$. ∎

Lemma 7.1 shows that in both the CDA and the DDA games the set of RI equilibria in preference strategies coincides with the set of Nash equilibria in preference strategies. Henceforth, when considering preference strategies we will omit the prefix RI or Nash and simply refer to *equilibrium*.

Lemma 7.1 identifies a natural class of pure-strategy equilibria of the DDA game in which only preference strategies are employed (below, Theorem 7.6 shows that such equilibria always exist.) Corollary 6.1 implies that at such an equilibrium, the selection of firms within a play of the DDA game does not influence the outcome. Thus when players behave strategically, just as when they behave straightforwardly, the order in which firms make offers need not add any randomness to the final outcome of the observed vacancy chain.

We are now going to show that when the other agents choose the preference strategies corresponding to their true preferences, no group of agents containing at least one firm that is "active" in the plays of the game which occur under the true preferences can strictly improve each group member's outcome by choosing (preference or non-preference) strategies different from their true preferences. This result generalizes a theorem of Dubins and Freedman [11] for the case where $\mu' = \varnothing$; for details see [32, Theorem 4.10, p. 92]. This will imply that in the CDA game it is a dominant strategy for every firm to reveal its true preferences. This will then allow us to derive properties of equilibria in the DDA game as well.

We will use a Blocking Lemma which generalizes a result of Gale and Sotomayor [13]; see [32, Lemma 3.5, p. 56]. The strategy of the proof will be to use the Blocking Lemma to identify a blocking pair which would exist if some active coalition of firms and workers could all profit from misrepresenting their preferences. The existence of this blocking pair would contradict the μ'-stability of the outcome (established in Lemma 3.5).

We will require some new notation. Given two matchings μ and μ'', let

$$F(\mu > \mu'') \equiv \{f \in F: \mu(f) >_f \mu''(f)\}$$

$$F(\mu = \mu'') \equiv \{f \in F: \mu(f) = \mu''(f)\}.$$

Also, given a (reference) matching μ' let

$$F^{\mathrm{I},\mu'}(\mu) \equiv \{f \in F\colon \mu(f) = \mu'(f) \in W\}.$$

If μ' is some initial matching and $\mu = \mathrm{DA}(\mu')$ then the set $F^{\mathrm{I},\mu'}$ is the set of *inactive* firms, who make no offers in the course of the deferred acceptance procedure. Note also that if μ is μ'-stable and either $\mu(f) <_f f$ or f is in a blocking pair for μ, then necessarily $f \in F^{\mathrm{I},\mu'}(\mu)$ (see Lemma 3.3). Thus, $F^{\mathrm{I},\mu'}(\mu)$ is the set of firms that are the only potential candidates to violate acceptability or be part of a blocking pair for μ'-stable matchings, and in this spirit, the superscript I could also stand for *instability*.

The next lemma identifies useful partitions of F.

LEMMA 7.2. *Let μ, μ' and μ'' be matchings. Then the sets $F(\mu > \mu'') \cap F^{\mathrm{I},\mu'}(\mu'')$, $F(\mu > \mu'') \backslash F^{\mathrm{I},\mu'}(\mu'')$, $F(\mu'' > \mu) \cap F^{\mathrm{I},\mu'}(\mu)$, $F(\mu'' > \mu) \backslash F^{\mathrm{I},\mu'}(\mu)$ and $F(\mu = \mu'')$ form a partition of F.*

LEMMA 7.3. (Blocking Lemma). *Let μ' be a matching and let $\mu \equiv DA(\mu')$. Assume μ'' is a matching such that*

1. $\mu'' \geqslant_W \mu'$.
2. *$\mu'(w) \in F$ implies $\mu''(w) \in F$, for each $w \in W$.*
3. $F(\mu'' > \mu) \backslash F^{\mathrm{I},\mu'}(\mu) \neq \varnothing$.

Then there exists a blocking pair (f, w) for μ'' with $\mu(w) >_w \mu''(w)$, $\mu(f) \geqslant_f \mu''(f)$ and $f \notin F^{I,\mu'}(\mu'')$.

Proof. Let

$$F' \equiv [F(\mu'' > \mu) \backslash F^{\mathrm{I},\mu'}(\mu)] \cup [F(\mu > \mu'') \cap F^{\mathrm{I},\mu'}(\mu'')];$$

then by assumption (3), $F' \neq \varnothing$. Let $\mu(F') \equiv \{\mu(f')\colon \mu(f') \in W,\ f' \in F'\}$ and $\mu''(F') \equiv \{\mu''(f')\colon \mu''(f') \in W,\ f' \in F'\}$. As μ is μ'-stable (Lemma 3.5), we have that if $\mu(f') <_{f'} f'$ then $\mu'(f') = \mu(f') \in W$. So, $\mu(f') \geqslant_{f'} f'$ for each $f' \in F \backslash F^{\mathrm{I},\mu'}(\mu)$. Thus, if $f' \in F(\mu'' > \mu) \backslash F^{\mathrm{I},\mu'}(\mu)$, then $\mu''(f') >_{f'} \mu(f') \geqslant_{f'} f'$, assuring that $\mu''(f') \in W$. Further, for each $f' \in F(\mu > \mu'') \cap F^{\mathrm{I},\mu'}(\mu'')$, $\mu''(f') = \mu'(f') \in W$. It follows that $\mu''(f') \in W$, for each $f' \in F'$. So, $|\mu''(F')| = |F'| \geqslant |\mu(F')|$. We next consider two cases:

Case (i)—$\mu''(F') \neq \mu(F')$: As $|\mu''(F')| \geqslant |\mu(F')|$ we have that $\mu''(F') \backslash \mu(F') \neq \varnothing$. Let $w \in \mu''(F') \backslash \mu(F')$ and let $f' \equiv \mu''(w) \in F'$. We claim that $\mu(w) >_w \mu''(w)$. Indeed, if $f' \in F(\mu'' > \mu) \backslash F^{\mathrm{I},\mu'}(\mu)$ then the μ'-stability of μ and the fact that $f \notin F^{\mathrm{I},\mu'}(\mu)$ imply that (f', w) is not a blocking pair for

μ. As $w = \mu''(f') >_{f'} \mu(f')$ it follows that $\mu(w) >_w f' = \mu''(w)$. If, alternatively, $f' \in F(\mu > \mu'') \cap F^{\mathrm{I},\mu'}(\mu'')$, then $\mu(f') \neq \mu''(f') = \mu'(f') = w$, implying that $\mu(w) \neq \mu'(w) = \mu''(w) = f'$. Thus, the μ'-acceptability of μ implies that $\mu(w) >_w \mu'(w) = \mu''(w)$. We next show that $\mu(w) \in F$. If $\mu'(w) \in F$ then the rules of the DA Algorithm imply that $\mu(w) \in F$ (as a matched worker rejects her current position only for accepting an alternative proposal); if, alternatively, $\mu'(w) = w$ then the established inequality $\mu(w) >_w \mu''(w)$ and the assumption $\mu'' \geqslant_W \mu'$ imply that $\mu(w) >_w \mu''(w) \geqslant_w \mu'(w) = w$. So, indeed, $\mu(w) \in F$. Let $f \equiv \mu(w) \in F$. As $\mu(f) = w \notin \mu(F')$, $f \notin F'$. Further, as $f' = \mu''(w) \in F'$ it follows that $f \neq f'$, implying that $\mu''(f) \neq \mu''(f') = w = \mu(f)$. So, $f \notin F(\mu = \mu'')$. Also, the established inequality $\mu(w) >_w \mu''(w)$ and assumption (1) imply that $\mu(w) >_w \mu''(w) \geqslant_w \mu'(w)$. Hence, $f = \mu(w) \neq \mu'(w)$, implying that $f \notin F^{\mathrm{I},\mu'}(\mu)$. So, $f \notin F'$, $f \notin F(\mu = \mu'')$ and $f \notin F^{\mathrm{I},\mu'}(\mu)$. By Lemma 7.2 we conclude that $f \in F(\mu > \mu'') \backslash F^{\mathrm{I},\mu'}(\mu'')$. So, $w = \mu(f) >_f \mu''(f)$ and $f \notin F^{\mathrm{I},\mu'}(\mu'')$. As it was already shown that $f = \mu(w) >_w \mu''(w)$, we have that (f, w) is a blocking pair for μ'' that satisfies the corresponding properties (in particular, $f \notin F^{\mathrm{I},\mu'}(\mu'')$).

