Incentives in Random Matching Markets

Joana Pais

A Thesis Supervised by:
Jordi Massó

Submitted to
The Department of Economics
at
Universitat Autònoma de Barcelona

in Partial Fulfillment of the Requirements for the Degree of
Doctor in Economics

May, 2005

Acknowledgements

My first words are dedicated to my mentor, Jordi Massó. As an advisor, Jordi proved to have all the qualities one can wish for: boundless enthusiasm for research, endless patience to read and comment on my work, and always being able to find time to guide me in my research. I am deeply indebted to him for his incessant support and motivation.

I extend my sincere gratitude and appreciation to all the members and academic visitors of the Departament d'Economia i d'Història Econòmica at Universitat Autònoma de Barcelona for precious suggestions that made this thesis take its current shape. I would like to thank Guillaume Haeringer, Bettina Klaus, Francesc Obiols, and Nuria Rodriguez. Especially, I am grateful to Antoni Calvó and Flip Klijn, for their valuable insights and observations. Gabrielle Demange, Antonio Romero-Medina, and Alvin Roth were also kind enough to offer their comments.

In the last four years I had the opportunity to know a lot of people whom I regard with affection. I will never forget Helena and Marc, Agi, Annia, Aida, Coralio, Danilo, Helmut, and all the students from my year, with whom I shared periods of hard-studying but also joyful moments. I just hope that I will be able to enjoy their company in the future. Finally, I thank Isabel for being the best friend one can have.

I gratefully acknowledge the financial support from the Fundação para a Ciência e Tecnologia under grant SFRH/BD/5308/2001 and from the research project BEC 2002-02130 financed by the Ministerio de Ciencia y Tecnología.

Contents

Chapter 1
Introduction

There is now a vast literature on two-sided matching markets. The term *two-sided* refers to the fact that agents belong to one of two disjoint sets and can never interchange roles. Each agent has ordinal preferences over the other side of the market and the prospect of being unmatched and the *matching* problem reduces to the process that brings the members of these two sets together.

Matching is a pervasive phenomenon arising in several economic and social settings. The admission of students to colleges, the assignment of civil servants to civil service positions, entry-level labor markets—as the widely explored market for graduating physicians—, or even the assignment of kidneys to patients needing a kidney transplant are among the matching situations that have gained attention in the last four decades. The working of centralized and deterministic matching procedures, along with strategic issues that confront individuals in these contexts, have been scrutinized. Such matching markets typically work by having each agent submit a rank ordered preference list of acceptable partners to a central clearinghouse, which then produces a matching by processing all the preference lists according to an algorithm. Usually, this algorithm delivers a unique outcome for each preference profile and is, therefore, deterministic. Apart from a few notable exceptions, the study of random mechanisms, which assign to each preference profile a probability distribution on the set of matchings, has been neglected.

The importance of allowing for random matching in markets that are centralized lies in equity considerations. In fact, in discrete problems, any deterministic mechanism is bound to favor a subset of the agents involved. This problem is even more pertinent in two-sided matching models where the opposition of interests between the two sides of the market is particularly acute.

On the other hand, in most labor markets no central planning authority exists to assign workers to firms. The process of finding the best-suited worker for a job or the ideal firm to work for is in the huge majority of cases organized along decentralized lines. Firms and workers place advertisements, go to employment agencies, mobilize local networks, and read newspapers when looking for the perfect partner; then, offers are tendered using the telephone, by mail, or through the Internet. Decentralized decision making in such complex environments may introduce some randomness into which assignments are made: which workers fill which positions may depend on the order in which proposals are issued. The exact set of rules that governs a centralized market, making it particularly amenable to analysis, is no longer present when matching is organized in a decentralized way.

Moreover, strategic issues mount in complexity: in a decentralized market agents do not merely choose which list of preferences to submit; instead, they can decide, after each interview or telephone call what to do next. The size of the strategy space is thus extremely large and has precluded analysis by means of standard matching tools.

The purpose of this thesis is to explore the functioning of labor markets where workers are assigned to firms by means of random processes, either centralized or decentralized, using the simple matching tools. It is divided into four chapters that should be regarded as self-contained papers. In what follows, I will consider each paper separately and briefly describe the main results.

1.1 On Random Matching Markets: Properties and Equilibria

In Chapter 2, the starting point of the analysis is an algorithm that starts with any matching situation and proceeds by creating, at each step, a provisional matching. At each moment in time, a firm is randomly chosen and the best worker on its list of preferences is considered. If this worker is already holding a firm he prefers, the matching goes unchanged. Otherwise, they are (temporarily) matched, pending the possible draw of even better firms willing to match this worker. Some features of this algorithm are explored; namely, it encompasses other algorithms in the literature, as Gale-Shapley's famous deferred-acceptance algorithm. I then analyze the incentives facing agents in the revelation game induced by the proposed algorithm. The random order in which firms are selected when the algorithm is run introduces some uncertainty in the output reached. Since agents' preferences are ordinal in nature, I use ordinal Nash equilibria, based on first-order stochastic dominance. This guarantees that in equilibrium each agent plays his best response to the others' strategies for every utility representation of the preferences.

1.2 Incentives in Decentralized Random Matching Markets

I take a further step further in Chapter 3 by considering a sequential game that represents the functioning of a decentralized labor market. The original feature is that available strategies exhaust all possible forms of behavior: agents act in what they perceive to be their own best interest throughout the game, not necessarily according to a list of possible matches. The game starts with a move by Nature that determines the order of play, reflecting the inherently uncertain features of a decentralized market. Then, firms are selected according to the drawn order and given the opportunity to offer their positions. In order to account for the dynamic nature of the game, I characterize subgame perfect ordinal Nash equilibria. In the main result of the paper, I show that every play of such an equilibrium where firms best reply by acting according to their true preferences leads to a stable matching. And stable matchings are those that we expect to see in practice: if a matching is unstable, there is an agent or a pair of agents (consisting of a firm and a worker) with incentives to circumvent the matching. This provides an explanation for the success of some decentralized labor markets.

1.3 Random Stable Mechanisms in the College Admissions Problem

Following a different approach, in Chapter 4, I consider the game where agents from the two sides of the market meet bilaterally in a random fashion and, upon meeting, match if this is consistent with their strategies, and separate otherwise. Strategies are lists but, contrary to Chapter 2, all matchings achieved with positive probability are stable for the revealed preferences. In this paper, I characterize ordinal Nash equilibria, providing simultaneously some results that extend to deterministic mechanisms. In particular, a matching can be obtained as the outcome of a play of the game where firms reveal their true preferences if and only if it is stable with respect to the true preferences. In closing, I relate equilibrium strategy profiles in the games induced by both random and deterministic stable mechanisms: for any random stable mechanism that always assigns positive probability to two different matchings, I show that a strategy profile is an ordinal Nash equilibrium if and only if it has a unique stable matching and there exists a deterministic stable mechanism where it is a Nash equilibrium.

1.4 Giving Advice and Perfect Equilibria in Matching Markets

It is well-known that in the game induced by a stable mechanism, every individually rational matching can be sustained in equilibrium. In this last Chapter I attempt to provide a better prediction of the outcomes of such games by imposing additional rationality requirements. Hence, Chapter 5 is a preliminary analysis of ordinal perfect equilibria in matching markets. I show that, in the game induced by a random stable mechanism, an ordinal perfect equilibrium strategy is exhaustive, listing all the acceptable partners. It follows that some individually rational matchings cannot be sustained in an ordinal perfect equilibrium. When either the firm-optimal or the worker-optimal mechanisms are considered, truth telling is the unique ordinal perfect equilibrium that may emerge. It is thus apparent that ordinal perfect equilibria rarely exist; in fact, truth telling is an ordinal perfect equilibrium if and only if it is a Nash equilibrium in dominant strategies.

From a different point of view, the results in this paper allow for advising agents about how to participate in these markets. I show that agents who are poorly informed and aim at minimizing the probability of being unmatched, should list all the acceptable partners in the game induced by a random stable mechanism. In the game induced by the firm-optimal or the worker-optimal stable mechanisms, I go farther to suggest the honest revelation of one's preferences as a sensible form of behavior.

Chapter 2
On Random Matching Markets: Properties and Equilibria

2.1 Introduction

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

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

In this paper we consider an extension to Gale and Shapley's algorithm or, to be precise, to the version proposed by McVitie and Wilson (1970). We start from an arbitrary matching and the algorithm proceeds by creating, at each step, a provisional matching. Hence, at each moment in time, a firm is randomly chosen and the best worker on its list of preferences is considered. If this worker is already holding a firm he prefers, the matching goes unchanged and this particular worker is removed from the firm's list. Otherwise, they are (temporarily) matched, pending the possible draw of even better firms willing to match this worker. McVitie and Wilson's algorithm is an instance of the one we are proposing, when the initial matching is the empty matching. Moreover, it also encompasses the algorithm pro-

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

posed by Blum, Roth, and Rothblum (1997) to explore the vacancy chain problem when the input matching is firm-quasi-stable (*i.e.*, a matching whose stability was disrupted by the emergence of a new position or the retirement of a worker).

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

We proceed as follows. In Section 2.2 we present the simple marriage model and introduce notation. We formally describe the algorithm in Section 2.3, showing that it captures other algorithms. In addition, some of its features are explored. In Section 2.4 we turn our attention to a different class of questions, related to individual decision making. The matching process is modeled as a game and its equilibria are characterized. Some concluding remarks follow in Section 2.5.

2.2 The Marriage Model

Consider two finite and disjoint sets $F = \{f_1, ..., f_n\}$ and $W = \{w_1, ..., w_p\}$, where F is the set of firms and W is the set of workers. Let $V = F \cup W$. Sometimes we refer to a generic agent by v and we use f and w to represent a generic firm and worker, respectively. Each agent has a strict, complete, and transitive preference relation over the agents on the other side of the market and remaining unmatched. The preferences of a firm f, for example, can be represented by $P_f = w_3, w_1, f, w_2, ..., w_4$, indicating that f's first choice is to be matched to w_3, its second choice is w_1 and it prefers remaining unmatched to being assigned

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

to any other worker. Sometimes it is sufficient to describe only f's ranking of workers it prefers to remaining unmatched, so that the above preferences can be abbreviated as $P_f = w_3, w_1$. Let $P = (P_{f_1}, ..., P_{f_n}, P_{w_1}, ..., P_{w_p})$ denote the profile of all agents' preferences; we sometimes write it as $P = (P_v, P_{-v})$ where P_{-v} is the set of preferences of all agents other than v. Further, we may use P_U, where $U \subseteq V$, to denote the profile of preferences $(P_v)_{v \in U}$. We write $v' P_v v''$ when v' is preferred to v'' under preferences P_v and we say that v *prefers* v' to v''. We write $v' R_v v''$, when v likes v' *at least as well as* v'' (it may be the case that v' and v'' are the same agent). Formally, a *marriage market* is a triple (F, W, P). Let $A(P_f)$ denote the set of workers that are *acceptable* to firm f, *i.e.*, $A(P_f) = \{w \in W : w P_f f\}$; $A(P_w)$ is defined analogously. A pair $(f, w) \in F \times W$ is acceptable if f and w are acceptable to each other.

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

2.3 The Algorithm

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

Gale and Shapley (1962) showed that a stable matching exists for every marriage market. Their proof is in fact an algorithm for finding such a matching. Starting from a situation where no agent is matched, in the "deferred-acceptance" algorithm (DA-algorithm), firms propose to workers who can hold at most one unrejected offer at any time. At any step of the algorithm, every rejected firm proposes to its next choice, as long as there are acceptable workers on its list to whom it has not made an offer yet. The algorithm stops after the step in which every rejected firm has proposed to all of its acceptable workers.

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

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

2.3.1 Definition of the DA$^{\mu^I}$-Algorithm

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

Definition 2.1 *The Generalized Deferred-Acceptance Algorithm (DA$^{\mu^I}$-algorithm):*

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

1. (a) For all $f \in F,\ A_f^0 = A(P_f) \cup \{f\}$;
   (b) $\mu^0 = \mu^I;\ i := 1$;
2. If, for all $f \in F$, $\mu^{i-1}(f) = \max_{P_f} A_f^{i-1}$, then stop with μ^{i-1}.
3. Else, take any firm f such that:
   (a) either $\max_{P_f} A_f^{i-1} = f$ and $\mu^{i-1}(f) \neq f$, leading to $\mu^i = \mu^{i-1} \setminus \{(f, \mu^{i-1}(f))\}$;
   (b) or $\max_{P_f} A_f^{i-1} = w$ and $\mu^{i-1}(f) \neq w$, in which case:
      I. if $\mu^{i-1}(w) P_w f$, then $\mu^i = \mu^{i-1}$ and $A_f^i = A_f^{i-1} \setminus \{w\}$, $A_{f'}^i = A_{f'}^{i-1}$, for all $f' \neq f$;
      II. else:
         (i) if $\mu^{i-1}(f) = f$ and $\mu^{i-1}(w) = w$, then $\mu^i = \mu^{i-1} \cup \{(f, w)\}$ and $A_{f'}^i = A_{f'}^{i-1}$, for all $f' \in F$;
         (ii) if $\mu^{i-1}(f) \neq f$ and $\mu^{i-1}(w) = w$, then $\mu^i = (\mu^{i-1} \cup \{(f, w)\}) \setminus \{(f, \mu^{i-1}(f))\}$ and $A_{f'}^i = A_{f'}^{i-1}$, for all $f' \in F$;
         (iii) if $\mu^{i-1}(f) = f$ and $\mu^{i-1}(w) \neq w$, then $\mu^i = (\mu^{i-1} \cup \{(f, w)\}) \setminus \{(\mu^{i-1}(w), w)\}$ and $A_{\mu^{i-1}(w)}^i = A_{\mu^{i-1}(w)}^{i-1} \setminus \{w\}$, $A_{f'}^i = A_{f'}^{i-1}$, for all $f' \neq \mu^{i-1}(w)$;
         (iv) if $\mu^{i-1}(f) \neq f$ and $\mu^{i-1}(w) \neq w$, then $\mu^i = (\mu^{i-1} \cup \{(f, w)\}) \setminus \{(f, \mu^{i-1}(f)), (\mu^{i-1}(w), w)\}$ and $A_{\mu^{i-1}(w)}^i = A_{\mu^{i-1}(w)}^{i-1} \setminus \{w\}$, $A_{f'}^i = A_{f'}^{i-1}$, for all $f' \neq \mu^{i-1}(w)$;
4. $i := i + 1$; go to 2.