Case (ii)—$\mu''(F') = \mu(F')$: In this case $|\mu''(F')| = |\mu(F')| = |F'|$ and therefore $\mu(f') \in W$ for each $f' \in F'$. Consider an execution of the DA Algorithm with input μ'. Assumption (3) assures that for some $f' \in F'$, $f \notin F^{\mathrm{I},\mu'}(\mu)$. In particular, as $\mu(f') \in W$ it follows that $\mu(f') \neq \mu'(f')$; hence f' switches mates along the considered execution of the DA Algorithm. Let f^* be the last firm in F' to make a successful offer to a worker, i.e., a proposal that is accepted, and let w be the worker who receives this proposal. As $\mu(f') \in W$ for each $f' \in F'$, at that stage all other firms of F' are matched, and further, $w = \mu(f^*)$. Also, as $w \in \mu(F') = \mu''(F')$, we have that $f'' \equiv \mu''(w) \in F'$, and as $F' \cap F(\mu = \mu'') = \emptyset$, we have that $w = \mu(f^*) \neq \mu''(f^*)$, implying that $f^* = \mu(w) \neq \mu''(w) = f''$. We next show that f'' has been rejected by w along the execution, when or before she accepts the offer of f^*. As $f'' \in F'$ we consider two alternatives. If $f'' \in F(\mu > \mu'') \cap F^{\mathrm{I},\mu'}(\mu'')$, then $\mu'(f'') = \mu''(f'') = w = \mu(f^*) \neq \mu(f'')$, so w is initially matched to f'' (under μ') and eventually she rejects f'' at the first time she switches a mate. Alternatively, if $f'' = \mu''(w) \in F(\mu'' > \mu) \backslash F^{\mathrm{I},\mu'}(\mu)$, then $w = \mu''(f'') >_{f''} \mu(f'')$ and f makes proposals along the considered execution. Therefore the rules of the DA Algorithm imply that f'' is rejected by w along the execution. So, in both cases we have that f'' has been rejected by w along the execution. But, the definition of f^* as the firm in F' which makes the last successful offer to a worker, the fact that $f'' \in F'$ and the fact that $\mu(f') \in W$ for each $f' \in F'$ assure that the rejection of f'' by w occurs when or before the offer of f^* to w takes place. Now, by assumption (1), $f'' = \mu''(w) \geqslant_w \mu'(w)$ and the rejection of f'' by w must occur for an alternative preferred firm (the cases $f'' = \mu'(w)$ and $f'' >_w \mu'(w)$ must be considered separately). As the rules of the DA Algorithm assure that w is matched at all stages following such a rejection, we conclude that w is matched to a firm when receiving and accepting the offer of f^*, say to f, and further, the firm f is weakly preferred by w to f''. We then have that $f^* = \mu(w) >_w f \geqslant_w f''$. But, as f is unmatched after its rejection by w (for f^*), as that stage was the last in which an offer of a member of F' is accepted and as $\mu(f') \in W$ for each $f' \in F'$, we conclude that $f \in F \backslash F'$; in particular $f \neq f''$ and $f >_w f'' = \mu''(w)$. Also, as f is rejected by w (for f^*), $w >_f \mu(f)$. Now, since $f \notin F'$, Lemma 7.2 implies that f belongs to exactly one of the sets $F(\mu = \mu'')$, $F(\mu > \mu'') \backslash F^{\mathrm{I}, \mu'}(\mu'')$ or $F(\mu'' > \mu) \cap F^{\mathrm{I}, \mu'}(\mu)$. As f is unmatched at some stage of the execution (namely, after it is rejected by w), either $\mu(f) \neq \mu'(f)$ or $\mu(f) = \mu'(f) = f$; hence, $f \notin F^{\mathrm{I}, \mu'}(\mu)$. Therefore $f \notin F(\mu'' > \mu) \cap F^{\mathrm{I}, \mu'}(\mu)$. Further, if $f \in F(\mu = \mu'')$ then $f \notin F^{\mathrm{I}, \mu'}(\mu'')$ since otherwise $f \in F^{\mathrm{I}, \mu'}(\mu)$ in contradiction to the fact that $f \notin F^{\mathrm{I}, \mu'}(\mu)$; so, if either $f \in F(\mu = \mu'')$ or $f \in F(\mu > \mu'') \backslash F^{\mathrm{I}, \mu'}(\mu'')$ we have that $f \notin F^{\mathrm{I}, \mu'}(\mu)$. It follows that $\mu(f) \geqslant_f \mu''(f)$, implying that $w >_f \mu(f) \geqslant_f \mu''(f)$. As it was shown that $f >_w \mu''(w)$, we have that (f, w) is a blocking pair for μ''. Further, the asserted requirements $\mu(w) >_w \mu''(w)$, $\mu(f) \geqslant_f \mu''(f)$ and $f \notin F^{\mathrm{I}, \mu'}(\mu'')$ have been established. ∎

We next show that no coalition U of players containing at least one active firm can all profit by adopting (preference or non-preference) strategies which differ from their true preference strategies.

Theorem 7.4. *Let μ' be a matching and $\mu \equiv DA^P(\mu')$. Let $U \subseteq V$ be a group of agents such that*

$$U \cap [F \backslash F^{I, \mu'}(\mu)] \neq \varnothing.$$

Let $\sigma = (P_{V \backslash U}, \sigma_U'')$ be some strategy profile, $o \in (\mathcal{O})$ be some sample point and μ'' be the output of the play of the game with strategy profile σ'' under the order of selections induced by o. Then $\mu(v) \geqslant_v^P \mu''(v)$ for some $v \in U$.

Proof. We first consider the case that $\sigma_U'' = P_U''$ consists entirely of preference strategies. In this case $\mu'' = \mathrm{DA}^{P''}(\mu')$. Assume, by way of contradiction, that $\mu''(v) >_v^P \mu(v)$ for each $v \in U$. Then the assumption that $U \cap [F \backslash F^{\mathrm{I}, \mu'}(\mu)] \neq \varnothing$ implies that $F(\mu'' > \mu) \backslash F^{\mathrm{I}, \mu'}(\mu) \neq \varnothing$. Further, the rules of the DA Algorithm imply that for each $w \in W$, $\mu''(w) \in F$ whenever $\mu'(w) \in F$. Also, $\mu = \mathrm{DA}^P(\mu') \geqslant_W^P \mu'$ (Lemma 3.5) and therefore $\mu'' >_{W \cap U}^P \mu \geqslant_{W \cap U}^P \mu'$. In addition, $\mu'' = \mathrm{DA}^{P''}(\mu') \geqslant_W^{P''} \mu'$ (Lemma 3.5) and by our assumption $P_{W \backslash U} = P''_{W \backslash U}$; hence $\mu'' \geqslant_{W \backslash W'}^P \mu'$. So, $\mu'' \geqslant_W^P \mu'$. We established that the three conditions of Lemma 7.3 are satisfied. By the conclusions of the lemma there exists a blocking pair (f, w) for μ'' under P where $\mu(w) >_w^P \mu''(w)$, $\mu(f) \geqslant_f^P \mu''(f)$ and $f \notin F^{\mathrm{I}, \mu'}(\mu'')$. Now, by the assumption that $\mu''(v) >_v^P \mu(v)$ for each $v \in U$ we have that $f, w \notin U$. Thus $P_f'' = P_f$ and $P_w'' = P_w$; hence (f, w) is a blocking pair for μ'' under P'' where

$f \notin F^{\mathrm{I},\mu'}(\mu'')$. This conclusion contradicts the μ'-stability of μ'' under P'' (see Lemma 3.3).