2.3.2 Properties of the DA$^{\mu^I}$-Algorithm

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

Example 2.1 *The outcome depends on the selection of the order by which firms propose.*

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

$$\begin{array}{ll} P_{w_1} = f_2,\ f_1 & P_{f_1} = w_1,\ w_2 \\ P_{w_2} = f_1,\ f_2 & P_{f_2} = w_2,\ w_1. \end{array}$$

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

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

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

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

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

Although a random element appears each time a firm is chosen, all the uncertainty is fully translated into a probability distribution over the set of matchings. For each input matching and for each profile of preferences, a lottery over matchings is obtained. Hence, fix a probability distribution on Σ and take an initial matching μ^I, a preference profile P, and an arbitrary worker w. We will let $\widetilde{DA}^{\mu^I}[P]$ denote the probability distribution over the set of matchings induced by the DA$^{\mu^I}$-algorithm and $\widetilde{DA}^{\mu^I}[P](w)$ be the distribution that $\widetilde{DA}^{\mu^I}[P]$ induces over $F \cup \{w\}$. The expression $\Pr\{\widetilde{DA}^{\mu^I}[P] = \mu\}$ represents the probability that μ is the output of the DA$^{\mu^I}$-algorithm with preferences P. Observe that this probability rests on the probability distribution on Σ, but all results hold regardless of this lottery. Finally, for all $w \in W$, $v \in F \cup \{w\}$, the subset of all possible orders leading to an output matching where w is assigned to v is denoted by $\Sigma_{v,w}$.

In the particular case that the input matching is the empty matching, $\emptyset$, a degenerate probability distribution over the set of matchings is obtained. In fact, it turns out that, when $\mu^I = \emptyset$, the DA$^{\mu^I}$-algorithm specializes to McVitie and Wilson's algorithm and the firm-optimal stable matching is obtained with probability one. For illustration, consider the matching market in Example 2.1 and assume the algorithm starts with the empty matching. If f_1 is the first firm to propose, it invites w_1 and w_1 accepts this proposal. Then, f_2 follows and proposes to w_2, who also accepts. If we reverse the order of events and f_2 is the first to move, w_2 accepts its proposal, given that he is currently unmatched; f_1 invites the best worker on its list, w_1, who also accepts. Thus, we always reach μ_F for every order of proposals.

Proposition 2.1 *For any matching market* (F, W, P)*,* $\Pr\{\widetilde{DA}^{\emptyset}[P] = \mu_F\} = 1$.

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

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

Another case worth describing is when the input matching is a firm-quasi-stable matching, as defined by Sotomayor (1996) and Blum, Roth, and Rothblum (1997). In Proposition 2.2 we show that when the initial matching is firm-quasi-stable, the same stable output matching is obtained, independently of the order in which firms propose.

Remark 2.1 turns out to be crucial in what follows.

Remark 2.1 *The DA$^{\mu^I}$-algorithm implies that once a firm proposes to a worker and he accepts, this firm cannot fire him nor exchange him for another worker. In fact, when the proposal is made, the firm reveals that this particular worker is the best among all who have not rejected it. If the worker accepts, the only occasion under which the firm can make a proposal again is when the worker it holds accepts an offer from a different firm.*

Proposition 2.2 *Let (F, W, P) be a matching market. For all $\mu^I \in QS(P)$, there is some $\mu \in S(P)$ such that* $\Pr\{\widetilde{DA}^{\mu^I}[P] = \mu\} = 1$.

Proof. Take $\mu^I \in QS(P)$. For every order of proposals, the first firm to have its offer accepted must be unmatched at μ^I. In fact, by definition of firm-quasi-stability, if (f, w) blocks μ^I, f must be unmatched at μ^I. Assume hence that f proposes to w and that this proposal is the first to be accepted. It follows that, after this acceptance, w is strictly better off and every other worker is holding its initial partner. The rest of the proof now follows using an induction argument.

Suppose that up to step i in the algorithm only firms with vacancies have had their proposals accepted. Let μ^i be the matching at the beginning of step $i+1$. Assume that all workers are weakly better off at μ^i than at the initial matching μ^I. We will show, by way of contradiction, that the next firm to be accepted by some worker must be unmatched. So assume that f is matched to w at μ^i, it proposes to w' and this proposal is accepted. Thus, at μ^{i+1}, f and w' are matched to each other and their former partners are unmatched. By Remark 2.1, if f is matched to w at μ^i and it is willing to propose to another worker, it must be the case that $\mu^I(f) = w$. Now, by assumption, $\mu^i(w') R_{w'} \mu^I(w')$. Since $f P_{w'} \mu^i(w')$, we have $f P_{w'} \mu^I(w')$. Further, $w' P_f \mu^I(f) = w$. Thus, (f, w') form a blocking pair to μ^I and $\mu^I(f) \neq f$, contradicting the fact that $\mu^I \in QS(P)$.

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

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

Example 2.2 *An output matching may not be stable.*

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

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

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

An execution of the DA$^{\mu^I}$-algorithm with arbitrary input matching need not be stable. Further, any worker involved in instability of the output matching μ

must have been matched under the input matching.[3] And, if some firm is part of a blocking pair for μ, it must have been rejected by the worker with whom it forms a blocking pair for μ along the execution of the algorithm.[4]

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

Proposition 2.3 *Let* (F, W, P) *be a marriage market and* $\mu^I \in IR(P)$. *Let* $\mu \neq \mu^I$ *be such that, for all* $w \in W$, $\mu^I(w) R_w \mu(w)$. *Then,* $\Pr\{\widetilde{DA}^{\mu^I}[P] = \mu\} = 0$.

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

No unmatched worker would accept to fill a position in an unacceptable firm. Therefore, a worker who is strictly worse off in the output matching μ must have started matched. Moreover, he must have been fired by his initial partner. So, assume w_1 is the first worker to be fired by $\mu^I(w_1)$. This implies that either $\mu^I(w_1)$ fired w_1 to be alone or it proposed to another worker, say w_2, and he accepted. In the former case the individual rationality of μ^I is contradicted. In the latter case, since by assumption w_2 is still holding $\mu^I(w_2)$, we must have $\mu^I(w_1) P_{w_2} \mu^I(w_2)$. By Remark 2.1, w_2 will never end up worse off in the output matching, contradicting the definition of μ. ■

A slightly weaker result holds for the firms. An output matching where *every* firm is matched to a worker ranked lower than its initial partner in its preference list cannot be reached with positive probability. Example 2.3 shows that the requirement of having every firm strictly worse off in the output matching is necessary. Subsequently, we state the result.

---

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

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

Example 2.3

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

$$\begin{array}{ll} P_{w_1} = f_1,\ f_2 & P_{f_1} = w_2,\ w_1 \\ P_{w_2} = f_2,\ f_3,\ f_1 & P_{f_2} = w_1,\ w_2 \\ & P_{f_3} = w_2, \end{array}$$

and let the input matching be $\mu^I = \{(f_1,\ w_2),\ (f_2,\ w_1)\}$. Every execution of the algorithm leads to the matching $\mu = \{(f_1,\ w_1),\ (f_2,\ w_2)\}$. In fact, for every order in which firms propose, when f_3 is given the opportunity to act, it makes a successful offer to w_2, who is still holding f_1 at that point. Later, f_1 is forced to propose to w_1 and f_2 ends up matched to w_2. Hence, $\mu \neq \mu^I$ such that $\mu^I(f) R_f \mu(f)$ for every $f \in F$ is reached with probability one. ◇

Proposition 2.4 *Let (F, W, P) be a marriage market, and let μ^I be an arbitrary input matching. Let μ be such that $\mu^I(f) P_f \mu(f)$, for all $f \in F$. Then,* $\Pr\{\widetilde{DA}^{\mu^I}[P] = \mu\} = 0.$

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

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

Proof. Notice that every worker who is initially assigned to a firm cannot end up alone in the output matching μ. Assume not and, without loss of generality, let us say w such that $\mu^I(w) \in F$ is unmatched under μ. This implies that $\mu^I(w)$ fired w. In addition, it follows from Remark 2.1, that no firm, including $\mu^I(w)$, proposed to w later on. But if this is so, $\mu^I(w)$ must end up matched to a worker ranked higher than w in its list of preferences. This contradicts the fact that $\mu^I(f) P_f \mu(f)$, for all $f \in F$.

Claim 2 Every firm is matched under μ.

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

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

Proof. This follows from Remark 2.1 and Claim 1.

Consider the last step at which a proposal is made by a firm f and accepted by a worker w. (Note that if no proposal is accepted along the execution, then $\mu = \mu^I$, contradicting the definition of μ.) At the last step of the algorithm, w must be unmatched when he accepts f's proposal. Otherwise, the firm held by w would be unmatched under μ, which contradicts Claim 2.

By Claim 3, w must be matched under μ^I, let us say $\widehat{f} = \mu^I(w)$. Firm f is not w's initial partner, or else $\mu^I(f) = \mu(f)$, contradicting the definition of μ. By Claim 2 and given that we are considering the last step of the algorithm, $\widehat{f}$ is matched at this stage. Given that every firm is worse off under the output matching, it must be the case that $\widehat{f}$ is matched to a worker ranked lower than w in $P_{\widehat{f}}$. As a consequence, $\widehat{f}$ must have proposed to w and this proposal was rejected. By Remark 2.1, this implies that w is matched to a firm preferred to $\widehat{f}$ at this last step of the algorithm and we get another contradiction: w was not alone when he accepted f's proposal. ■

2.4 The Game

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

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

To address strategic questions we need to develop ideas about what constitutes a "best decision" to be taken by an agent. With this purpose in mind, take two probability distributions over the set of matchings, $\widetilde{\mu}$ and $\widetilde{\mu}'$. Without loss of generality, consider $w \in W$ (what follows also holds for a representative firm, with the obvious modifications); $\widetilde{\mu}(w)$ and $\widetilde{\mu}'(w)$ denote the distributions induced over w's set of assignments by $\widetilde{\mu}$ and $\widetilde{\mu}'$, respectively. We say that $\widetilde{\mu}(w)$ first order stochastically P_w-dominates $\widetilde{\mu}'(w)$ if $\Pr\{\widetilde{\mu}(w) R_w v\} \geq \Pr\{\widetilde{\mu}'(w) R_w v\}$, for all

$v \in F \cup \{w\}$. Thus, for all $v \in F \cup \{w\}$, the probability of w being assigned to v or to a strictly preferred agent is higher under $\widetilde{\mu}(w)$ than under $\widetilde{\mu}'(w)$. Now, consider the problem that player w would face if the strategy choices Q_{-w} of the other players were known. In this case, any strategy Q_w by w would determine the probability distribution induced by the mechanism over the set of matchings. Therefore, a particular strategy choice Q_w is preferred if the induced probability distribution over the set of matchings stochastically dominates the one induced by any other alternative strategy.

Definition 2.2 *Given Q_{-w} and the preferences P_w, we say that a strategy Q_w* stochastically P_w-dominates *another strategy $\widehat{Q}_w$ if, for all $v \in F \cup \{w\}$, $\Pr\{\widetilde{DA}^{\mu^I}[Q_w, Q_{-w}]\,(w) R_w v\} \geq \Pr\{\widetilde{DA}^{\mu^I}[\widehat{Q}_w, Q_{-w}](w)\ R_w\ v\}$. In a similar way, given Q_{-f} and the preferences P_f, we define* stochastic P_f-dominance.

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

Definition 2.3 *Given an initial matching μ^I and the preferences P_v, a* dominant strategy *for $v \in V$ is a strategy Q_v that, for every Q_{-v}, stochastically P_v-dominates every alternative strategy $\widehat{Q}_v$.*

In Example 2.1, we have shown that the outcome of the generalized deferred-acceptance algorithm may depend on the random order in which firms' lists are considered. Thus, the study of Nash equilibria in the game induced by the mechanism we have described would require us to consider not merely agents' preferences over riskless outcomes, but also over lotteries. Since agents' preferences are ordinal and no natural utility representation of these orderings exists, we will adopt the following equilibrium notion.

Definition 2.4 *Given an initial matching μ^I and a profile of preferences P, the profile of strategies Q is an* ordinal Nash equilibrium *(ON equilibrium) if, for each player v in V, Q_v stochastically P_v-dominates every alternative strategy $\widehat{Q}_v$, given Q_{-v}.*

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

2.4.1 Strategic Questions

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

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

Example 2.4 *Revealing the true preferences is not a dominant strategy for all firms.*

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

$$\begin{array}{rr} P_{w_1} = f_2 & P_{f_1} = w_2 \\ P_{w_2} = f_3,\ f_1 & P_{f_2} = w_1 \\ P_{w_3} = f_3 & P_{f_3} = w_3,\ w_2. \end{array}$$

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

2.4.2 Ordinal Nash Equilibria

We have observed that faithfully transmitting the true preferences is not necessarily an ordinal Nash equilibrium. Therefore, we need to ask whether ordinal Nash equilibria always exist in the revelation game induced by the DA$^{\mu^I}$-algorithm. Proposition 2.6 will show that they do: when μ^I is individually rational, every element of a non-empty subset of $IR(P)$ can be sustained in equilibrium with probability one.

Definition 2.5 *Let μ^I be an arbitrary matching. We say that μ is* individually rational with respect to μ^I *if $\mu \in IR(P)$ and if, for all $f \in F$, $w' = \mu^I(f) P_f \mu(f)$, implies $\mu(w') \neq w'$.*

We will denote by $IR^{\mu^I}(P)$ the set of all individually rational matchings with respect to μ^I. For illustration, in the particular case that μ^I is the empty matching, the set of all individually rational matchings with respect to this initial matching coincides with the set of individually rational matchings (*i.e.*, $IR^{\varnothing}(P) = IR(P)$).[5] In what follows we show that this set is always non-empty.

Proposition 2.5 *Let μ^I be an individually rational matching for (F, W, P). Then, $S(P)$ is a subset of $IR^{\mu^I}(P)$.*

Proof. Consider $\mu \in S(P)$. We will prove that $\mu \in IR^{\mu^I}(P)$ using a contradiction argument. Assume that $\mu \in S(P)$. By definition of stability, this implies $\mu \in IR(P)$, but assume that there exists a firm f such that $w' = \mu^I(f) P_f \mu(f)$ and $\mu(w') = w'$. Stability of μ implies that $w' P_{w'} f$ and we get a contradiction: μ^I is not individually rational. Therefore, every stable matching is an element of $IR^{\mu^I}(P)$. ■

[5] This holds since if $f P_f \mu(f)$, then $\mu(f) \neq f$.

Proposition 2.6 *Let μ^I be an individually rational matching for (F, W, P) and let $\mu \in IR^{\mu^I}(P)$. Then, there exists an ordinal Nash equilibrium Q in the revelation game induced by the DA$^{\mu^I}$-algorithm that leads to μ. Furthermore,* $\Pr\{\widetilde{DA}^{\mu^I}[Q] = \mu\} = 1$.

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

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

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

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

Theorem 2.1 *Let μ^I be an individually rational input matching for $(F, W, (Q_F, P_W))$. Assume that the strategy profile Q is an ordinal Nash equilibrium in the revelation game induced by the DA$^{\mu^I}$-algorithm. Then, the probability distribution obtained over the set of matchings is such that every element in its support is a member of $S(Q_F, P_W)$.*

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

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

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

and Q_w is not a best reply to Q_{-w}.

We have proved that μ_i is individually rational. Thus, there must exist a blocking pair for μ_i when the preference profile (Q_F, P_W) is considered. Let us say (f, w) blocks μ_i, *i.e.*, $fP_w\mu_i(w)$ and $wQ_f\mu_i(f)$. This implies that f proposed to and was rejected by w in the course of every execution leading to μ_i. By Remark 2.1, either $\mu_i(w)Q_w f$ (case *(i)*) or, if not, w must have rejected f while he was still holding $\mu^I(w)$ and $\mu^I(w)Q_w f$ (case *(ii)*).

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

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

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

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

contradicting that Q is an *ON* equilibrium.

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

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

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

This completes the proof. ■

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

The following result is an important special case of Theorem 2.1.

Corollary 2.1 *Let μ^I be an individually rational input matching for (F, W, P). Assume (P_F, Q_W) is an ordinal Nash equilibrium in the revelation game induced by the DA$^{\mu^I}$-algorithm. Then, the probability distribution obtained over the set of matchings is such that every element in its support is a member of $S(P)$.*

Proof. Immediate from Theorem 2.1 with $Q_F = P_F$. ■

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

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

Definition 2.6 *Let* $\mu \in S(P)$. *Let* μ^I *be an arbitrary matching. We say that* μ *is* stable with respect to μ^I *if, for all* $f \in F$ *such that* $\mu^I(f)P_f\mu(f)$, *we can define a non-empty subset of firms* $\widehat{F}(f) = \{f_1, f_2, ..., f_r\}$, $r \leq n$, *for which the following conditions hold:*

1. $\mu(f_{i+1}) = \mu^I(f_i)$, for all $i = 1, ..., r-1$, and $\mu(f_1) = \mu^I(f_r)$;
2. $\mu(\mu^I(f)) \in \widehat{F}(f)$;
3. $\mu(f_i)P_{f_i}\mu^I(f_i)$, for some $i = 1, ..., r$.

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

Example 2.5 *(Example 2.3 continued)*

In the matching market of Example 2.3, the only stable matching is $\mu = \{(f_1,\ w_1),\ (f_2,\ w_2)\}$. Comparing μ with the initial matching $\mu^I = \{(f_1,\ w_2),$

$(f_2,\ w_1)\}$, it is clear that no firm is strictly better off under μ than under μ^I. Hence, condition 3 is not fulfilled and $S^{\mu^I}(P)$ is empty. $\diamond$

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

Proposition 2.7 *Let μ^I be an individually rational input matching for (F, W, P). Let $\mu \in S^{\mu^I}(P)$. Then, there exists an ordinal Nash equilibrium (P_F, Q_W) in the revelation game induced by the DA$^{\mu^I}$-algorithm that leads to μ. Moreover,* $\Pr\{\widetilde{DA}^{\mu^I}[P_F, Q_W] = \mu\} = 1$.

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

Let us now prove that, for every firm f, P_f stochastically P_f-dominates every other strategy Q_f. We will consider the most general case, assuming that $\mu^I(f)$ and $\mu(f) \in W$ and $\mu^I(f) \neq \mu(f)$ (the proofs for other cases follow easily from this one). Given that the only worker who is willing to accept f is $\mu(f)$, by choosing its strategy appropriately, f can either be alone, hold $\mu(f)$ or, eventually, remain with $\mu^I(f)$ under the output matching. By stability of μ, $\mu(f) P_f f$. If, additionally, $\mu(f) P_f \mu^I(f)$, firm f can do no better than obtaining $\mu(f)$ and truth telling guarantees $\mu(f)$ is assigned to f with probability one. Otherwise, if

$\mu^I(f)P_f\mu(f)$, f is not able to retain $\mu^I(f)$. In fact, given the definition of $S^{\mu^I}(P)$, $\mu^I(f)$ is matched to some firm under μ and obtains a proposal from this firm. Thus, f cannot do better than being assigned to $\mu(f)$ and P_f stochastically P_f-dominates every other strategy Q_f.

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

Proposition 2.7 showed that there are ordinal Nash equilibria at which firms reveal their true preferences and the output is stable for the true preferences. These equilibria involve misrepresentation by the workers. Further, by misstating their preferences "appropriately," workers can compel the best achievable stable matching. However, as the following example shows, the above proposition does not exhaust all ordinal Nash equilibria.

Example 2.6 *(Example 2.3 continued) There may be more ordinal Nash equilibria than those given in Proposition 2.7.*

Recall that in the matching market in Example 2.3, when $\mu^I = \{(f_1,\ w_2),\ (f_2,\ w_1)\}$ is considered, every execution of the algorithm with P leads to $\mu = \{(f_1,\ w_1),\ (f_2,\ w_2)\}$. Under μ, workers obtain the best possible positions and firms cannot improve by deviating. No manipulation will enable f_1 and f_2 to keep the workers they hold under μ^I, given the presence of f_3. As a result, P is an ordinal equilibrium, even though $S^{\mu^I}(P)$ is empty. ◇

2.5 Concluding Remarks

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

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

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

Chapter 3
Incentives in Decentralized Matching Markets

3.1 Introduction

The study of centralized markets has been privileged in the two-sided matching literature. The introduction of centralized matching procedures in markets that experienced certain kinds of failures is partially responsible for such dedication. In fact, a number of markets—for physicians, lawyers, dentists, and osteopaths, among others—have adopted central clearinghouses after periods of uncontrolled unraveling of appointment dates and chaotic recontracting.[6] These markets now work by having each agent of the two sides of the market submit a rank ordered preference list of acceptable matches to the central clearinghouse, which then produces a matching by processing all the preference lists according to an algorithm. Roth (1984a, 1991) showed that the algorithms used in most of the successful clearinghouses roughly follow the lines of Gale and Shapley's deferred acceptance algorithm (Gale and Shapley, 1962). This procedure generates a matching of workers to positions that is stable in terms of the submitted preferences in the sense that no worker and firm that are not matched to each other would prefer to be so matched.[7]

In contrast, decentralized markets have received relatively little attention.[8] The exact set of rules that governs a centralized market, making it particularly amenable to analysis, is no longer present when matching is organized in a decentralized way. Moreover, decentralized markets involve different strategic issues from those of centralized markets. In fact, when a clearinghouse exists, agents must simply decide what preference lists to submit to the matchmaker, after which the match is created. However, in a decentralized market agents do not submit lists; instead, they can decide, after each interview or telephone call what to do next. The size of the strategy space is thus extremely large and has precluded analysis by means of standard matching tools.

The purpose of this paper is to apply the extremely simple marriage model to the study of decentralized labor markets. The starting point of the analysis is any matching situation, providing a framework to the study of both entry-level

---

[6] See Roth and Xing (1994) and Niederle and Roth (2003).

[7] See Roth and Sotomayor (1990) for a comprehensive study of two-sided matching markets.

[8] There are notable exceptions, namely Blum, Roth, and Rothblum (1997), Haeringer and Wooders (2004), Roth and Vande Vate (1991), Roth and Xing (1997), among others.

and senior level markets. The matching process is then modeled as an extensive form game, where firms sequentially offer their positions. Clearly, decentralized decision making in complex environments may introduce randomness in the order in which offers are made. The speed of the mail, the telephone network, or the internal structure of firms making some react faster than others determine the success in establishing communication with the desired workers. Such inherently uncertain features of the market are modeled here as chance moves that determine the order of play. Hence, at each moment in time, any firm—even if already matched—is randomly selected and given the opportunity to offer its position to a worker. This worker may reject the new offer or he may (temporarily) hold it, pending the possible arrival of even better offers. We assume that, once rejected, the firm is not willing to propose to the same worker again, but it may obviously offer its position to a different worker when given the opportunity to act.[9]

In our setting, there is more to a job than just a salary. Hence, we consider that monetary transfers are embodied in agents' preferences and these preferences are ordinal in nature. Furthermore, the random order in which firms are selected introduces some uncertainty in which matchings are achieved. In fact, it may happen that starting with the same initial matching, different plays of the game yield different outcomes for the same strategy profile. It follows that, in order to compare different probability distributions over matchings, we use a solution concept based on first-order stochastic dominance. The notion of ordinal Nash equilibrium guarantees that each agent is an expected utility maximizer for every utility representation of his preferences.[10] We go beyond this concept to account for the dynamic nature of the game and characterize subgame perfect ordinal Nash equilibria. Despite the strength of this concept, we prove the existence of subgame perfect ordinal Nash equilibria and, in particular, equilibria where firms use preference strategies (*i.e.*, strategies that can, up to some point, be identified with a list of preferences). On the other hand, every such equilibrium delivers matchings that are stable with respect to a particular profile of preferences. This has two appealing implications. First, for any equilibrium where firms adhere to preference lists, all outcomes are such that the set of unmatched agents is the same. Second, in the particular case that firms act according to their true preferences, stability with respect to the true preferences is guaranteed in a subgame perfect ordinal Nash equilibrium. This provides an explanation for the success of some decentralized labor markets. In fact, if we expect equilibria where firms act straightforwardly to prevail, only stable matchings are obtained and no individual agent or pair of agents (consisting of a firm and a worker) will have the incentive to circumvent

9 This assumption guarantees that every play of the game ends in a finite number of steps. It does not appear that allowing for any finite number of repeated proposals would materially change the validity of the results that follow.

10 This concept was introduced in d'Aspremont and Peleg (1988); it has been used in the context of voting theory in Majumdar and Sen (2004) and in matching markets in Ehlers and Massó (2003), and Majumdar (2003).

the matching. Moreover, revealing the true preferences can be easily justified. The decisions of a firm do not usually reflect the opinion of a single individual; instead, such actions embody a complex process of assembling the opinions of several individuals. We may conjecture that establishing a list of candidates and using it as guidance is a more plausible form of behavior than deciding, at each moment in time, whom to propose to. In addition, in some settings firms obey objective criteria to admit workers, so that strategic behavior on the firms' side looses its meaning. The (partially) converse statement holds when we start from a situation where all agents are unmatched: every stable matching for the true preferences can be reached as the outcome of an equilibrium play where firms act straightforwardly according to their true preferences.

The paper is organized as follows. In Section 3.2 we introduce the matching model, and review some results on matching markets. We formally present the model in Section 3.3. In Section 3.4 we turn our attention to questions related to individual decision making and characterize equilibria. Some results and underlying assumptions are discussed in Section 3.5. We conclude in Section 3.6. Some proofs can be found in the Appendix.

3.2 The Marriage Model

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

Formally, a *marriage market* is a triple (F, W, P). An outcome for a marriage market, a *matching*, is a function $\mu : V \longrightarrow V$ satisfying the following: *(i)* for each f in F and for each w in W, $\mu(f) = w$ if and only if $\mu(w) = f$; *(ii)* if $\mu(f) \neq f$ then $\mu(f) \in W$; *(iii)* if $\mu(w) \neq w$ then $\mu(w) \in F$. If $\mu(v) = v$, then v is *unmatched* under μ, while if $\mu(w) = f$, we say that f and w are *matched* to one another. A description of a matching is given by $\mu = \{(f_1, w_2), (f_2, w_3)\}$, indicating that f_1 is matched to w_2, f_2 is matched to w_3 and the remaining agents in the market are unmatched. A matching μ is *individually rational* if each agent is acceptable to its partner, *i.e.*, $\mu(v)R_v v$, for all $v \in V$. We denote the set of all individually rational matchings by $IR(P)$. Two agents f and w form a *blocking pair* for μ if they prefer each other to the agents they are actually assigned to under μ, *i.e.*, $fP_w\mu(w)$ and $wP_f\mu(f)$. A matching μ is *stable* if it is individually rational and it is not blocked by any pair of agents. We denote the set of all stable matchings by $S(P)$.

3.3 The Decentralized Job Matching Game

3.3.1 Description of the Game

In this section, we define the Decentralized Game. The game is given by a market (F, W, P) and an initial matching μ^I. In general, we consider μ^I to be individually rational under the true preferences. The rules of the game are as follows.

The game begins with a node at which nature chooses a sequence of firms at random. Each sequence corresponds to an order at which firms are given the opportunity to make proposals. Following nature's move, the first firm in the selected sequence has the chance to make a proposal. If unmatched under μ^I, the firm may propose to any worker or pass its turn. If matched under μ^I, it may simply fire its initial partner, propose to a different worker, or pass its turn and keep the initial partner.

In the case that a proposal is actually made, the game continues by having the proposed worker deciding whether to accept or to reject the offer. If he accepts, a new matching is formed where this worker and the proposing firm are together and their previous partners, if any, are unmatched. If he rejects, μ^I goes on unchanged. In the case that the firm simply chose to fire its initial worker, a new matching is formed where the firm and its former partner are unmatched, whereas if the firm chose "pass," the initial matching is preserved.

The second firm then moves and the game continues by giving firms the opportunity to make offers, in accordance with the order of the sequence. Each time a firm is called to play, the available moves depend on whether its position is vacant or not. If vacant, the firm may propose to any worker to whom it has not proposed before or simply pass its turn. Otherwise, it may fire the worker it holds, propose to a worker different from its current match and from any worker it has already proposed to, or pass its turn. When a worker receives a proposal, he may accept the offer or reject it and keep his former partner. Note that a matched worker is only allowed to reject his current position if he obtains and accepts an alternative offer.

The game continues as long as there is at least one firm wishing to make a new offer or to fire the incumbent worker. As soon as every firm in the market sequentially passes its turn, the game ends.

To complete the description of the game, we still have to specify the information that each agent possesses throughout the game. It is sensible to assume that in labor markets where myriads of firms and workers interact, each agent only becomes aware of events as they directly impinge on him. In the particular case of a firm, this means that it learns only if the proposal it made was accepted or rejected, or if its position became vacant. Hence, a firm's information set is defined by its initial partner and an ordered list of workers to whom it proposed, along with their reactions. Similarly, a worker is only aware of events that directly affect him. A worker's information set is identified by his initial position, as well as an ordered list of proposals received, his own responses, and firings. The initial chance move is never observable.[11]

3.3.2 Chance Move

Let us now focus on nature's move. At each moment in time, a randomly selected firm is given the chance to play. This random selection should not be interpreted merely as every firm having equal probability of proposing at each step. It may reflect some institutional—and perhaps inherently uncertain—features of the market which are not modeled. In fact, in decentralized markets matching is performed over the telephone network, using the mail, or through the Internet. In such environments, randomness determines the order in which agents communicate: it may depend on which telephone call goes through, on the speed of the mail, or on how fast firms react to eventual proposals. Or it may even be the case that there exists a natural order in which firms are expected to propose—firms that have potentially more to gain will certainly devote more resources into finding the right worker for their position and are, therefore, more likely to make offers.