Finally, consider the case that σ''_U does not consist entirely of preference strategies and suppose that the conclusion of the theorem did not hold (for the sample point $o \in \mathcal{O}$). Then, for this sample point, the same outcome would also be produced if each non-preference strategy in σ''_U were replaced by (any of) the preference strategies P''_U "revealed" in the course of the play of the game at that sample point. But these preference strategies would then violate the conclusions of the theorem (at every sample point, by Corollary 6.1), and we have just proved that for preference strategies this cannot happen, so it also cannot happen when non-preference strategies are employed. ▮

The following example shows that the assumption that $U \cap (F \backslash F^{\mathrm{I},\mu'}(\mu)) \neq \emptyset$ cannot be relaxed to $U \cap F \neq \emptyset$, i.e. that it is necessary to consider coalitions U which include an active firm.

EXAMPLE 2. A marriage market and (*non*-firm-quasi-stable) initial matching where one worker and one firm can both profit by deviating from their "true" preference orders, while the others play the "true" preferences: Consider a market with 2 firms and 3 workers where

$$P_{f_1} = w_3, w_1, w_2 \qquad P_{w_1} = f_1, f_2$$

$$P_{f_2} = w_1, w_2 \qquad P_{w_2} = f_2, f_1$$

$$P_{w_3} = f_1.$$

Let the initial matching be $\mu' = \{(f_1, w_2)\}$. Then $\mu \equiv \mathrm{DA}^P(\mu') = \{(f_1, w_2), (f_2, w_1)\}$. Now, with $P''_{f_1} = w_1, w_2$, $P''_{w_1} = f_1$ and $P'' \equiv (P_{-\{f_1, w_1\}}, P''_{w_1}, P''_{f_1})$ we have that $\mu'' \equiv \mathrm{DA}^{P''}(\mu') = \{(f_1, w_1), (f_2, w_2)\}$ where both f_1 and w_1 prefer μ'' to μ. Hence, f_1 and w_1 (who also formed a blocking pair for the initial matching) have incentives to arrange an agreement to both deviate from P to P''.

Note, however, that once f_1 and w_1 reach an agreement to switch their preferences and w_1 plays the preference strategy P''_{w_1}, f_1 has no incentive to keep its part in the agreement and it will be better by playing its true preference order. In this case the output matching is $\{(f_1, w_3), (f_2, w_2)\}$ which f_1 prefers to μ''. In fact, we will see in Theorem 7.5 below that it is always a best response for each firm to play its true preferences when other agents play preference strategies.

Call a strategy *dominant* for an agent v if for every selection of strategies by the other agents and for every selection of the order by which the firms make their offers, no improvement of the outcome for v is possible through a change of its/her strategy. It is known that in the centralized deferred acceptance game with the empty matching as input it is a dominant strategy for every firm to use the preference strategy corresponding to its true preferences [25] or [32, Theorem 4.7]. We next generalize this result for any (firm-quasi-stable) initial matching.

THEOREM 7.5. *In the* CDA *game with arbitrary initial matching, it is a dominant strategy for every firm to play the preference strategy that corresponds to its true preferences.*

Proof. Let μ' be a matching and let $f \in F$ and consider some preference strategy profile P'_{-f} of the other agents. Let $P^* \equiv (P'_{-f}, P_f)$ and $\mu^* \equiv \mathrm{DA}^{P^*}(\mu')$. If $f \in F^{\mathrm{I},\mu'}(\mu^*)$, that is, $\mu^*(f) = \mu'(f) \in W$ then f never makes an offer during any execution of the DA Algorithm; hence the preference order of f does not influence the output matching. Thus, for each preference list P''_f for f and $P'' \equiv (P'_{-f}, P''_f)$, $\mathrm{DA}^{P''}(\mu') = \mathrm{DA}^{P^*}(\mu')$ and so f cannot profit by deviating from the strategy corresponding to its true preferences.

Next, assume that $f \in F \backslash F^{\mathrm{I},\mu'}(\mu^*)$. Then Theorem 7.4 with $U = \{f\}$ and P^* and μ^* standing, respectively, for P and μ, implies that for each preference list P''_f of f and $P'' \equiv (P'_{-f}, P''_f)$, $[\mathrm{DA}^{P^*}(\mu')](f) \geqslant_f^{P^*} [\mathrm{DA}^{P''}(\mu')](f)$. As $P^*_f = P_f$, it follows again that f cannot profit by deviating from the strategy corresponding to its true preferences. ∎

Note that the above theorem is defined in terms of a game whose strategy space really makes the most sense if the initial matching μ' is firm-quasi-stable with respect to the true preferences P. (If μ' could be unacceptable with respect to the true preferences, for example, it would be unreasonable to have modeled the strategy sets of the firms in a way that prevents them from summarily firing unacceptable, but incumbent, workers.) However the proof of the theorem does not depend on the assumption of firm-quasi-stability.

The arguments of the proof of Theorem 7.5 show that in the DDA game each firm's true preference remains its best response when all workers play preference strategies. The next example shows that truth revealing need not be a dominant strategy in the DDA game and a firm may benefit by deviating from its true preferences when non-preference strategies are used by the other players. The example illustrates that the conclusion is not true even for the special case where the input is the empty matching.

EXAMPLE 3. A game in which true preferences are not a dominant strategy for all firms:

Consider a market with 3 firms and 2 workers where

$$P_{f_1} = w_1, w_2 \qquad P_{w_1} = f_3$$
$$P_{f_2} = w_2 \qquad P_{w_2} = f_1, f_2$$
$$P_{f_3} = w_1$$

Let the initial matching be $\mu' = \emptyset$, and suppose the order in which the firms make offers is the (degenerate random) cyclic order $f_1, f_2, f_3, f_1, \ldots$ Suppose further that worker w_1 and firms f_2 and f_3 play according to their true preferences, and that worker w_2 plays according to the (non-preference) strategy $\sigma_{w_2} =$ "accept whichever of $\{f_1, f_2\}$ proposes first, and reject all other offers." Then if f_1 plays according to its true preferences, it will be unmatched at the outcome resulting from the strategy choices (P_{-w_2}, σ_{w_2}). But, if f_1 plays according to the preference strategy $P'_{f_1} = w_2$ then it will be matched to w_2 at the outcome resulting from the strategy choices $(P_{-\{w_2, f_1\}}, P'_{f_1}, \sigma_{w_2})$. Thus f_1 does better by playing P'_{f_1} than by playing according to its true preferences.

This is a good place to discuss for a moment the non-preference strategies of the workers, which, Example 1 in Section 6 and the above example show, make the DDA game potentially quite different from the CDA game. We conjecture that in the present model it might even be possible to eliminate non-preference strategies by some equilibrium refinement (e.g., perhaps non-preference strategies can be deleted in the course of the iterated elimination of dominated strategies). But in richer environments, like those of the entry level markets studied in [35, 36] (e.g., for American lawyers and psychologists, and for graduates of elite Japanese universities), we occasionally see phenomena strikingly like non-preference strategies. For example, there are sometimes incentives for firms to make "exploding offers" which force workers to accept the first reasonable offer they receive, much like strategy σ_{w_2} in Example 3. We will briefly return to this in the conclusion.

Nevertheless, as we have seen in Lemma 7.1, there is a natural class of equilibria of the DDA game in which no non-preference strategies are employed. We will show below (in Theorem 7.6) that this is quite a large class of equilibria. Lemma 7.1 further suggests that, for empirical purposes, this will be a natural class of equilibria to investigate at least initially, since it will yield hypotheses that are not sensitive in critical ways to the unobservable parts of the market which are modeled as random.