To be precise, the game starts with a lottery prescribing a sequence of firms that defines the subsequent moves. A sequence corresponds to one of the innumerable possible orders in which firms are allowed to act. We assume that every sequence is infinite and that, in each sequence, every single firm appears infinitely many times. We also assume that every sequence has positive probability of occuring. The sample space over which this probability distribution is defined is denoted by O and o is an arbitrary sample point, a sequence of firms.

[11] Such low information environment may be enriched. It may be the case that agents learn of the actions of the others, even though they are not immediatly affected by them. The validity of the results that follow will be discussed for broader information structures.

Note that, even though we consider infinite sequences of firms, every play of the game ends in finite time. In fact, as firms are only allowed to propose to each worker once and, obviously, firing is possible only if matched, the moment comes when every firm chooses to pass its turn if called to play, either because it is happy with the incumbent worker, or because passing is the only available action. Moreover, the fact that a firm appears infinitely many times in each sequence guarantees that this moment comes and the firm is actually called to play. The end is then reached for every play of the game.

3.3.3 The Strategy Space

In what follows, we will describe agents' strategies and introduce some notation. A player's strategy in the Decentralized Game complies with the usual definition of behavioral strategy in an extensive form game, *i.e.*, a plan of action for each information set where he is called to act. However, in the context of a matching market there is a class of strategies worth emphasizing, strategies that resemble those used in a centralized market. Following Blum, Roth, and Rothblum (1997) we will call these strategies "preference strategies." Such strategies obey a consistency criterion in which agents decide how to move at any information set basing on a list of preferences, including those information sets that would not be reached had that list actually been used. Hence, deviations are regarded as temporary mistakes and further moves fit in the original list. To make things clear, when using a preference strategy, a firm selects an ordered list of potential matches and, whenever called to propose, makes the offer to the best worker on its list to whom it has not proposed before; likewise, a worker decides whether to accept or to reject a new proposal by comparing it with his current position on his list.

Even though the lists of preferences that serve as guidance do not have to faithfully reveal agents' true preferences, the set of preference strategies represents merely a small part of the set of feasible strategies.[12] For example, a worker w's strategy of accepting only the first proposal he gets and rejecting all the others is not consistent with any list of preferences. In fact, different plays of the game induce different orders of proposals; thus, depending on the play of the game, w's first proposal may be from, say, f and f'. It follows that f may be revealed preferred to f' by w or vice-versa, which clearly cannot be consistent with a preference list.

As for notation, actions are taken at decision nodes, typically denoted by x. A strategy profile σ specifies a strategy for each agent; 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. Preference strategies will be denoted by the corresponding preference profile—Q_v, for example, is a preference strategy for v—while P_v

[12] We refer to Chapter 2 for the analysis of a job matching game where the strategy space is confined to the set of preference strategies.

always denotes v's true preferences. A sequence of firms o and a strategy profile σ determine a play of the game, denoted by π.

3.3.4 Random Matching and Ordinal Nash Equilibria

In the Decentralized Game, different plays of the game with the same strategy profile may yield different output matchings, depending on the order of proposals. This applies even in the case that agents use preference strategies, as the following example illustrates.

Example 3.1 *The outcome depends on the selection of the order by which firms propose.*

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

$$\begin{array}{ll} P_{w_1} = f_2,\ f_1 & P_{f_1} = w_1,\ w_2 \\ P_{w_2} = f_1,\ f_3,\ f_2 & P_{f_2} = w_2,\ w_1 \\ & P_{f_3} = w_2. \end{array}$$

Note that the unique stable matching for this market is $\mu = \{(f_1,\ w_2), (f_2,\ w_1)\}$. Now consider the Decentralized Game with $\mu^I = \{(f_1,\ w_2)\}$ when agents play according to their true preferences P.

Start by considering the case in which f_3 is the first to make an offer. Given that f_3 is using P_{f_3}, it proposes to the only acceptable worker, w_2, and w_2 rejects this proposal, as he is initially matched to f_1, the best firm on his list. Then, it may be the case that either f_1's or f_2's opportunity comes. Let us say f_1 makes an offer; it proposes to w_1, the first worker in P_{f_1}, who is currently unmatched and thus accepts the proposal. Once this proposal is accepted, w_2 is left unmatched. Hence, when f_2 is given the chance to propose, w_2 accepts its offer. In the following moves every firm passes its turn, so that the game ends with the final non-stable matching $\hat{\mu} = \{(f_1,\ w_1), (f_2,\ w_2)\}$.

Nevertheless, if the first randomly chosen firm is f_2, its proposal to w_2 is refused, as this worker is still matched to f_1 and f_1 is preferred to f_2 in P_{w_2}. The next firm to propose can either be f_1, f_2, or f_3. Assume f_2 is the first to propose. It proposes to w_1, the second worker on its list, and w_1 accepts. Next, if f_1's turn comes, it proposes to w_1, who rejects this offer, since he is matched to his top choice f_2. So imagine f_1 is called to propose once more, tendering an offer to w_2, who accepts it. When finally f_3 proposes to w_2, he rejects the offer, given that he is already holding the highest ranked firm in his preference list. This play of the game terminates when the three firms are given the chance to pass their turns and the matching $\mu = \{(f_1,\ w_2), (f_2,\ w_1)\}$ is reached as the outcome of the game. ◇

Given an initial matching and a strategy profile, all the uncertainty on the order of play as described above is fully translated into a probability distribution over the set of matchings. Hence, fix a probability distribution on O and take an initial matching μ^I, a preference profile P, and an arbitrary worker w (what follows also holds for a representative firm, with obvious modifications). We will let $\widetilde{DG}^{\mu^I}[\sigma]$ denote the probability distribution over the set of matchings induced by the Decentralized Game starting from μ^I when the strategy profile σ is used and $\widetilde{DG}^{\mu^I}[\sigma](w)$ is the distribution that $\widetilde{DG}^{\mu^I}[\sigma]$ induces over $F \cup \{w\}$. The expression $\Pr\{\widetilde{DG}^{\mu^I}[\sigma] = \mu\}$ represents the probability that μ is the final matching of the Decentralized Game with the strategy profile σ. Moreover, $\Pr\{\widetilde{DG}^{\mu^I}[\sigma](w)R_w v\}$ is the probability that, in the Decentralized Game, w obtains a partner at least as good as v when σ is adopted. Observe that these probabilities rest on the probability distribution on O, but all the results that follow hold regardless of this lottery.

To address strategic questions we need to develop ideas about what constitutes a "best decision" to be taken by an agent. With this purpose in mind, let σ be a strategy profile and again consider $w \in W$. We say that, given σ_{-w}, the strategy σ_w *stochastically* P_w*-dominates* σ'_w in the Decentralized Game if, for all $v \in F \cup \{w\}$, $\Pr\{\widetilde{DG}^{\mu^I}[\sigma_w, \sigma_{-w}](w)R_w v\} \geq \Pr\{\widetilde{DG}^{\mu^I}[\sigma'_w, \sigma_{-w}](w)R_w v\}$. Thus, for any level of satisfaction, the probability that w's match exceeds that level of satisfaction is greater under $\widetilde{DG}^{\mu^I}[\sigma_w, \sigma_{-w}]$ than under $\widetilde{DG}^{\mu^I}[\sigma'_w, \sigma_{-w}]$. This provides the basis for the solution concepts we will adopt throughout the paper.

Definition 3.1 *Let (F, W, P) be a matching market and let μ^I be the initial matching. The profile of strategies σ is an* ordinal Nash equilibrium *(*ON equilibrium*) in the Decentralized Game if, for each player v in V, σ_v stochastically P_v-dominates every alternative strategy σ'_v given σ_{-v}.*

Thus, by using a strategy other than σ_v, v will not be able to strictly increase the probability of obtaining any v' (an agent with whom it may end up matched) and all agents ranked higher than v' in its true preference list, P_v. This means that we will be concerned in finding a profile of strategies σ with the property that, once adopted by the agents, no one can profit by unilaterally deviating for all possible utility representations of the agents' preferences.

Finally, the notion of ordinal Nash equilibrium can be refined to account for the dynamic nature of the Decentralized Game.

Definition 3.2 *Let (F, W, P) be a matching market and let μ^I be the initial matching. The profile of strategies σ is an* ordinal subgame perfect Nash equilibrium *(*OSPN equilibrium*) in the Decentralized Game if it induces an ordinal Nash equilibrium in every subgame of the Decentralized Game.*

3.4 Equilibrium Analysis

We begin this section by exploring the relationship between ordinal Nash and subgame perfect ordinal Nash equilibria.

Proposition 3.1 *Let $|F| \geq 2$. Then, no information set is a singleton.*

Lemma 3.1 *Let $|F| \geq 2$. Let x and x' be the two last decision nodes of the play of the game π, such that x' precedes x. Then, x and x' belong to two different firms and both firms choose the action "pass" at these nodes.*

Proof. First, notice that the game ends when every firm has sequentially chosen "pass." Given that x and x' precede the terminal node reached with π and that $|F| \geq 2$, it follows that the action taken at these nodes must be "pass." Now suppose, by contradiction, that both x and x' are firm f's decision nodes. Since, when π is considered, the game ends after f chooses "pass" at x, every firm other than f must have chosen "pass" in the nodes that precede x. Hence, every firm other than f has passed its turn in the nodes that precede x'. The rules of the game thus imply that the game ends immediately after f chooses "pass"at x' and we reach a contradiction: x is not a decision node. ■

Lemma 3.2 *Let $|F| \geq 2$. Let π be a play of the game and let x be a node of f reached along π, such that the game does not end after f's choice at x along π. Then, there exists a firm f' that still has a chance to act in π.*

Proof. Immediate from Lemma 3.1. ■

Proof. [Proof of Proposition 3.1] Let x be a node that belongs to f in π when nature draws the sequence o. Let f's move at x correspond to the kth element of o. We will prove that there exists a sequence o' and a node x' reached when nature draws o', such that x and x' belong to the same information set.

First, assume that the game does not end after f's choice at x along π. By Lemma 3.2, there exists a firm f' that still has the chance to act along π. Now let o' be a sequence whose k first elements are the same as those in o, but that differs from o in that f is inserted in position $k+1$ and all the remaining elements are identical. Consider any play of the game where nature draws o' and every agent chooses exactly the same actions as along π up to the point where o'_{k+1} is called to play. Let x' be the node corresponding to f's move in position k of the sequence o'. It is clear that x' belongs to the same information set as x, since every action, except for the unobservable nature's move, is the same along π and π'.

Now let x be a node of f, reached along π, such that f's action at x is the last action in π. By Lemma 3.1, there exists a firm $f' \neq f$ that has had the chance to move immediately before f moves at x, *i.e.*, in position o_{k-1} of the sequence, and both have chosen "pass." Now let o' be a sequence whose first $k-1$ elements coincide with those of o, but where f' occupies the position o_k and f occupies the position o_{k+1}. Consider the play of the game π' where nature draws o', every agent up to the element $k-1$ in the sequence chooses exactly the same action as in π, and f' chooses "pass" when called to play at the kth position of the sequence. Let x' be the node reached in π' where f acts in position $k+1$. Since f cannot observe nature's moves nor f''s action, it holds exactly the same information in both x and x'. Hence, x' belongs to the same information set as x.

Now consider π where nature draws o and along which some worker w may accept or reject a proposal made by firm f. Let x be the node where w acts and let f's proposal correspond to the kth element of o. Lemma 3.1 ensures that the game does not end after w's move at x. Hence, let o' be any sequence whose k first elements are the same as those in o, but such that the elements in position $k+1$ are different. Define π' as a play of the game in which nature draws o' and every other player chooses the same actions as along π up to the point where w reacts to f's proposal. Let x' be the node where w takes such decision. Since nature's draws are not observable, w's information is exactly the same in x and in x'. It follows that the information set containing x is not a singleton. ■

An immediate implication of this result is that the set of ordinal Nash and subgame perfect ordinal Nash equilibria coincide. In fact, given that all information sets are non-singletons, the Decentralized Game has no proper subgames. It may be conjectured that this is due to the low information environment we have assumed. And there are labor markets in which agents may become aware of events that do not affect them directly—acquaintances and social networks in general may play an important role. However, considering an enriched information environment where agents perceive all the offers that are made, as well as the proposed workers' reactions, the arguments in the above proof remain valid, as long as nature's move remains unobservable. Roughly speaking, for every decision node x along some play of the game that includes a draw of nature o, it is always possible to find a decision node x' belonging to the same information set of x by building a different play of the game in the following way: add a single firm to o, let it choose "pass" in its new decision node, and let agents choose exactly the same actions as in the original play in every other node. The conclusion follows since every proposal, acceptance, and rejection is made respecting the original order. Hence, even in this extreme case, ordinal subgame perfect Nash coincide with ordinal Nash equilibria. In what follows, we will refer to these concepts indistinctly as ordinal equilibria.

The following theorem is the main result of this section. Individual rationality is an obvious necessary condition that every ordinal equilibrium outcome must fulfill. Here, we state that under every ordinal equilibrium play of the Decentralized Game where firms use lists of preferences, *some* form of stability is preserved. To be more precise, every matching that can be obtained under such a play is stable for the same profile of preferences. The following remark is used in the proof of the theorem.

Remark 3.1 *When using a preference strategy, a firm will not fire a worker it proposed to nor exchange him for another worker along any play of the Decentralized Game. In fact, when a proposal is made, the firm reveals that this particular worker is the best among all who have not rejected it. If the worker accepts, the only occasion under which the firm makes a proposal again is when the worker it holds resigns from his position.*

Theorem 3.1 *Let μ^I be an individually rational input matching for $(F, W, (Q_F, P_W))$. Assume that the strategy profile $\sigma = (Q_F, \sigma_W)$ is an ordinal equilibrium in the Decentralized Game. Then, the probability distribution obtained over the set of matchings is such that every element in its support is a member of $S(Q_F, P_W)$.*

Proof. Suppose that $\{\mu_1, ..., \mu_k\}$ is the support of the distribution induced over the set of matchings when agents use σ. Assume that for some $i \in \{1, ..., k\}$, $\mu_i \notin S(Q_F, \sigma_W)$. We will prove that σ is not an ordinal equilibrium.

Let π be a play of the game that results in μ_i. To start, notice that for every firm f it must be the case that its assignment, $\mu_i(f)$, is acceptable with respect to Q_f. In fact, once using Q_f, f never proposes, under any play of the game, to a worker that, according to Q_f, is considered worse than being unmatched. On the other hand, every worker must consider his partner acceptable with respect to P. Assume that this is not the case and that there exists a worker, say w, such that $w P_w \mu_i(w)$. Individual rationality of the matching μ^I implies $\mu_i(w) \neq \mu^I(w)$.

Hence, σ_w must include, at some point along π, accepting $\mu_i(w)$'s proposal. Now take an alternative strategy $\hat{\sigma}_w$ according to which no offer is accepted by w. By using $\hat{\sigma}_w$, w may end up unmatched or, if initially matched, keep his original partner $\mu^I(w)$, but he is never assigned to a firm considered unacceptable under P_w. Thus, the following holds:

$$1 = \Pr\{\widetilde{DG}^{\mu^I}[\hat{\sigma}_w, \sigma_{-w}](w)R_w w\} > Pr\{\widetilde{DG}^{\mu^I}[\sigma](w)R_w w\}$$

and σ_w is not a best reply to σ_{-w}.

We have proved that μ_i is individually rational. Thus, there must exist a blocking pair for μ_i when the preference profile (Q_F, P_W) is considered. Let us say (f, w) blocks μ_i, *i.e.*, $fP_w\mu_i(w)$ and $wQ_f\mu_i(f)$. This implies that f proposed to w and, by Remark 3.1, was rejected by w in the course of π. Hence, σ_w includes rejecting f in at least one of w's information sets. Now, define $\hat{\sigma}_w$ as the strategy according to which w chooses the same actions as under σ_w at every information set, except for those that lead to rejecting f. When using $\hat{\sigma}_w$, if f proposes to w, w accepts this proposal and rejects every subsequent offer.

First, let us prove that the probability of being assigned to f is strictly higher under $\hat{\sigma}_w$ than under σ_w. Recall that π is a play of the game leading to μ_i and let o be nature's move in π. We know that firm f must have proposed to w along π. If, instead of using σ_w, w deviates and acts according to $\hat{\sigma}_w$, by Remark 3.1, w will end up matched to f when nature draws o and σ_{-w} is used. Now let π' be the play of the game in which nature draws o', with $o' \neq o$, and players use $(\hat{\sigma}_w, \sigma_{-w})$. If f does not propose to w along π', w acts exactly as if using σ_w and ends up matched to the same partner as when nature draws o and players use σ. Otherwise, f and w are matched in the final matching. It follows that the probability of having f and w matched is strictly increased when w uses $\hat{\sigma}_w$.

In order to prove σ_w is not a best reply to σ_{-w}, assume, without loss of generality, that $P_w = f_1, f_2, ..., f_{m-1}, f, f_{m+1}, ..., w, ..., f_n$. Consider a firm f_j, with $j = 1, ..., m-1$, and consider all the plays of the game where σ is used and where w and f_j end up together in the final matching. Some of these plays may not give f_j assigned to w when he deviates and acts according to $\hat{\sigma}_w$. However, the only occasion under which this happens is when w obtains a proposal from f and ends up matched to f. Hence, the probability of having w matched to f or to a firm he considers better than $f$is strictly increased when w uses $\hat{\sigma}_w$. We have

$$\Pr\{\widetilde{DG}^{\mu^I}[\hat{\sigma}_w, \sigma_{-w}](w)R_w f\} > Pr\{\widetilde{DG}^{\mu^I}[\sigma](w)R_w f\},$$

contradicting that σ is an ordinal equilibrium. ∎

The importance of this result lies in two of its implications. Since the set of unmatched agents is the same for every matching that is stable in a matching market (McVitie and Wilson, 1970, and Roth, 1982), the same agents remain unmatched in every possible outcome of an ordinal equilibrium where firms use lists of workers to guide their decisions. Moreover, when we focus on equilibria where firms act according to their true preferences, stability with respect to the true preferences is guaranteed. Such straightforward form of behavior can be easily justified. In some settings, firms obey objective criteria when selecting whom to hire (*e.g.*, universities select students according to their grades, some firms choose their workers basing on scores given by a recruiting agency, student placement mechanisms assign students to public schools according to the area of residence,...). Even when firms are not constrained to follow such rules, hiring new workers embodies a process of aggregating the opinions of different individuals that compose a recruiting committee; hence, we may expect that a list of workers is fixed and all decisions are taken basing on that list. Having to decide what to do next at each moment in time seems to be a less plausible form of behavior. Finally, reverting to the true preferences is always an easy resort, given the multiplicity of available strategies and the complexity of the environment.

Ordinal equilibria always exist when the initial matching is individually rational. In particular, the following results show the existence of ordinal equilibria where firms use preference strategies.

Definition 3.3 *Let μ^I be an arbitrary matching. We say that μ is* individually rational with respect to μ^I *if $\mu \in IR(P)$ and if, for all $f \in F$, $w' = \mu^I(f) P_f \mu(f)$, implies $\mu(w') \neq w'$.*

We will denote by $IR^{\mu^I}(P)$ the set of all individually rational matchings with respect to μ^I. This set is always non-empty since it includes $S(P)$, the set of stable matchings (see Proposition 2.5).

Proposition 3.2 *Let μ^I be an individually rational matching for (F, W, P). Then, $S(P)$ is a subset of $IR^{\mu^I}(P)$.*

Proof. Consider $\mu \in S(P)$. We will prove that $\mu \in IR^{\mu^I}(P)$ using a contradiction argument. Assume that $\mu \in S(P)$. By definition of stability, this implies $\mu \in IR(P)$, but assume that there exists a firm f such that $w' = \mu^I(f) P_f \mu(f)$ and $\mu(w') = w'$. Stability of μ implies that $w' P_{w'} f$ and we get a contradiction: μ^I is not individually rational. Therefore, every stable matching is an element of $IR^{\mu^I}(P)$. ■

Since a stable matching exists for every marriage market (Gale and Shapley, 1962), $IR^{\mu^I}(P)$ is not empty for every individually rational matching μ^I.

Proposition 3.3 *Let μ^I be an individually rational matching for (F, W, P) and let $\mu \in IR^{\mu^I}(P)$. Then, there exists an ordinal equilibrium $\sigma = (Q_F, \sigma_W)$ in the Decentralized Game that leads to μ with probability one.*

Proof. Define $Q_f = \mu(f)$, for every firm f and let $\sigma_w = Q_w = \mu(w)$. It is clear that every play of the game with the profile σ will lead to the output matching μ.

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

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

One particular case worth exploring is the case in which the starting point is the empty matching. The Decentralized Game then becomes a stylized model of an entry-level labor market without commitment, where cohorts of vacant positions and cohorts of candidates become simultaneously available, and decisions are taken in a decentralized way. It turns out that starting from the empty matching allows us to take the analysis farther.

Proposition 3.4 *Let μ^I be the empty matching and let $\mu \in S(P)$. Then, there exists an ordinal equilibrium in the Decentralized Game where firms reveal their true preferences that yields μ with probability one.*

Proof. Let $\sigma = (P_F, \sigma_W)$ and define σ_w as follows. For every worker w matched under μ, σ_w is the strategy of accepting only $\mu(w)$ and rejecting every other proposal, while it leads to the rejection of all proposals, without exception, when w is unmatched under μ.

We start by showing that the profile of strategies σ always leads to the matching μ, *i.e.*, $\Pr\{\widetilde{DG}^{\mu^I}[\sigma] = \mu\} = 1$. If this is not the case, then there exists a play of the game leading to $\widehat{\mu} \neq \mu$. But this is equivalent to having a firm, say f, that ends up matched to a partner, $\widehat{\mu}(f)$, different from $\mu(f)$ for some instance of the game. Given that f is unmatched in the initial matching and that the only worker willing to accept f is $\mu(f)$, we must have $\widehat{\mu}(f) = f$ (as long as $f \neq \mu(f)$; otherwise it must be the case that $\widehat{\mu}(f) = \mu(f)$ and we have a contradiction). So assume that $\widehat{\mu}(f) = f$. Since $\mu(f)$ would accept f's proposal and f is acting according to its true preferences, it must be the case that $f P_f \mu(f)$. Hence, $\mu(f)$ is not acceptable and the stability of μ is contradicted.

Let us now prove that, for every firm f, P_f stochastically P_f-dominates every other strategy σ_f. We will start by considering the case in which $\mu(f) \neq f$. Given that the only worker who is willing to accept f is $\mu(f)$, by choosing its strategy appropriately, f can either be alone or hold $\mu(f)$ under the output matching. By stability of μ, $\mu(f) P_f f$; since truth telling guarantees that $\mu(f)$ is assigned to f with probability one, f cannot improve by switching its strategy. In the case that $\mu(f) = f$, no worker accepts its proposal, and the best it can achieve is staying unmatched. It follows that f cannot do better than being assigned to $\mu(f)$ and P_f stochastically P_f-dominates every other strategy σ_f.

Now take the case of an arbitrary worker, w. Suppose, by contradiction, that σ_w does not stochastically P_w-dominate strategy $\hat{\sigma}_w$. This implies that $\Pr\{\widetilde{DG}^{\mu^I}[P_F, \hat{\sigma}_w, \sigma_{-w}](w) R_w \mu(w)\} = 1$ and that there exists a firm, say f, such that the following holds: $\Pr\{\widetilde{DG}^{\mu^I}[P_F, \hat{\sigma}_w, \sigma_{-w}](w) = f\} > 0$ and $f P_w \mu(w)$. But this means that, for some draw of nature, f approaches w before making an offer to $\mu(f)$. In fact, it cannot be the case that f proposes to $\mu(f)$ first and he does not accept it, as $\mu(f)$ is acting according to his original strategy, $\sigma_{\mu(f)}$, defined above. Thus, f must prefer w to $\mu(f)$. However, in this case (f, w) forms a blocking pair for μ, contradicting the fact that μ is stable. ■

Hence, in a decentralized entry-level labor market every stable matching can be reached as the outcome of an ordinal equilibrium play of the game where firms stick to their true rankings. The (partially) converse statement is given by Theorem 3.1, ensuring that every such ordinal equilibrium guarantees stability. These results may be viewed as an extension of some known features of the game induced by Gale and Shapley's centralized mechanism (Roth, 1984b), where the underlying strategy space is confined to the set of preference strategies.

3.5 Discussion

In this section we put our results in perspective and discuss some of the underlying assumptions.

As mentioned in the Introduction, centralized procedures have been introduced in many matching markets in response to certain market failures. It has been argued that the stability of the mechanisms employed is crucial for their success. In fact, those centralized procedures that achieved stable outcomes resolved the market failures, while those producing unstable outcomes continued to fail.[13] Since many matching markets do not employ centralized matching procedures, and yet are not observed to experience such problems, we can suspect that some markets may reach stable outcomes by means of decentralized decision making without commitment. Theorem 3.1 provides support to this conjecture. To make things clear, let us return to Example 3.1. We have seen that some plays of the game lead to unstable outcomes for the true preferences (the matching $\hat{\mu}$ is not stable). Nevertheless, Theorem 3.1 implies that if we expect agents to use equilibrium strategies and, by best replying, firms faithfully reveal their true preferences, then a stable matching is reached. Hence, if equilibrium predictions are to be taken seriously, the success of some decentralized markets is explained.

It is now probably worth discussing the robustness of the results to some changes in the rules of the game. First, throughout the game matchings are formed and dissolved as agents act in what they perceive to be their own best interest. We may think of this as a mere negotiation process, where no contracts are signed and where these temporary matchings would be the ones prevailing should negotiations suddenly stop. Alternatively, considering that provisional matchings are indeed consummated amounts to assuming that agents are free to recontract without any restrictions whatsoever. In the other extreme, we can consider that it is too costly to fire a worker. Hence, only firms with vacancies will actually make proposals and the Decentralized Game falls in the realms of Blum, Roth, and Rothblum's analysis. Blum, Roth, and Rothblum (1997) study how markets for senior positions may be re-stabilized after new firms have been created or workers have retired. In fact, stability for the true preferences is achieved in every equilibrium where firms act according to their true preferences, as long as the starting point is a firm-quasi-stable matching, *i.e.*, a matching whose stability has been disrupted by the creation of a new position or the retirement of a worker. Hence, Theorem 3.1, which allows for having any individually rational as an initial matching, no longer holds. The validity of Proposition 3.3 is also compromised: if we start form an initial matching where every firm is matched, no firm will be allowed to hire a new worker and the initial matching situation will be preserved, independently of the strategies used. In this setting, the majority of the results on equilibria depend on having a firm-quasi-stable matching as a starting point.

A different issue concerns providing workers with the initiative to propose. In some real labor markets, not only firms, but also workers may defy their preferred firms and we may account for this in the Decentralized Game. Hence, suppose that

[13] See, for example, Roth (1984a, 1990, 1991).

at each moment in time, an agent, either a firm or a worker, is randomly selected and makes an offer to someone in the other side of the market to whom it has never proposed to nor received a proposal from. The agent that receives the offer can only accept, or reject and keep his former partner, if a former partner existed. The game ends when every agent in the market passes its turn. It turns out that, starting from an arbitrary matching, every individually rational matching can be obtained in an ordinal equilibrium play of the game, so that the scope of Proposition 3.3 is enhanced. In what Theorem 3.1 and its implications are concerned, stability is robust to sophisticated behavior by the one side of the market, provided that the other side acts in accordance with the true preferences. To be precise, every matching sustained at an ordinal equilibrium is stable with respect to the true preferences whenever firms (respectively workers) faithfully transmit their preferences and workers (respectively firms) behave strategically by using strategies that may reveal different orderings of the other side of the market in different executions of the algorithm. Finally, in the particular case that the initial matching is the empty matching, every stable matching can be reached with probability one in an equilibrium where one side of the market truthfully reveals its preferences.[14]

3.6 Concluding Remarks

The present paper attempts to extend the two-sided matching theory by constructing a game that mimics the behavior of decentralized labor markets. Equilibrium analysis in a random context is performed at the expense of using an ordinal equilibrium concept—justified by the ordinal nature of agents' preferences—but that allows for obtaining some interesting results. Namely, equilibria where firms use preference strategies always exist and lead to matchings that preserve stability for a particular profile of preferences. Furthermore, when we consider an ordinal equilibrium where firms act truthfully, stability for the true preferences is achieved in every outcome matching. This fact may account for the success of some decentralized labor markets. A case of particular interest has the empty matching as the starting point of the game. Here we give a fairly complete characterization of ordinal equilibria.

It is natural to ask to what extent the stylized model constructed here can serve as a description of a real decentralized labor market. The marriage model is perhaps too simple. Aside from the assumption that each firm has a unique position to fill, the important unrealistic feature lies in considering that the salary associated with each position is a fixed part of the job description, rather than something to be negotiated between each firm and prospective worker. It thus remains important to explore models where these assumptions are relaxed, even though we believe that

[14] A formal statement of these results and their proofs are given in the Appendix, in Propositions 3.5, 3.6, and 3.7.

the present analysis provides a good starting point to understand the functioning of some decentralized labor markets.