We consider next the incentives facing workers. When the initial matching is the empty one, there is at least one worker who can profitably misrepresent her preferences whenever there is more than one stable matching [32, Theorems 4.6, 4.7 and Corollary 4.12]. Further, assuming that the other agents choose to play according to their true preferences, any worker or group of workers with more than one achievable outcome can reveal preferences to compel any jointly achievable outcome (by "jointly" we mean a set of achievable outcomes for a given group of workers that can be simultaneously achieved in a single stable matching). In the forthcoming Theorem 7.6 we discuss generalizations of these results for the unmatched workers. However for the matched workers these conclusions no longer hold. The next two examples illustrate the different situation facing the matched workers when there is a non-empty firm-quasi-stable initial matching.

EXAMPLE 4. A market with two stable matchings preferred by the (matched) workers to the initial matching μ', but nevertheless, no (matched) worker or set of (matched) workers can all profit by deviating from their true preferences:

Consider a market with 4 firms and 4 workers where

$$P_{f_1} = w_2, w_1, w_4, w_3 \qquad P_{w_1} = f_2, f_1, f_3$$
$$P_{f_2} = w_3, w_2, w_4, w_1 \qquad P_{w_2} = f_3, f_2, f_1$$
$$P_{f_3} = w_1, w_3, w_4, w_2 \qquad P_{w_3} = f_1, f_3, f_4, f_2$$
$$P_{f_4} = w_3, w_4 \qquad P_{w_4} = f_4, f_3, f_2, f_1.$$

Let the initial firm-quasi-stable matching be $\mu' \equiv \{(f_1, w_2), (f_2, w_3), (f_3, w_1)\}$. In this market there are two stable matchings:

$$\mu_1 = \{(f_1, w_3), (f_2, w_1), (f_3, w_2), (f_4, w_4)\}$$

and

$$\mu_2 = \{(f_1, w_1), (f_2, w_2), (f_3, w_3), (f_4, w_4)\}.$$

It is easy to verify that μ_2 results from the DA Algorithm when the true preferences P are played. As w_4 gets her most preferred outcome, she can never do better by changing her selected strategy. We next show that any coalition of workers that does not contain w_4 can do no better than μ_2 when w_4 plays according to her true preferences. Assuming that the firms play according to their true preferences the first step of each play is a proposal of f_4 to w_3. If w_3 rejects that proposal then at the next step f_4 will propose to w_4, she will accept the proposal and the matching $\mu' \cup \{(f_4, w_4)\}$ would result. In that case w_3 deviates from her true preferences but her outcome does not improve. Assume now that w_3 accepts the proposal of f_4. Then in the second step f_2 proposes to w_2. If w_2

rejects that proposal than at the next step f_2 will propose to w_4, she will accept the proposal and the matching

$$\{(f_1, w_2), (f_2, w_4), (f_3, w_1), (f_4, w_3)\}$$

would result. In that case w_2 deviates from her true preferences but her outcome does not improve. So, assume that w_2 accepts the proposal of f_2. Similar arguments show that at the third step, when f_1 offers w_1, she accepts its offer (otherwise the matching $\{(f_1, w_4), (f_2, w_2), (f_3, w_1), (f_4, w_3)\}$ will result and w_1 does not profit) and at the forth step, when f_3 offers w_3 she also accepts its offer (otherwise the matching $\{(f_1, w_1), (f_2, w_2), (f_3, w_4), (f_4, w_3)\}$ will result). Now, at the fifth step f_4 oAAers w_4, who accepts its offer, and the game ends with output μ_2.

Note however that if w_4 cooperates with the other workers and chooses the preference strategy in which only f_4 is acceptable, the other workers can reveal preferences such that μ_1 will result (by choosing preference strategies where only $\mu'(w)$ and $\mu_1(w)$ as acceptable). But as we have seen, w_4 has no incentive to do this.

The above example shows that matched workers may be unable to profit by misrepresentation even when none of them starts with an achievable outcome. The next example shows that initial stability is not a good guide in either direction.

EXAMPLE 5. A market in which the (matched) workers have incentives for misrepresentation even though the initial matching is stable:

Consider a market with 3 firms and 2 workers where

$$P_{f_1} = w_1, w_2 \qquad P_{w_1} = f_2, f_1$$
$$P_{f_2} = w_2, w_1 \qquad P_{w_2} = f_1, f_2, f_3.$$
$$P_{f_3} = w_1, w_2$$

Let the (stable) initial matching be $\mu' \equiv \{(f_1, w_1), (f_2, w_2)\}$. If w_2 reveals her true preferences, μ' is the final outcome. However, if w_2 chooses the preference strategy $P'_{w_2} = f_1, f_3, f_2$, the output of the DA Algorithm is the matching $\{(f_1, w_2), (f_2, w_1)\}$ which w_2 prefers to μ'.

Call a preference strategy for an agent a *truncation* if the acceptable part of the declared preference is an upper part of the acceptable set of the agent's true preferences (and the preferences over the unacceptable part of the declared preference list are the same as in the true preferences). The above example illustrates a difference between the CDA game starting with the empty matching and the CDA game starting with an arbitrary firm-quasi-stable matching. In the first case, for every worker w and given some fixed preference strategies P'_{-w} of the other agents, a truncation strategy is w's best response. But, the above example shows that this is not true with an arbitrary firm-quasi-stable initial matching. In the example, the two preference strategies for w_2 that are her best response are $P'_{w_2} = f_1, f_3, f_2$ and $P'_{w_2} = f_1, f_3$, and neither is a truncation strategy.[24]

When the initial matching is empty, any stable matching can result from some equilibrium in preference strategies. This result is proved for the CDA game in [32, Theorem 4.15]; hence, by Lemma 7.1 it holds for the DDA game as well. In particular, an equilibrium in pure strategies always exists in this case. We next generalize this result to the centralized and the decentralized games starting with general firm-quasi-stable initial matchings. It is no longer the case that any stable matching can be achieved at equilibrium as the initial matching constrains what is possible, but any jointly achievable outcome for the *unmatched* workers can result from an equilibrium in which the firms reveal their true preferences. There is also an equilibrium that yields the outcome obtained when all agents reveal their true preferences.

THEOREM 7.6. *Let $\mu' \in Q(P)$. Then $\underline{S}^{\mu'}_W(P) \neq \emptyset$ and for every $\mu \in \underline{S}^{\mu'}_W(P)$ there is a preference strategy profile $P'' \equiv (P_F, P''_W)$ such that:*

1. *P'' is an equilibrium in both the CDA and DDA games with initial matching μ',*
2. *$DA^{P''}(\mu') \in \underline{S}^{\mu'}_W(P)$, and*
3. *$[DA^{P''}(\mu')](w) = \mu(w)$ for each $w \in W$ with $\mu'(w) = w$.*

In particular, if $\mu = DA^P(\mu') = \wedge_W \underline{S}^{\mu'}_W(P)$, then $DA^{P''}(\mu') = \mu$.

Proof. We consider only the CDA game in which workers have only preference strategies as Lemma 7.1 implies that all equilibria constructed for the CDA game are also equilibria for the DDA game.

By Theorem A.6 of the Appendix, $\underline{S}^{\mu'}_W(P) \neq \emptyset$. Let $\mu \in S^{\mu'}_W(P)$, let $W' \equiv \{w \in W: \mu'(w) = w\}$ and let $P' \equiv (P_{V \setminus W'}, P'_{W'})$ where for each $w \in W'$, P'_w is obtained from P_w by truncation just below $\mu(w)$. (Note that the set W' may be empty.) The matching μ is clearly acceptable under P'. Also, as $\mu'(w) = w$ whenever $P_w \neq P'_w$ we have that μ' is acceptable under P'. Next, as P' is a restriction of P, the firm-quasi-stability of μ' and the stability of μ under P imply their firm-quasi-stability and stability, respectively, under P'. Further, by the definition of $\underline{S}^{\mu'}_W(P)$, $\mu \geqslant^P_W \mu'$; hence, as $P_{W \setminus W'} = P'_{W \setminus W'}$, trivially, $\mu \geqslant^{P'}_{W \setminus W'} \mu'$. Also, for $w \in W'$, $\mu(w) \geqslant^{P'}_w w = \mu'(w)$. So,

[24] See Roth and Vande Vate [34], particularly Theorem 3, for discussion of truncation strategies in connection with random matching of the sort considered in Roth and Vande Vate [33].