3.7 Appendix

In this section we extend some of the above results to the case in which both sides of the market are able to tender offers. First, it can easily be shown that no information set is a singleton, as long as there are at least two agents in the market, *i.e.*, $|V| \geq 2$.[15] It follows that subgame perfect ordinal Nash equilibria coincide with ordinal Nash equilibria. The remaining results are proved in what follows.

Proposition 3.5 *Let μ^I be an arbitrary matching in (F, W, P) and let $\mu \in IR(P)$. Then, there exists an ordinal equilibrium in the Decentralized Game that leads to μ with probability one.*

Proof. Let $Q_v = \mu(v)$, for all $v \in V$. Clearly, every play of the game with strategy profile Q leads to μ. We will show that Q is an ordinal equilibrium. In the case that v is such that $\mu(v) \neq v$, the only agent that proposes to or accepts a proposal from v is $\mu(v)$. Hence, no deviation will improve v's match. Otherwise, for v such that $\mu(v) = v$, no agent is willing to match v. As before, by switching strategy, v cannot end up matched and improve his position. ■

In what follows, we extend Theorem 3.1. The result is stated for equilibria in which firms use preference strategies and workers are allowed to have other forms of behavior. Note however that we restrict to equilibria where a worker's strategy is consistent with a list of preferences along each play of the game (even though it may correspond to incompatible lists when different plays of the game are considered). A similar result, where the roles of firms and workers are interchanged, can be proved.

Proposition 3.6 *Let μ^I be an arbitrary matching in (F, W, P). Assume that the strategy profile $\sigma = (Q_F, \sigma_W)$ is an ordinal equilibrium in the Decentralized Game, where σ_w is consistent with a list of preferences in each play of the game, for all $w \in W$. Then, the probability distribution obtained over the set of matchings is such that every element in its support is a member of $S(Q_F, P_W)$.*

Proof. Suppose that $\{\mu_1, ..., \mu_k\}$ is the support of the distribution induced over

[15] The reasoning behind the proof of Proposition 3.1 remains valid, but instead of analysing decision nodes that belong to firms and to workers as separate cases, the distinction to be made is between nodes where proposals are issued, and those where acceptances or rejections take place.

the set of matchings when agents use σ and assume that for some $i \in \{1, ..., k\}$, $\mu_i \notin S(Q_F, \sigma_W)$. We will prove that σ is not an ordinal equilibrium.

We will denote by π a play of the game leading to μ_i. To start, notice that for every firm f it must be the case that its assignment, $\mu_i(f)$, is acceptable with respect to Q_f. In fact, once using Q_f, f never proposes to nor accepts a proposal from a worker that, according to Q_f, is considered worse than being unmatched. On the other hand, every worker must find his partner acceptable. Assume that this is not the case and that there exists a worker, say w, such that $wP_w\mu_i(w)$. Now take an alternative strategy $\hat{\sigma}_w$ according to which w resigns from $\mu^I(w)$—if w is initially matched—and accepts no offers. By using $\hat{\sigma}_w$, w ends up unmatched in every play of the game. Hence, $1 = \Pr\{\widetilde{DG}^{\mu^I}[\hat{\sigma}_w, \sigma_{-w}](w)R_w w\} > Pr\{\widetilde{DG}^{\mu^I}[\sigma](w)R_w w\}$ and σ_w is not a best reply to σ_{-w}.

Individual rationality of μ_i of (Q_F, P_W) is proven. Thus, there must exist a blocking pair for μ_i when the preference profile (Q_F, P_W) is considered. Let us say (f, w) blocks μ_i, *i.e.*, $fP_w\mu_i(w)$ and $wQ_f\mu_i(f)$. This implies that, in the course of π, either *(i)* f proposed to w or *(ii)* w proposed to f. If *(i)* holds, by Remark 3.1, f was rejected by w and we can prove that σ_w is not a best reply to σ_{-w} using the same arguments as in the proof of Theorem 3.1. Otherwise, in case *(ii)*, since w uses a strategy that is consistent with a list of preferences under π, f must have rejected w. (The reasoning behind this relies in arguments similar to those of Remark 3.1.) In this case we can find a successful deviation for f. In fact, define $\hat{\sigma}_f$ as the strategy according to which f chooses the same actions as under σ_f at every information set, except for those that lead to rejecting w when w proposes. Hence, when using $\hat{\sigma}_f$, if w proposes to f along a play of the game, f accepts this proposal and holds it until the end of this play. For every play of the game in which w does not propose to f, f acts exactly as when using σ_f.

First, we will prove that the probability of being assigned to w is strictly higher under $\hat{\sigma}_f$ than under σ_f. Recall that π is a play of the game leading to μ_i and let o be nature's move in π. We know that w must have proposed to f along π. Once f deviates and acts according to $\hat{\sigma}_f$, f will end up matched to w when nature draws o and σ_{-f} is used. Now let π' be the play of the game in which nature draws o', with $o' \neq o$, and players use $(\hat{\sigma}_f, \sigma_{-f})$. If w does not propose to f along π', f ends up matched to the same partner as when nature draws o and players use σ. Otherwise, f and w are matched in the final matching. It follows that the probability of having f and w matched is strictly increased when f uses $\hat{\sigma}_f$.

In order to complete the proof that σ_f is not a best reply to σ_{-f}, assume, without loss of generality, that $P_f = w_1, w_2, ..., w_{l-1}, w, w_{l+1}, ..., f$. Consider a worker w_j, with $j = 1, ..., l-1$, and consider all the plays of the game where σ is used and where f and w_j end up together in the final matching. Such plays may not give w_j assigned to f when f switches to $\hat{\sigma}_f$. However, the only occasion under which this happens is when f obtains a proposal from w and ends up matched to him. Hence, the probability of having f matched to w or to a worker it considers better than w is strictly increased when f uses $\hat{\sigma}_f$. We have

$$\Pr\{\widetilde{DG}^{\mu^I}[\hat{\sigma}_f, \sigma_{-f}](f) R_f w\} > Pr\{\widetilde{DG}^{\mu^I}[\sigma](f) R_f w\},$$

contradicting that σ is an ordinal equilibrium. ■

Proposition 3.7 *Let μ^I be the empty matching and let $\mu \in S(P)$. Then, there exists an ordinal equilibrium in the Decentralized Game where one side of the market reveals its true preferences that yields μ with probability one.*

Proof. We analyze the case in which firms act according to their true preferences; the same arguments hold, with the roles of firms and workers reversed, when workers act straightforwardly. Hence, consider $\sigma = (P_F, \sigma_W)$ and define σ_w as follows: if w is matched under μ, σ_w is the strategy of always choosing "pass" when called to propose and accepting only $\mu(w)$'s proposal; while if w is such that $\mu(w) = w$, no proposal is made nor accepted by w.

We start by showing that the profile of strategies σ always leads to the matching μ. If this is not the case, then there exists a play of the game leading to $\widehat{\mu} \neq \mu$. But this is equivalent to having a firm, say f, that ends up matched to a partner, $\widehat{\mu}(f)$, different from $\mu(f)$ for some instance of the game. Given that workers make no proposals and that the only one willing to accept f is $\mu(f)$, we must have $\widehat{\mu}(f) = f$ (as long as $f \neq \mu(f)$; otherwise it must be the case that $\widehat{\mu}(f) = \mu(f)$ and we have a contradiction). So assume that $\widehat{\mu}(f) = f$. Since $\mu(f)$ would accept f's proposal and f is acting according to its true preferences, it must be the case that $f P_f \mu(f)$. Hence, $\mu(f)$ is not acceptable and the stability of μ is contradicted.

Let us now prove that, for every firm f, P_f stochastically P_f-dominates every other strategy σ_f. We will start by considering the case in which $\mu(f) \neq f$. Given that workers do not issue offers and that the only worker who is willing to accept f is $\mu(f)$, by choosing its strategy appropriately, f can either be alone or hold $\mu(f)$ under the output matching. By stability of μ, $\mu(f) P_f f$; since truth telling guarantees that $\mu(f)$ is assigned to f with probability one, f cannot improve by deviating. In the case that $\mu(f) = f$, no worker accepts its proposal nor proposes to f, and the best it can achieve is staying unmatched. It follows that f cannot do better than being assigned to $\mu(f)$ and P_f stochastically P_f-dominates every other strategy σ_f.

Now take the case of an arbitrary worker, w. Suppose, by contradiction, that σ_w does not stochastically P_w-dominate a different strategy $\hat{\sigma}_w$. Then, $\Pr\{\widetilde{DG}^{\mu^I}[P_F,$

$\hat{\sigma}_w, \sigma_{-w}](w) R_w \mu(w)\} = 1$ and that there exists a firm, say f, such that the following holds: $\Pr\{\widetilde{DG}^{\mu^I}[P_F, \hat{\sigma}_w, \sigma_{-w}](w) = f\} > 0$ and $f P_w \mu(w)$. Let π be a play of the game where f and w are matched. By stability of μ, $\mu(f) P_f w$, so that f proposes to $\mu(f)$ in the course of π. Given the outcome matching, $\mu(f)$ rejects f's proposal. This contradicts the definition of $\mu(f)$'s strategy. ■

Chapter 4
Random Stable Mechanisms in the College Admissions Problem

4.1 Introduction

The study of two-sided matching has been mainly devoted to centralized markets. These matching markets work by having each agent of the two sides of the market submit a rank ordered preference list of acceptable matches to a central clearinghouse, which then produces a matching by processing all the preference lists according to some algorithm. Typically, such mechanisms are deterministic in the sense that the outcome depends on the submitted lists in a way that involves no element of chance. As a consequence, the existing results do not generally allow us to address behavior in many labor markets and other two-sided matching situations where lotteries ultimately determine the outcome. In discrete problems where agents have opposite interests randomization is surely one of the most practical tools to achieve procedural fairness.[16] Hence, equity considerations provide an important justification for the introduction of chance in many instances of centralized matching. On the other hand, lotteries are especially attractive as a means of representing the frictions of a decentralized market. Indeed, in the extremely complex environment of a real life market, decentralized decision making will often lead to uncertain outcomes: the question of who will match with whom depends on the realization of random events—random meetings.

This paper studies a class of matching mechanisms that are random: given agents' behavior, chance determines the final outcome. These mechanisms may be used in centralized markets as a means to promote procedural fairness. Or they may arise in the context of decentralized decision making: starting from an arbitrary matching, agents from the two sides of the market meet bilaterally in a random fashion. We assume that each individual has preferences over the other side of the market and the prospect of being unmatched; however, they are not compelled to behave in a straightforward manner, according to these true preferences. Instead, agents are confronted with a game in which they act in what they perceive to be their own best interest. Hence, upon meeting, the paired agents match if this is consistent with their strategies, and separate otherwise. Since one of the clearest lessons from the study of deterministic procedures is that

[16] At least to move towards procedural fairness. A random matching mechanism is procedurally fair whenever the sequence of moves for the agents is drawn from a uniform distribution. See Moulin (1997, 2003).

understanding such incentives is crucial to understand the behavior of the market, the paper is devoted to equilibrium analysis.

Our study was largely motivated by Roth and Vande Vate (1990, 1991). In the context of the marriage problem where matching is one-to-one, Roth and Vande Vate (1990) proved that, starting from an arbitrary matching, the decentralized decision making process of allowing randomly chosen blocking pairs to match will converge to a stable matching with probability one. Under a stable matching no individual or pair of agents has incentives to circumvent the matching. It is argued that such process can be thought of as an approximation to real life dynamics. In the related paper Roth and Vande Vate (1991), strategic considerations are made for the marriage market, focusing on the class of truncation strategies, *i.e.*, strategies that are order-consistent with true preferences, but may regard fewer partners as acceptable. In a one-period game in which every agent states a list of preferences and then a matching stable with respect to those preferences is selected at random, it is shown that all stable matchings can be reached as equilibria in truncations. However certain unstable matchings can also arise in this way. A multi-period extension is then considered to rule out such undesirable outcomes.

As in Roth and Vande Vate (1991) we assume that random meeting among agents will eventually converge to a stable matching with respect to the chosen strategy profile. Hence, such process induces a lottery exclusively over stable outcomes. However, the present paper extends their contribution in two ways. First, we take equilibrium analysis further, going beyond the analysis of truncations. A concept of equilibrium based on first-order stochastic dominance is used, given that preferences are ordinal in nature and probability distributions over matchings are to be compared. The notion of ordinal Nash equilibrium guarantees that each agent plays his best response to the others' strategies for every utility representation of the preferences.[17]

Second, the analysis is conducted in the context of the college admissions model. In this model, agents belonging to two disjoint sets (henceforth firms and workers) have preferences over the other side of the market; in addition, each firm can employ at most some fixed number of workers, while each worker can fill only one position. Strategic issues in this context have been studied for a deterministic stable matching mechanism. Roth (1985) shows that no stable mechanism exists that makes it a dominant strategy for all players to report their true preferences. Moreover, he proves that there are equilibrium misrepresentations that generate any individually rational matching with respect to the true preferences.[18] Ma (2002) shows that in order to obtain stability with respect to true preferences, we

[17] This concept was introduced in d'Aspremont and Peleg (1988); it has been used in the context of voting theory in Majumdar and Sen (2004) and in matching markets in Ehlers and Massó (2003), and Majumdar (2003).

[18] For a detailed explanation of these and other results see Roth and Sotomayor (1990), a comprehensive treatment of the matching problem.

have to use a refinement of the Nash equilibrium concept and restrict to truncations at the match point (*i.e.*, strategies that preserve the ordering of the true preferences, but rank as unacceptable all the agents that are less preferred than the current match). More precisely, all strong equilibria in truncations at the match point produce stable outcomes. Further, he establishes that every Nash equilibrium profile admits at most one stable matching with respect to the true preferences; if, indeed, such a matching is admitted, it will always be achieved.

In this paper we characterize equilibria arising in the game induced by a random stable mechanism, providing simultaneously some results that extend to deterministic mechanisms. First, we show that when ordinal Nash equilibria are considered, a unique matching is obtained as the outcome of the random process. In addition, this outcome is individually rational with respect to the true preferences. Since every individually rational matching for the true preferences can be achieved as an equilibrium outcome, we establish that a matching can be reached at an ordinal Nash equilibrium if and only if it is individually rational for the true preferences. We then turn our attention to equilibria where firms behave straightforwardly. In fact, there are reasons to contemplate truth telling as a salient form of behavior in situations involving uncertainty; further, sophisticated strategic play does not even make sense in settings where firms follow an objective criterion to fill their positions. We prove that, even though workers may not play straightforwardly, stability with respect to the true preferences holds for any matching that results from a play of equilibrium strategies in which firms reveal their true preferences. Conversely, every matching that is stable for the true preferences can be achieved as an equilibrium outcome. In closing, we relate the equilibrium strategy profiles in the games induced by both random and deterministic stable mechanisms. In particular, for any random stable mechanism that always assigns positive probability to two different stable matchings (when they exist), we show that a strategy profile is an ordinal Nash equilibrium if and only if it has a unique stable matching and there exists a deterministic stable mechanism where it is a Nash equilibrium.