$\mu \geqslant_W^{P'} \mu'$, and therefore $\mu \in \underline{S}_W^{\mu'}(P')$. As $\mu' \in Q(P')$, Theorem 4.3 implies that $\mu'' \equiv \mathrm{DA}^{P'}(\mu') = \wedge_W \underline{S}_W^{\mu'}(P')$; in particular, $\mu \geqslant_W^{P'} \mu''$. We next argue that $\mu(w) \leqslant_w^{P'} \mu''(w)$, for each $w \in W'$. This conclusion is trivial when $\mu(w) = w$. Alternatively, if $\mu(w) \in F$, then Theorem 2.2 (applied to P') implies that $\mu''(w) \in F$; as $\mu(w)$ is the least-preferred firm by w under P', we have that $\mu(w) \leqslant_w^{P'} \mu''(w)$. The inequality $\mu \leqslant_{W'}^{P'} \mu''$ combines with the established inequality $\mu \geqslant_W^{P'} \mu''$ to show that $\mu(w) = \mu''(w)$ for all $w \in W'$. We next claim that μ'' is stable under the true preferences P. Trivially, as μ'' is acceptable under P' it is acceptable under P. Also, as $\mu'' \in S(P')$, no pair $(f, w) \in A(P') \subseteq A(P)$ is a blocking pair for μ'' under P', immediately implying that such a pair cannot be a blocking pair for μ'' under P. Further, if $(f, w) \in A(P) \backslash A(P')$, then $w \in W'$ and $f <_w^P \mu(w)$; but as $\mu''(w) = \mu(w)$ it follows that $f <_w^P \mu''(w)$ and (f, w) can not be a blocking pair for μ'' under P. Thus there is no blocking pair for μ'' under P and, indeed, $\mu'' \in S(P)$. Now, by Lemma 3.5, $\mu'' \geqslant_W^{P'} \mu'$ and as $\mu''(w) = \mu(w)$ for each $w \in W'$, the construction of P' from P implies that $\mu'' \geqslant_W^P \mu'$ (the cases where $\mu'(w) = w$ and $\mu'(w) >_w w$ must be considered separately). So, $\mu'' \in S(P)$ and $\mu'' \geqslant_W^P \mu'$; hence, $\mu'' \in \underline{S}_W^{\mu'}(P)$.

Next, consider the profile $P'' \equiv (P_F, P''_W)$ where for each $w \in W$, P''_w is the same as P'_w except that all firms that are preferred (under P'_w) to $\mu''(w)$ are no longer acceptable; in particular, if $\mu''(w) = w$ then the set of acceptable firms for w under P''_w is empty, and if $w \in W'$ with $\mu(w) = \mu''(w) \in F$, then $\mu(w) = \mu''(w)$ is the only acceptable firm for w under P''. As $\mu'' \geqslant_W^{P'} \mu'$ and as μ' is acceptable under P', the matching μ' is acceptable under P''. Further, since P'' is a restriction of P', the firm-quasi-stability of μ' under P' implies its firm-quasi-stability under P''. So, $\mu' \in Q(P'')$. We claim that $\mathrm{DA}^{P''}(\mu') = \mu'$. Indeed, consider some execution of Algorithm $\mathrm{DA}^{P'}$ with initial matching μ'. As the outcomes for workers can only improve along such executions and as the output matching under this execution is μ'' it follows that no worker w receives a proposal from a firm which she prefers to $\mu''(w)$. As P'' coincides with P' on the firm's preferences and on the lower part of the worker's preferences bounded from above by the matching μ'', it follows that this execution of Algorithm $\mathrm{DA}^{P'}$ coincides with a corresponding execution of $\mathrm{DA}^{P''}$. Thus, $\mathrm{DA}^{P''}(\mu') = \mathrm{DA}^{P'}(\mu') = \mu'$. In particular, $\mu'' = \mathrm{DA}^{P''}(\mu') \in \underline{S}_W^{\mu'}(P)$ and $[\mathrm{DA}^{P''}(\mu')](w) = \mu'(w) = \mu(w)$ for each $w \in W'$.

It remains to prove that no agent can profit by deviating from P'' (for the moment, to another preference strategy). Since P'' is a preference strategy profile, Theorem 7.5 implies that no firm f can profit by deviating from $P''_f = P_f$. Assume now that some worker w can profit by revealing alternative preferences P^*_w, that is, with $P^* \equiv (P''_{-w}, P^*_w)$ and $\mu^* \equiv \mathrm{DA}^{P^*}(\mu')$ we have that $f \equiv \mu^*(w) >_w^P \mu''(w) \geqslant_w^P \mu'(w)$ (where the right inequality was established earlier). Now, as $f >_w^P \mu''(w) \geqslant_w^P w$ and $w = \mu^*(f) \geqslant_f f$, the stability of μ'' under P implies that $w' \equiv \mu''(f) >_f^P w = \mu^*(f) \geqslant_f f$ and that $w' \in W$; as $P_f = P_f^*$, the latter inequality implies that $w' >_f^{P^*} \mu^*(f)$. Now, since $w' >_f^P w$, then $w' \neq w$ and $f = \mu''(w')$ is the w'-most-preferred firm under $P_{w'}^* = P_{w'}''$; hence, $f \geqslant_{w'}^{P^*} \mu^*(w')$ and as $f \neq \mu^*(w')$, $f >_{w'}^{P^*} \mu^*(w')$. So, (f, w') is a blocking pair for μ^* under P^*. Thus by the μ'-stability of μ^* under P^*, $w = \mu^*(f) = \mu'(f)$; but the above shows that $f \neq \mu'(w)$. This contradiction proves that worker w cannot profit by deviating from P_w''.

Finally, suppose $\mu = \mathrm{DA}^P(\mu') = \wedge_W \underline{S}_W^{\mu'}(P)$ (see Theorem 4.3). As $\mu'' \in \underline{S}_W^{\mu'}(P)$, $\mu'' \geqslant_W^P \mu = \mathrm{DA}^P(\mu') = \wedge_W \underline{S}_W^{\mu'}(P)$. It was also shown that $\mu \geqslant_W^{P'} \mu''$ and that $\mu(w) = \mu''(w)$ for each $w \in W'$; hence, the construction of $P' = (P_{-W}, P'_W)$ implies that $\mu \geqslant_W^P \mu''$. As $\mu \geqslant_W^P \mu''$ and $\mu'' \geqslant_W^P \mu$, we conclude that $\mu = \mu''$. ∎

Arguments similar to those used in the first part of the proof of Theorem 7.6 (the construction of P') show that any set of unmatched workers can compel any jointly achievable outcome for them in the set $\underline{S}_W^{\mu'}(P)$ (when the other agents reveal their true preferences). In particular, as $\mathrm{DA}^P(\mu') = \wedge_W \underline{S}_W^{\mu'}(P)$, it follows that there is a worker who has incentive to deviate from her true preferences whenever there is an unmatched worker who has at least two achievable outcomes in the set $\underline{S}_W^{\mu'}(P)$.

Theorem 7.6 showed that there are always equilibria in preference strategies for the DDA game at which firms reveal their true preferences and the output is stable for the true preferences. The following theorem shows that stability with respect to the original preferences holds for any matching that results from a play of equilibrium strategies in which firms reveal their true preferences. This result holds even though workers will not in general reveal their true preferences, and even when they employ non-preference strategies. This theorem generalizes a known result for the CDA game with initial empty matching; see [27] or [32, Theorem 4.16]. Of course, in the CDA game any equilibrium defines a single outcome, but in the DDA game the output may depend on a particular play (see Example 1).