We proceed as follows. In Section 4.2 we present the college admissions model and introduce notation. We describe the random matching mechanism and the equilibrium concept in Section 4.3. In Section 4.4 we turn our attention to individual decision making. The matching process is modeled as a one-period game and its equilibria are then characterized. In Section 4.5 we briefly discuss equilibria in the context of a sequential game. Some concluding remarks follow in Section 4.6.

4.2 The Model

The agents in the college admissions model consist of two finite and disjoint sets, the set $W = \{w_1, ..., w_p\}$ of workers and the set $F = \{f_1, ..., f_n\}$ of firms. We let $V = W \cup F$ and sometimes refer to a generic agent by v, while w and f represent a generic worker and firm, respectively. Each worker w can fill at most one position and has a strict preference relation P_w over the set $F \cup \{w\}$. Each firm f has a *quota* q_f, the maximal number of workers it may employ, and a strict preference ordering P_f over the set $W \cup \{f\}$. For example, the preferences of f can be represented by $P_f = w_3, w_1, f, w_2, ..., w_4$, indicating that the best worker for f is w_3, its second choice is w_1 and it prefers having a position unfilled to hiring any other worker. A worker is *acceptable* if the firm prefers to employ him rather than having a position unfilled; in the above example, the set of acceptable workers is $A(P_f) = \{w_1, w_3\}$. Similarly, given P_w we can define an acceptable firm and the set $A(P_w)$. In general, we describe only rankings of acceptable agents, so that the above preferences are abbreviated as $P_f = w_3, w_1$. Let $P = (P_{f_1}, ..., P_{f_n}, P_{w_1}, ..., P_{w_p})$ denote the profile of all agents' preferences; we sometimes write it as $P = (P_v, P_{-v})$ where P_{-v} is the set of preferences of all agents other than v. We write $v' P_v v''$ when v' is preferred to v'' under preferences P_v and we say that v *prefers* v' to v''. We write $v' R_v v''$, when v likes v' *at least as well as* v'' (note that v can only be indifferent if $v' = v''$).

Formally, a *matching market* is a triple (F, W, P). Define an unordered family of elements of V to be a collection of elements in which the order is immaterial. Let $\bar{V}$ denote the set of unordered families of elements of V. An outcome for the matching market (F, W, P) is a *matching*, a function μ from the set V to the set $\bar{V}$ satisfying the following: *(i)* $|\mu(w)| = 1$ for every $w \in W$ and $\mu(w) \in F$ whenever $\mu(w) \neq w$; *(ii)* $|\mu(f)| = q_f$ for every $f \in F$ and if $|W \cap \mu(f)| < q_f$ then $\mu(f)$ is filled to q_f with copies of f; *(iii)* $\mu(w) = f$ if and only if $w \in \mu(f)$.

Each worker's preferences over matchings correspond precisely to his preferences over his own assignments at the matchings. Similarly, firms' preferences over matchings are tantamount to the preferences over its assignments. Hence, in order to compare matchings, each firm with quota greater than one must be able to compare groups of workers. Following Roth (1985), it will be sufficient for the purpose of this paper to consider only the firms' preferences over individual workers, as long as their preferences over outcomes are *responsive* to the preferences over single agents. We say that a preference $\bar{P}_f$ for f over sets of workers is responsive to its preference P_f over single workers if, for any two matchings μ and μ', whenever $\mu(f) = \mu'(f) \cup \{w\} \backslash \{\sigma\}$ for $\sigma \in \mu'(f)$ and $w \notin \mu'(f)$, then $\mu(f) \bar{P}_f \mu'(f)$ if and only if $w P_f \sigma$. Responsive preferences are assumed throughout the paper. Finally, for any firm f with quota q_f, its *choice set* from any set $\bar{W}$ of workers, denoted by $Ch_{P_f}(\bar{W})$, is the set of f's q_f most preferred acceptable workers in $\bar{W}$ if $|\bar{W} \cap A(P_f)| \geq q_f$, and the entire set $\bar{W} \cap A(P_f)$ if $|\bar{W} \cap A(P_f)| < q_f$.

A matching μ is *individually rational* if for every $w \in W$, $\mu(w) R_w w$, and if for every firm f and w in $\mu(f)$, $w P_f f$. A firm f and a worker w are a *blocking pair* for μ if they are not matched under μ but prefer one another to one of their assignments, *i.e.*, $f P_w \mu(w)$ and $w P_f \sigma$ for some σ in $\mu(f)$. A matching μ is *stable* if it is individually rational and if there is no blocking pair for μ. Note that the stability of μ depends on preferences over individuals, irrespectively of the responsive extension that is being used. We let $IR(P)$ and $S(P)$ denote the set of all individually rational and the set of all stable matchings respectively with respect to a profile P. A firm f and a worker w are *achievable* for each other if f and w are matched under some stable matching.

The proof of existence of stable matchings in Gale and Shapley (1962) is constructed by means of the deferred-acceptance algorithm. For a given a preference profile P, proposals are issued by one side of the market accordingly, while the other side merely reacts to such offers by rejecting all but the best in P. In the case that firms make job offers, the algorithm arrives at the firm-optimal stable matching μ^F, with the property that all firms are in agreement that it is the best stable matching. The deferred-acceptance algorithm with workers proposing produces the worker-optimal stable matching μ^W with corresponding properties. Further, the optimal stable matching for one side of the market is the worst stable matching for every agent on the other side of the market, a result presented in Knuth (1976) but attributed to John Conway.

Finally, a *matching mechanism* φ is a function from preference profiles to matchings and we say that φ is *stable* if it produces a stable outcome for every preference profile, *i.e.*, $\varphi[P] \in S(P)$ for every preference profile P. We let φ^F and φ^W denote the matching mechanisms that yield μ^F and μ^W, respectively.

4.3 Random Matching and Ordinal Nash Equilibria

Many matching markets do not employ centralized procedures. Agents are free to issue offers and make acceptations and rejections as they please and matching is performed over the telephone network, using the mail, or through the Internet. In such environments, randomness determines the order in which agents communicate: it may depend on which telephone call goes through, on the speed of the mail, or on how fast firms react to eventual proposals. When a central clearinghouse does exist, chance is widely used to restore procedural fairness—any deterministic mechanism is bound to favor a subset of the agents involved. In two-sided matching markets, the need for compromise solutions is especially intense given the strong polarization of interests of agents reflected in the structure of the stable set. Some real life applications of random procedures concern allocation problems as on-campus housing, namely in American universities, or public housing.[19] Student placement mechanisms that assign students to colleges are another example of mechanisms where randomness plays a role, as well as procedures used to match students to optional courses or even children to summer camps.[20] Finally, randomness is present in any matching mechanism where the position in a queue or the order of arrival may influence assignments.

Formally, a random mechanism is a mapping from preference profiles to lotteries over the set of matchings. A random mechanism $\tilde{\varphi}$ and a preference profile Q induce a random matching $\tilde{\varphi}[Q]$. Throughout the paper, we only consider random stable matchings. Hence, $\tilde{\varphi}[Q]$ denotes the probability distribution induced over the set of stable matchings $S(Q)$ and $\tilde{\varphi}[Q](v)$ is the probability distribution induced over agent v's achievable matches. We will use, for example, $\Pr\{\tilde{\varphi}[Q](v)R_v\hat{v}\}$ to denote the probability that v obtains a partner at least as good as $\hat{v}$ according to v's true preferences when the profile Q is used in the mechanism $\tilde{\varphi}$. Whenever the probability distribution $\tilde{\varphi}[Q]$ is degenerate, we will abuse the notation slightly by letting $\tilde{\varphi}[Q]$ denote the unique outcome matching. Observe however that in general the support of $\tilde{\varphi}[Q]$, denoted by $supp\tilde{\varphi}[Q]$, is a subset of the set of stable matchings $S(Q)$.

In a matching market (F, W, P), we consider a game induced by $\tilde{\varphi}$ in which agents are each faced with the decision of what strategies to act on. As a first approach, we examine a one-period game where the strategy space of a player in the game is the set of all possible preference lists. Hence, given the true preference ordering P_v, each player v may eventually reveal a different order Q_v over the players on the other side of the market, and then a matching μ stable with respect to the stated preferences Q is selected at random among all the potential matchings, *i.e.*, the elements of $supp\tilde{\varphi}[Q]$, with probability $Pr\{\tilde{\varphi}[Q] = \mu\}$. An extension of the results obtained to a more complex setting is discussed in Section 5.

To address strategic questions we need to develop ideas about what constitutes a "best decision" to be taken by an agent. With this purpose in mind, let $\hat{Q}$ be a strategy profile and consider $w \in W$ (what follows also holds for a representative firm, with obvious modifications). Given a random stable mechanism $\tilde{\varphi}$, we say that, given $\hat{Q}_{-w}$, the strategy Q_w *stochastically* P_w*-dominates* Q'_w if, for all $v \in F \cup \{w\}$, $\Pr\{\tilde{\varphi}[Q_w, \hat{Q}_{-w}](w)R_w v\} \geq \Pr\{\tilde{\varphi}[Q'_w, \hat{Q}_{-w}](w)R_w v\}$. Thus, for all $v \in F \cup \{w\}$, the probability of w being assigned to v or to a strictly preferred agent is higher under $\tilde{\varphi}[Q_w, \hat{Q}_{-w}](w)$ than under $\tilde{\varphi}[Q'_w, \hat{Q}_{-w}](w)$. Hence, if we consider the problem that player w faces given the strategy choices $\hat{Q}_{-w}$ of the other players, a particular strategy choice Q_w may be preferred if, given $\hat{Q}_{-w}$, it stochastically dominates every other alternative strategy. This provides the basis for the solution concept we will adopt throughout the paper.

19 See Abdulkadiroglu and Sönmez (1999).

20 See Abdulkadiroglu and Sönmez (2003) for a description of student assignment mechanisms.

Definition 4.1 *Given the profile of preferences P, the profile of strategies Q is an* ordinal Nash equilibrium *(*ON equilibrium*) in the game induced by $\tilde{\varphi}$ if, for each player v in V, Q_v stochastically P_v-dominates every alternative strategy Q'_v given Q_{-v}.*

Thus, by using a strategy other than Q_v, v will not be able to strictly increase the probability of obtaining any v' (an agent with whom it may end up matched) and all agents ranked higher than v' in its true preference list, P_v. This means that we will be concerned in finding a profile of strategies Q with the property that, once adopted by the agents, no one can profit by unilaterally changing its strategy for all possible utility representations of the agents' preferences.

4.4 Equilibrium Analysis

We now turn to characterize ordinal Nash equilibria in the game induced by a random stable mechanism $\tilde{\varphi}$. Proposition 4.1 asserts that no ordinal equilibrium supports more than one stable matching. Using the decentralized interpretation, we can say that the outcome in equilibrium is immune to the order in which agents meet when players behave strategically, even though truth revealing often leads to a lottery over matchings. Agents manipulate to protect themselves against uncertainty.

Proposition 4.1 *Let Q be an ordinal Nash equilibrium in the game induced by a random stable mechanism $\tilde{\varphi}$. Then, a single matching is obtained with probability one.*

Proof. By contradiction, assume that Q is an ON equilibrium and $|supp\tilde{\varphi}[Q]| \geq 2$. Then, there exists a worker $w \in W$ and matchings μ, $\hat{\mu} \in supp\tilde{\varphi}[Q]$ such that $\mu(w) \neq \hat{\mu}(w)$. Let $\mu'(w)$ be the best match among all given by the elements of $supp\tilde{\varphi}[Q]$, *i.e.*, $\mu'(w) R_w \mu(w)$, for all $\mu \in supp\tilde{\varphi}[Q]$. Define $Q'_w = \mu'(w)$ and $Q' = (Q'_w, Q_{-w})$. Note that μ' is stable for Q and, once w changes his strategy, it remains stable for Q' (it remains individually rational and no blocking pairs emerge). Further, since the set of matched agents is the same under every stable matching, w is matched to $\mu'(w)$ under every matching in $S(Q')$. Then, $1 = Pr\{\tilde{\varphi}[Q'](w) R_w \mu'(w)\} > Pr\{\tilde{\varphi}[Q](w) R_w \mu'(w)\}$ and Q_w does not stochastically P_w-dominate Q'_w. It follows that Q is not an equilibrium. ■

As a consequence, in the particular case that the random mechanism always assigns positive probability to at least two different matchings (if such matchings exist), the set of stable matchings of each ordinal Nash equilibrium is a singleton.

In general, however, the set of stable matchings of an ordinal Nash equilibrium may contain several elements. As proved in Ma (2002) for a deterministic stable mechanism, the random stable mechanism then chooses the matching that is unanimously preferred among all the stable matchings with respect to the submitted profile.

Lemma 4.1 *Let Q be an ordinal Nash equilibrium in the game induced by a random stable mechanism $\tilde{\varphi}$. Then, for any matching $\mu \in S(Q)$,*

1. $\tilde{\varphi}[Q](w) R_w \mu(w)$ *for every* $w \in W$ *and*
2. $\tilde{\varphi}[Q](f) \bar{R}_f \mu(f)$ *for every* $f \in F$ *and every responsive extension* $\bar{R}_f$ *of* R_f.

Proof. By Proposition 4.1, $\tilde{\varphi}[Q]$ is a single matching. The result then follows from Lemma 6 in Ma (2002). ■

For illustration, consider the following example.

Example 4.1

Let $F = \{f_1, f_2\}$, $W = \{w_1, w_2\}$, and $q_{f_1} = q_{f_2} = 1$. Suppose that the true preferences are as follows:

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

Define $Q_{w_1} = f_2,\ f_1$ and $Q_{w_2} = f_1,\ f_2$ and note that the preference profile $Q = (Q_{w_1}, Q_{w_2},\ P_F)$ is an ordinal Nash equilibrium in the game induced by the mechanism that yields the firm-optimal stable matching. Now let $\tilde{\varphi}$ be a random mechanism that assigns probability 0.5 to both the worker-optimal and firm-optimal stable matchings. Clearly, the support of the probability distribution induced by $\tilde{\varphi}[Q]$ includes both $\mu^F[Q] = \{(f_1, w_1), (f_2, w_2)\}$ and $\mu^W[Q] = \{(f_1, w_2), (f_2, w_1)\}$. By Proposition 4.1, Q is not an ordinal Nash equilibrium in the game induced by $\tilde{\varphi}$. In fact, every worker can successfully deviate. For example, by using his true preferences, w_1 obtains his preferred firm f_1 with probability one. ◇

In the context of deterministic mechanisms, Roth (1985) shows that by suitably falsifying their preferences, agents can induce any individually rational matching with respect to the true preferences. Unfortunately, this is not a very illuminating result: the set of individually rational matchings includes all the matchings that are remotely plausible. Moreover, the possibility of sustaining matchings where agents hold non-acceptable partners is not ruled out, although individual rationality appears to be a minimum requirement for an equilibrium outcome.