THEOREM 7.7. *Let $\mu' \in Q(P)$ and let $\sigma = (P_F, \sigma'_W)$ be an* RI *equilibrium strategy profile of the* DDA *game with initial matching μ'. Suppose the play under some realization of this* RI *equilibrium results in the matching μ. Then $\mu \in \underline{S}_W^{\mu'}(P)$.*

Proof. Since μ' is acceptable under P, each agent can avoid an unacceptable outcome by playing the strategy associated with its/her true preferences. Also, a worker w can avoid outcomes she likes less than $\mu'(w)$ by playing the true preferences. So the assumption that σ is an RI equilibrium implies that μ is acceptable under P and $\mu \geqslant_W^P \mu'$. Suppose now that, under P, μ is blocked by a pair (f, w), i.e., $f >_w^P \mu(w)$ and $w >_f^P \mu(f)$.

In particular, $\mu'(w) \leqslant_w^P \mu(w) <_w^P f$. Now, if f is matched throughout the particular play of the game, then $\mu'(f) = \mu(f) \in W$ and $w >_f^P \mu(f) = \mu'(f)$. So, the pair (f, w) blocks μ' under P and $\mu'(f) \in W$, in contradiction to the firm-quasi-stability of μ' under P. Thus, at some stage of the game f was unmatched. The rules of the DA Algorithm and the fact that $\sigma_f = P_f$ then imply that w receives an offer from f at some point. But in this case consider the strategy profile $\sigma' \equiv (P_F, \sigma_{W \setminus \{w\}}, \sigma'_w)$ where σ'_w is the same as σ_w at every information set at which no offer has been received from f but differs from σ_w in that no offer from f is ever rejected and f is never abandoned. Then all the iterations of the given play of the game (with σ) up to the point where f makes an offer to w may be replayed in a play of the game with the strategy profile σ', at which point f makes an offer to w, and she ends the game matched to f. As $f >_w^P \mu(w)$, this contradicts the assumption that σ is an RI equilibrium. So, no pair blocks μ under P and hence μ is stable under P. As it was also shown that $\mu \geqslant_W \mu'$, we have that $\mu \in \underline{S}_W^{\mu'}(P)$. ▮

Theorems 7.6 and 7.7 identify a natural class of equilibria in which the firms reveal their true preferences and, although the workers need not in general reveal their true preferences, outcomes are nevertheless stable with respect to the true preferences.

8. CONCLUDING REMARKS

Senior level labor markets can be modeled as two-sided matching markets which are destabilized by retirements, and can return to stability by a process of offers and deferred acceptances. We have shown that this is a consistent point of view, as this kind of destabilization leads to firm-quasi-stable matchings, from which stability can again be achieved, both by straightforward and by equilibrium behavior (recall Theorems 4.1, 3.1, and 7.7).[25]

Our development here differs from the standard treatment of two-sided matching models both in that we generalize the model to consider matching beginning from arbitrary firm-quasi-stable matchings (instead of just from the empty matching at which all candidates and positions are available), and that our strategic results are obtained for decentralized

[25] We will argue elsewhere (Roth and Rothblum, in preparation) that when the low information which many market participants have about others' preferences is modeled, equilibrium behavior will often be well approximated by straightforward behavior. In this connection, note that Barbera and Dutta [4] have shown that there is a sense in which straightforward play can be regarded as a kind of generalized maximin, risk averse behavior.

markets. The phenomena we uncover have strong parallels to the standard treatment, with some important differences.

An important parallel with entry level matching is that, starting from an arbitrary firm-quasi-stable matching, the deferred acceptance procedure converges to a stable matching which is independent of the order in which firms make proposals. This result is one of the most similar to what is already known about entry level matching models, but the differences are instructive. The deferred acceptance procedure (with firms proposing), starting from the empty matching, produces the stable matching that is optimal, over the set of all stable matchings, for the firms. Starting from an arbitrary firm-quasi-stable matching, the deferred acceptance procedure produces the firm optimal stable matching over the set of stable matchings at which each worker does at least as well as at the initial matching. That this subset of stable matchings has a firm optimal matching is due to the fact that the initial matching in such a market is firm-quasi-stable. It would not be true for an arbitrary initial matching. Thus the connection between the initial matching (following retirements and new entries) and the stability of the market prior to retirements and entries plays a critical role.

It is this result which (through Theorems 3.6, 4.3, and Lemma 7.1) addresses the question about vacancy chains raised in the introduction. From a modeling perspective, Theorem 5.1suggests that these results may not be sensitive to whether retirements take place seasonally or are strung out over time, and Lemma 7.1 shows that equilibrium phenomena need not be sensitive to the details of agents' utility functions, which suggests that empirical work on decentralized markets may be more straightforward than if this were not the case.

Our strategic results reflect the fact that the decentralized game has more strategies than the centralized game. Hence there are fewer dominant strategies (Theorem 7.1 and Example 3), and hence fewer dominated strategies and potentially more equilibria in undominated strategies than in the centralized game. However the augmentation of the strategy sets does not destroy the stability of the matchings which arise from the most natural set of equilibria, in preference strategies (Theorems 7.6 and 7.7), and, in the opposite direction, Theorem 7.6 shows that the initial matching from which the game begins in the senior level model constrains which stable outcomes can be achieved for matched workers, in contrast to the model of entry level matching.

Let us not end without discussing what may be the chief limitation of the present model, which it shares with models of equilibration in general, namely the implicit assumption that the process of equilibration will continue to completion. In markets in which tendering offers is time consuming, or in which there are very many offers which would need to be made before the market would clear, we might find that many offers expire (and hence must be accepted or rejected) before the deferred acceptance process we consider here fully runs its course.[26] The model itself suggests why firms may have an incentive to make "exploding offers" which must be accepted or rejected before other offers can be considered. In particular, once a firm has tendered an offer to its most preferred candidate, if it could arrange things so as to compel immediate acceptance it would be better off than if the candidate could consider other offers. So in markets whose institutional arrangements make exploding offers practical (e.g., enforcible when accepted) we may also find that the deferred acceptance procedure is cut short. In such situations the present analysis suggests that we look for symptoms of the instability which can be expected to accompany the failure of the market to clear.[27]

In short, the present paper extends the theoretical analysis of two-sided matching models to markets which may be periodically destabilized, e.g., by retirements. Just as the standard model has provided a useful framework for the empirical analysis of entry level labor markets and other matching processes, we hope that the present analysis may provide a framework around which to organize similar studies of senior level labor markets.

APPENDIX

The Lattice Operators and the Set of Firm-Quasi-Stable Matchings

In this Appendix we prove some technical results about the set of firm-quasi-stable matchings that concern the lattice operators defined in Section 2. These results are key for many results of this paper.

Theorem 2.1 shows that the four binary operators $\vee_F$, $\wedge_F$, $\vee_W$ and $\wedge_W$ preserve stability. We next show that this property generalizes to firm-quasi-stability with regard to $\vee_W$, but not with regard to the other three operators.

THEOREM A.1. *Suppose $\mu, \mu' \in Q(P)$. Then $\mu \vee_W \mu' \in Q(P)$.*

Proof. Let $\mu^* \equiv \mu \vee_W \mu'$. Clearly, each worker is matched to at most one firm under μ^*. We next show that the same conclusion holds for the firms. Suppose to the contrary that there exists a firm f and two different

[26] In the context of the entry level labor market for clinical psychologists, Roth and Xing [36] consider just such a phenomenon, and the strategic behavior it induces in the market participants.

[27] Roth and Xing [35] describe several dozen entry level markets and submarkets in which, in one way or another, firms have acted on such incentives.

workers w and w' such that $\mu^*(w) = \mu^*(w') = f$. Without loss of generality assume that $w >_f w'$, $\mu(f) = w$ and $\mu'(f) = w'$. Now, $\mu^*(w) = f = \mu(w) \neq \mu'(w)$ and by definition, $\mu^*(w) \geqslant_w \mu'(w)$. Hence, $f = \mu^*(w) >_w \mu'(w)$ and as $w >_f w' = \mu'(f)$ it follows that (f, w) blocks μ'. Since $\mu'(f) = w' \in W$, we get a contradiction to the firm-quasi-stability of μ'. So, each firm is matched under μ^* to at most one worker and hence μ^* is a matching.

It remains to show that μ^* is firm-quasi-stable. As $\mu^* \subseteq \mu \cup \mu'$, the acceptability of μ and μ' guarantees the acceptability of μ^*. Suppose now that a pair (f, w) blocks μ^* where $\mu^*(f) \in W$. Without loss of generality assume $\mu^*(f) = \mu'(f)$. Then $f >_w \mu^*(w) \geqslant_w \mu'(w)$ and $w >_f \mu^*(f) = \mu'(f)$, which implies that (f, w) is a blocking pair for μ'. As before, since $\mu'(f) = \mu^*(f) \in W$ we get a contradiction to the firm-quasi-stability of μ'. So we proved that μ^* is an acceptable matching and that the firm in every blocking pair for it is unmatched, i.e., μ^* is firm-quasi-stable. ▌

Example 6. Two firm-quasi-stable matchings for which the binary operators $\wedge_W$, $\vee_F$, and $\wedge_F$ define correspondences between firms and workers that are not matchings: Consider a marriage market with 4 firms and 3 workers where

$$\begin{aligned} P_{f_1} &= w_1, w_2 & \quad P_{w_1} &= f_4, f_1 \\ P_{f_2} &= w_3, w_2 & \quad P_{w_2} &= f_1, f_2 \\ P_{f_3} &= w_3 & \quad P_{w_3} &= f_3, f_2 \\ P_{f_4} &= w_1, & & \end{aligned}$$

and consider the two matchings

$$\mu = \{(f_1, w_1), (f_2, w_2), (f_3, w_3)\}$$

and

$$\mu' = \{(f_1, w_2), (f_2, w_3), (f_4, w_1)\}.$$

Then both μ and μ' are firm-quasi-stable, but none of

$$\begin{aligned} \mu \wedge_W \mu' &= \{(f_1, w_1), (f_2, w_2), (f_2, w_3)\} \\ \mu \vee_F \mu' &= \{(f_1, w_1), (f_2, w_3), (f_3, w_3), (f_4, w_1)\} \\ \mu \wedge_F \mu' &= \{(f_1, w_2), (f_2, w_2)\} \end{aligned}$$

is a matching.

Theorems 2.1 and A.1 show that the operator $\vee_W$ maintains both stability and firm-quasi-stability. We next show that stability is achieved even when just one of the two underlying firm-quasi-stable matchings is stable. We shall need the following lemma which generalizes the decomposition lemma of Knuth; see Roth and Sotomayor [32, p. 42].

LEMMA A.2. *Let* $\mu \in S(P)$ *and* $\mu' \in Q(P)$. *Define* $F(\mu, \mu')$ *to be the set of firms* f *such that* $\mu(f) >_f \mu'(f)$ *and* f *does not belong to any blocking pair for* μ', *and define* $W(\mu', \mu)$ *to be the set of workers* w *such that* $\mu'(w) >_w \mu(w)$. *Then both* μ *and* μ' *are isomorphisms between* $F(\mu, \mu')$ *and* $W(\mu', \mu)$.

Proof. Suppose $f \in F(\mu, \mu')$. Then $\mu(f) >_f \mu'(f) \geqslant_f f$, so $\mu(f) \in W$; in particular, $w \equiv \mu(f) >_f \mu'(f)$. Since $f \in F(\mu, \mu')$, the pair (f, w) does not block μ', hence $\mu'(w) >_w f = \mu(w)$. So $w \in W(\mu', \mu)$. We have shown that μ maps $F(\mu, \mu')$ into $W(\mu', \mu) \subseteq W$, and since any matching is a one to one correspondence, it follows that $|F(\mu, \mu')| \leqslant |W(\mu', \mu)|$.

Suppose now $w \in W(\mu', \mu) \subseteq W$. Then $\mu'(w) >_w \mu(w) \geqslant_w w$ and so $\mu'(w) \in F$. In particular, $f \equiv \mu'(w) >_w \mu(w)$. Using the stability of μ we conclude that $\mu'(f) = w <_f \mu(f)$. Further, since $\mu'(f) = w \in W$, the firm-quasi-stability of μ' guarantees that f does not belong to any blocking pair for μ'. Hence $f \in F(\mu, \mu')$ and we have shown that μ' maps $W(\mu', \mu)$ into $F(\mu, \mu')$. Again, since any matching is a one to one correspondence, it follows that $|F(\mu, \mu')| \geqslant |W(\mu', \mu)|$. We conclude that $|F(\mu, \mu')| = |W(\mu', \mu)|$, and the claim of the lemma is obtained from the fact that one to one correspondences between sets of equal finite cardinality must be onto. ▮

THEOREM A.3. *Let* $\mu \in S(P)$ *and* $\mu' \in Q(P)$. *Then* $\mu \vee_W \mu' \in S(P)$.

Proof. Let $F(\mu, \mu')$ and $W(\mu', \mu)$ be the sets defined in Lemma A.2, and set $\mu^* \equiv \mu \vee_W \mu'$. By Theorem A.1, μ^* is firm-quasi-stable, in particular, μ^* is an acceptable matching. To show that μ^* has no blocking pair, suppose $f \in F$, and consider three cases:

(i) $f \in F(\mu, \mu')$: By Lemma A.2, $w' \equiv \mu'(f) \in W(\mu', \mu) \subseteq W$. Hence $\mu'(w') >_{w'} \mu(w')$; therefore, $\mu^*(w') = \mu'(w') = f$ and $\mu^*(f) = w' \in W$. The firm-quasi-stability of μ^* now implies that f does not belong to any blocking pair for μ^*.

(ii) $\mu(f) = f$ and $f \notin F(\mu, \mu')$: The stability of μ implies that for all $w \in A_f(P)$, $\mu(w) >_w f$, and by definition of μ^*, $\mu^*(w) \geqslant_w \mu(w) >_w f$. In particular, no pair that contains f blocks μ^*.

(iii) $\mu(f) \in W$ and $f \notin F(\mu, \mu')$: By Lemma A.2, for $w \equiv \mu(f)$, $f = \mu(w) \geqslant_w \mu'(w)$ (since otherwise, $w \in W(\mu', \mu)$ and $f \in F(\mu, \mu')$). So, $\mu^*(w) = \mu(w) = f$ and $\mu^*(f) = w \in W$. The firm-quasi-stability of μ^*, again, implies that f does not belong to any blocking pair for μ^*.

We proved that μ^* is an acceptable matching and that there is no blocking pair for it, so it follows that μ^* is a stable matching. ▮

The following corollary, due to Sotomayor [38], asserts that $\mu_W(P)$, the worker-optimal stable matching, is also worker-optimal within the set of firm-quasi-stable matchings.

COROLLARY A.4. *Let $\mu' \in Q(P)$. Then $\mu_W(P) \geqslant_W \mu'$.*

Proof. By Theorem A.3, $\mu^* \equiv \mu' \vee_W \mu_W(P) \in S(P)$ and by the optimality of $\mu_W(P)$ within $S(P)$ it follows that $\mu_W(P) \geqslant_W \mu^*$. Hence $\mu_W(P) \geqslant_W \mu^* \geqslant_W \mu'$. ∎

The next example demonstrates that $\mu_F(P)$, the firm-optimal stable matching, need not be firm-optimal in the set of firm-quasi-stable matchings.

EXAMPLE 7. A marriage market with a firm-quasi-stable matching that is preferred by some firm to the firm-optimal stable matching.

Consider any marriage market such that F contains a firm f with $A_f(P) \neq \emptyset$ and the f-most-preferred worker is not achievable for it. The matching where f is matched to its favorite worker, and all other agents are unmatched is firm-quasi-stable, and f prefers this matching to $\mu_F(P)$.

The next corollary of Theorem A.3 will be very useful in our development.

COROLLARY A.5. *Let $\mu' \in Q(P)$ and $w \in W$. Then either $\mu'(w)$ is achievable for w or $\mu'(w) <_w [\mu_F(P)](w)$.*

Proof. Assume $\mu'(w)$ is not achievable for w and let $\mu^* \equiv \mu' \vee_W \mu_F(P)$. By Theorem A.3, $\mu^* \in S(P)$; in particular, $\mu^*(w)$ is achievable for w. So, as $\mu^*(w) \in \{\mu'(w), [\mu_F(P)](w)\}$ and $\mu'(w)$ is not achievable for w, $[\mu_F(P)](w) = \mu^*(w) >_w \mu'(w)$. ∎

Given an acceptable matching μ', define

$$\underline{S}_W^{\mu'}(P) \equiv \{\mu \in S(P) \colon \mu \geqslant_W \mu'\},$$

that is, $\underline{S}_W^{\mu'}(P)$ is the set of stable matchings that are worker-superior to μ'.

THEOREM A.6. Let $\mu' \in Q(P)$. Then $\underline{S}_W^{\mu'}(P)$ is a non-empty sub-lattice of $S(P)$.

Proof. Corollary A.4 shows that the worker-optimal stable matching is always in $\underline{S}_W^{\mu'}(P)$; hence, $\underline{S}_W^{\mu'}(P) \neq \emptyset$. Further, if $\mu^1, \mu^2 \in \underline{S}_W^{\mu'}(P)$ then $\mu^1, \mu^2 \geqslant_W \mu'$ and by Theorem 2.1 $\mu^1 \vee_W \mu^2, \mu^1 \wedge_W \mu^2 \in S(P)$. By the worker-wise definition of $\vee_W$ and $\wedge_W$ it follows that $\mu^1 \vee_W \mu^2 \geqslant_W \mu'$ and

$\mu^1 \wedge_W \mu^2 \geqslant_W \mu'$. So, $\mu^1 \vee_W \mu^2$ and $\mu^1 \wedge_W \mu^2$ are in $\underline{S}_W^{\mu'}(P)$; in particular, $\underline{S}_W^{\mu'}(P)$ with lattice operators $\vee_W$ and $\wedge_W$ is a sub-lattice of $S(P)$. ∎

REFERENCES

1. A. Abbott, Vacancy models for historical data, *in* "Social Mobility and Social Structure," (R. L. Brieger, Ed.), Cambridge University Press, 1990, 80–102.
2. H. Abeledo and U. G. Rothblum, Courtship and linear programming, *Linear Algebra Its Appl.* 1993, to appear.
3. H. Abeledo and U. G. Rothblum, Paths to marriage stability, unpublished manuscript, 1992.
4. S. Barbera and B. Dutta, Protective behaviour in matching Modelsm, mimeo, Universitat Autonoma de Barcelona, 1991.
5. E. Bennett, Entry, divorce and re-marriage in matching markets, mimeo, 1991.
6. T. Bergstrom and M. Bagnoli, Courtship as a waiting game, *J. Polit. Econ.* **101** (1993), 185–202.
7. Y. Blum and U. G. Rothblum, Timing is everything' and marital bliss, mimeo, 1994.
8. V. P. Crawford, Comparative statics in matching markets, *J. Econ. Theory* **54** (1991), 389–400.
9. V. P. Crawford and E. M. Knoer, Job matching with heterogeneous firms and workers, *Econometrica* **49** (1981), 437–450.
10. G. Demange, D. Gale, and M. O. A. Sotomayor, A further note on the stable matching problem, *Discrete Appl. Math.* **16** (1987), 217–222.
11. L. E. Dubins and D. A. Freedman, Machiavelli and the Gale-Shapley algorithm, *Amer. Math. Monthly* **88** (1981), 485–494.
12. D. Gale and L. S. Shapley, College admissions and the stability of marriages, *Amer. Math. Monthly* **69** (1962), 9–15.
13. D. Gale and M. O. A. Sotomayor, Some remarks on the stable marriage problem, *Discrete Appl. Math.* **11** (1985), 223–232.
14. R. Gibbons and L. Katz, Layoffs and lemons, *J. Labor Econ.* **9** (1991), 351–80.
15. R. Gibbons and K. J. Murphy, Optimal incentive contracts in the presence of career concerns: Theory and evidence, *J. Polit. Econ.* **100** (1992), 468–505.
16. M. Granovetter, "Getting a Job," Harvard University Press, Cambridge, 1974.
17. M. Granovetter, "Afterword: Reconsiderations and a New Agenda, Getting a Job," Second Edition, University of Chicago Press, Chicago, 1995.
18. B. Holmstrom and P. Milgrom, The firm as an incentive system, *Amer. Econ. Rev.* **84** (Sept. 1994), 972–991.
19. A. S. Kelso, Jr., and V. P. Crawford, Job matching, coalition formation, and gross substitutes, *Econometrica* **50** (1982), 1483–1504.
20. D. E. Knuth, "Mariages Stables," Les Presses de l'Universite Montreal, Montreal, 1976.
21. D. McVitie and L. B. Wilson, Stable marriage assignment for unequal sets, *BIT* **10** (1970), 295–309.
22. J-p. Mo, Entry and structures of interest groups in assignment games, *J. Econ. Theory* **46** (1988), 66–96.
23. S. Mongell and A. E. Roth, Sorority rush as a two-sided matching mechanism, *Amer. Econ. Rev.* **81** (1991), 441–464.
24. R. A. Pollak, For better or worse: The roles of power in models of distribution within marriage, *Amer. Econ. Rev. Papers and Proceedings* **84** (May 1994), 148–152.

25. A. E. Roth, The economics of matching: Stability and incentives, *Math. of Oper. Res.* **7** (1982), 617–628.
26. A. E. Roth, The evolution of the labor market for medical interns and residents: A case study in game theory, *J. Polit. Econ.* **92** (1984), 991–1016.
27. A. E. Roth, Misrepresentation and stability in the marriage model, *J. Econ. Theory* **36** (1984), 383–387.
28. A. E. Roth, On the allocation of residents to rural hospitals: A general property of two-sided matching markets, *Econometrica* **54** (1986), 425–427.
29. A. E. Roth, New physicians: A natural experiment in market organization, *Science* **250** (1990), 1524–1528.
30. A. E. Roth, A natural experiment in the organization of entry level labor markets: Regional markets for new physicians and surgeons in the U.K., *Amer. Econ. Rev.* **81** (1991), 415–440.
31. A. E. Roth and U. G. Rothblum, The information requirements of strategic behavior in labor markets and other matching processes, in preparation.
32. A. E. Roth and M. O. A. Sotomayor, "Two Sided Matching: A Study in Game-Theoretic Modeling and Analysis," Econometric Society Monograph Series, Cambridge Univ. Press, Cambridge, UK, 1990.
33. A. E. Roth and J. H. Vande Vate, Random paths to stability in two sided matching, *Econometrica* **58** (1990), 1475–1480.
34. A. E. Roth and J. H. Vande Vate, Incentives in two-sided matching with random stable mechanisms, *Econ. Theory* **1** (1991), 31–44.
35. A. E. Roth and X. Xing, Jumping the gun: Imperfections and institutions related to the timing of market transactions, *Amer. Econ. Rev.* **84** (1994), 992–1044.
36. A. E. Roth and X. Xing, Turnaround time and bottlenecks in market clearing: Decentralized matching in the market for clinical psychologists, *J. Polit. Econ.* **105** (1997), 284–329.
37. D. R. Smith and A. Abbott, A labor market perspective on the mobility of college football coaches, *Social Forces* **61** (1983), 1147–1167.
38. M. O. A. Sotomayor, A non-constructive elementary proof of the existence of stable marriages, *Games Econ. Behav.* **13** (1996), 135–137.
39. H. C. White, "Chains of Opportunity: System Models of Mobility in Organizations," Harvard Univ. Press, Cambridge, 1970.