The results that follow establish that μ can be supported as an ordinal equilibrium if and only if it is individually rational. Hence, we provide a complete characterization of ordinal Nash equilibria outcomes in the game induced by $\tilde{\varphi}$. Furthermore, it can easily be shown that Proposition 4.3 can be extended to the deterministic case, providing a necessary condition for Nash equilibria in games induced by deterministic stable mechanisms.

Proposition 4.2 *Let μ be any individually rational matching for (F, W, P) and let $\tilde{\varphi}$ be a random stable mechanism. Then, there exists an ordinal Nash equilibrium Q that supports μ in the game induced by $\tilde{\varphi}$.*

Proof. Let $Q_w = \mu(w)$, for every $w \in W$, and let Q_f be such that $A(Q_f) = \mu(f) \cap W$, for every $f \in F$. Clearly, $S(Q) = \{\mu\}$ and μ is reached with probability one. Moreover, no agent can profitably deviate. To see this, take an arbitrary worker w. The only agent that accepts w is $\mu(w)$. Hence, w faces the choice of holding $\mu(w)$, if $\mu(w) \in F$, or being unmatched. Since $\mu(w) R_w w$, w has no profitable deviation. Now consider $f \in F$. Only those workers in $\mu(f)$ are willing to accept filling a position in f. Moreover, $\mu(f) \bar{R}_f W^S$, for every $W^S \subseteq \mu(f)$, by individual rationality of μ. It follows that f cannot improve upon $\mu(f)$ by deviating. Hence, Q is an ON equilibrium in $\tilde{\varphi}$. ∎

Proposition 4.3 *Let Q be an ordinal Nash equilibrium in the game induced by a random stable mechanism $\tilde{\varphi}$. Then, the unique equilibrium outcome $\tilde{\varphi}[Q]$ is individually rational for the true preferences P.*

Proof. By Proposition 4.1, a single matching is achieved in any equilibrium play of the game. Let us say $\tilde{\varphi}[Q] = \mu$. We will prove that μ is individually rational.

First, by contradiction, assume there exists a worker w such that $w P_w \mu(w)$. Suppose that, instead of acting according to Q_w, w uses $Q'_w = w$ and define $Q' = (Q'_w, Q_{-w})$. By considering every firm unacceptable, w is alone under every matching in $S(Q')$. Hence, $1 = Pr\{\tilde{\varphi}[Q'](w) R_w w\} > Pr\{\tilde{\varphi}[Q](w) R_w w\}$ and Q_w does not stochastically P_w-dominate Q'_w. It follows that Q is not an ON equilibrium.

Now suppose that there is a firm f such that $W^G = Ch_{P_f}(\mu(f) \cap W)$, for $W^G \subsetneq (\mu(f) \cap W)$. Consider Q'_f, an alternative strategy for f, where only the elements of W^G are considered acceptable. We will show that Q_f does not stochastically P_f-dominate Q'_f.

To start, consider the matching μ' such that $\mu'(f) = W^G$ and $\mu'(\hat{f}) = \mu(\hat{f})$, for every $\hat{f} \neq f$. Let $W^B = (\mu(f) \cap W) \backslash W^G$ (note that $W^B \neq \emptyset$) and $Q' = (Q'_f, Q_{-f})$. Now consider the matching market $(F, W \backslash W^B, Q'^R)$, where Q'^R is the profile as Q', but restricted to $W \backslash W^B$. We will prove that μ' is stable for Q'^R in this reduced market. Clearly, since μ is individually rational for Q, μ' is individually rational for Q'^R. Now suppose that $(\hat{f}, w)$ blocks μ', *i.e.*, $\hat{f} {Q'^R}_w \mu'(w)$ and $w Q'^R_{\hat{f}} \sigma$, with $\sigma \in \mu'(\hat{f})$. Since only the elements of $\mu'(f)$ are considered acceptable in Q'^R_f, we must have $\hat{f} \neq f$. Hence, ${Q'^R}_{\hat{f}} = Q^R_{\hat{f}}$. By definition of μ', we have $\mu'(\hat{f}) = \mu(\hat{f})$, for every $\hat{f} \neq f$, and $\mu'(w) = \mu(w)$, for every $w \in W \backslash W^B$. This implies that $\hat{f} Q^R_w \mu(w)$ and $w Q^R_{\hat{f}} \sigma$, with $\sigma \in \mu(\hat{f})$; hence, in the unrestricted market, $\hat{f} Q_w \mu(w)$ and $w Q_f \sigma$, with $\sigma \in \mu(\hat{f})$ and $(\hat{f}, w)$ block μ under Q, contradicting $\mu \in S(Q)$. Thus, μ' is stable in $(F, W \backslash W^B, Q'^R)$. Note that, since f is matched to W^G under a stable matching, it must hold exactly W^G under the firm-optimal stable matching for $(F, W \backslash W^B, Q'^R)$, by definition of ${Q'^R}_f$ and of the firm-optimal stable matching.

Suppose W^B join in. By Theorem 5.35 in Roth and Sotomayor (1990), every firm must be at least as well off in the new firm-optimal stable matching. Since only W^G are considered acceptable by f in the strategy Q'_f, f cannot improve upon W^G. Thus, it must be matched to W^G under the firm-optimal stable matching of the market (F, W, Q').

Finally, notice that, by definition of matching, $q_f = |\mu(f)|$. Since $W^B \neq \emptyset$, we have $q_f > |W^G|$. Hence, Theorem 5.13 in Roth and Sotomayor (1990) guarantees that f must hold the same workers under every stable matching in (F, W, Q'). Therefore, by deviating and acting according to Q'_f, f will get W^G with probability one instead of $\mu(f)$. Concluding, Q_f does not stochastically P_f-dominate Q'_f. ■

The above result is as uninformative as large the set of individually rational matchings may be. Ma (2002) shows that one way to make a sharper prediction of equilibrium outcomes and guarantee stability is to go as far as refining the notion of Nash equilibrium to strong Nash and require the use of a particular kind of strategies: truncations at the match point (*i.e.*, deleting the $(m+1)$th and less preferred partners when matched to the mth choice). We provide a different necessary condition for stability in the game induced by a random stable mechanism $\tilde{\varphi}$: every ordinal Nash equilibrium where firms behave straightforwardly is stable for the true preferences. Truth telling by firms is natural in markets where firms obey some kind of objective criterion to fill their positions (*e.g.*, universities admit students on the basis of examination scores, student placement mechanisms assign students to public schools according to the area of residence, firms hire workers according to scores given by recruiting agencies). Moreover, in situations involving uncertainty agents may have no clue about the form that effective strategies might have and straightforward behavior is always an easy resort.

Proposition 4.4 *Let $Q = (P_F, Q_W)$ be an ordinal Nash equilibrium in the game induced by a random stable mechanism $\tilde{\varphi}$. Then, the unique equilibrium outcome $\tilde{\varphi}[Q]$ is stable for the true preferences P.*

Proof. By Proposition 4.1, a unique matching is achieved as the outcome of an ordinal Nash equilibrium. Let us say that $\tilde{\varphi}[Q] = \mu$. By Proposition 4.3, $\mu \in IR(P)$. We will prove that $\mu \in S(P)$ by contradiction. Hence, suppose that (f, w) blocks μ when the true preferences are considered, *i.e.*, $f P_w \mu(w)$ and $w P_f \sigma$, for some $\sigma \in \mu(f)$. Consider the alternative strategy for w given by $Q'_w = f$ and define $Q' = (Q'_w, Q_{-w})$. We will prove that w is matched to f under every matching in $S(Q')$.

Let $\bar{\mu}$ be the matching that corresponds to μ in the related marriage market and let f_i denote the position of firm f that holds σ under $\bar{\mu}$. By Roth (1984a), under every stable matching for Q', w is either always unmatched or always matched to (possibly different) positions in firm f, the only positions he finds acceptable. Let us assume that w is unmatched. This implies that every position of firm f, in particular f_i, is matched to a worker better than w under every matching in $S(Q')$, in particular under the worker-optimal stable matching $\bar{\mu}_W[Q']$. Thus, $\bar{\mu}_W[Q'](f_i) P_{fi} w$. Since $w P_{f_i} \sigma$ and, by definition of worker-optimal stable matching, $\sigma R_{fi} \bar{\mu}_W[Q](f_i)$, we have $\bar{\mu}_W[Q'](f_i) P_{fi} \bar{\mu}_W[Q](f_i)$. Nevertheless, $\bar{\mu}_W[Q']$ is the worker-optimal stable matching in the reduced market $(F, W \backslash \{w\}, Q^R)$, with Q^R representing the same orderings of preferences as in Q, but restricted to $W \backslash \{w\}$. This contradicts Theorem 2.25 in Roth and Sotomayor (1990) since, under the worker-optimal stable matching, no firm can be matched to a better worker in the restricted market. Therefore, w must be matched to a position of firm f under every element of $S(Q')$.

In conclusion, by acting in accordance with Q'_w, w will be matched to f with probability one. Hence, Q_w does not stochastically P_w-dominate Q'_w and we have a contradiction. ■

Two remarks are in order. First, this result can easily be applied to games arising from deterministic stable mechanisms. Hence, stability for the true preferences is obtained in any Nash equilibrium where firms are truthful for any stable matching mechanism. Second, in accordance with the claims in Roth and Sotomayor

(1990) concerning deterministic mechanisms, the analogous result with workers telling the truth and firms acting strategically does not hold, although it would hold when all quotas equal one.[21] The college admissions problem, unlike the marriage problem, is not symmetric between the two sides of the market and there are substantial differences between the two when strategic issues are contemplated. Any firm with a quota greater than one resembles something like a coalition rather than an individual. Hence, allowing for manipulation on the firms side is similar to giving such powers to sets of agents in a marriage market and, in equilibria where workers tell the truth, stability is lost.

The converse result is given in Proposition 4.5, asserting that every stable matching for the true preferences can be supported as the outcome of an ordinal Nash equilibrium where firms act according to the true preferences. In fact, workers can compel any jointly achievable outcome in the game induced by a random stable mechanism, while firms behave straightforwardly.

Proposition 4.5 *Let μ be any stable matching for (F, W, P) and let $\tilde{\varphi}$ be a random stable mechanism. Then, there exists an ordinal Nash equilibrium $Q = (P_F, Q_W)$ that supports μ in the game induced by $\tilde{\varphi}$.*

Proof. Define $Q_w = \mu(w)$, for every $w \in W$. Clearly, $S(Q) = \{\mu\}$ and μ is reached with probability one.

Let us now prove that Q is an ON equilibrium. Take an arbitrary worker w and suppose that there exists a firm f such that $f P_w \mu(w)$. We claim that w cannot deviate to get matched to f. In fact, the stability of μ with respect to P implies that $\sigma P_f w$, for every $\sigma \in \mu(f)$. Since $\mu(\sigma) = f$ we have $Q(\sigma) = f$, for every $\sigma \in \mu(f)$, and f will end up matched to $\mu(f)$. Now consider firm f. The only workers to accept f are those in $\mu(f)$. Furthermore, individual rationality of μ implies that $\mu(f) \bar{R}_f W^S$, for every $W^S \subseteq \mu(f)$. It follows that f cannot improve upon $\mu(f)$ by deviating. In conclusion, Q is an ON equilibrium in $\tilde{\varphi}$. ■

Our next results establish a strong link between equilibria in games induced by random and by deterministic stable mechanisms. We start by pointing out that every ordinal Nash equilibria of the random process must be a simple Nash equilibrium of some mechanism where chance plays no role.

Proposition 4.6 *Let Q be an ordinal Nash equilibrium in the game induced by a random stable mechanism $\tilde{\varphi}$. Then, there exists a stable mechanism φ where Q is a Nash equilibrium.*

[21] See Roth (1985).

Proof. Assume that Q is an equilibrium that yields μ in the game induced by $\tilde{\varphi}$. Proposition 4.1 guarantees that μ is the only element in $supp\tilde{\varphi}[Q]$ and, by Proposition 4.3, $\mu \in IR(P)$. Now suppose, by contradiction, that there exists no stable matching mechanism where Q is a Nash equilibrium. In particular, consider any φ such that $\varphi[Q] = \mu$—such a mechanism exists since $\mu \in S(Q)$—and assume that some agent has a profitable deviation.

Let such agent be a worker, w. Then, there exists a strategy Q'_w such that $\varphi[Q'](w)P_w\mu(w)$, with $Q' = (Q'_w, Q_{-w})$. This implies that $\varphi[Q'](w) \in F$ since $\mu \in IR(P)$. Define $f = \varphi[Q'](w)$ and $Q''_w = f$. Observe that under any matching in $S(Q''_w, Q_{-w})$, w is matched to f ($\varphi[Q']$ is stable for (Q''_w, Q_{-w}) since it remains individually rational and no blocking pairs emerge once w uses Q''_w). Therefore, under every matching in $supp\tilde{\varphi}[(Q''_w, Q_{-w})]$, w holds f, and Q_w does not stochastically P_w-dominate Q''_w. We get a contradiction: Q is not an ON equilibrium in the game induced by $\tilde{\varphi}$.

Now assume that $f \in F$ can profit by deviating from Q_f. This means that there exists Q'_f such that $\varphi[Q'](f)\bar{P}_f\mu(f)$, with $Q' = (Q'_f, Q_{-f})$. Since $\mu \in IR(P)$, $\varphi[Q'](f) \cap W \neq \varnothing$. Define Q''_f such that $A(Q''_f) = \varphi[Q'](f) \cap W$. Since $\varphi[Q'] \in S(Q')$, once only the workers in $\varphi[Q'](f)$ are considered acceptable by f, we can guarantee that $\varphi[Q'] \in S(Q'')$. The definition of Q''_f and the fact that under every stable matching firms have the same number of positions filled (Theorem 5.12 in Roth and Sotomayor (1990)) imply that f holds $\varphi[Q'](f)$ in every element of $S(Q'')$. Therefore, $1 = Pr\{\tilde{\varphi}[Q''](f)R_f\varphi[Q'](f)\} > Pr\{\tilde{\varphi}[Q](f)R_f\varphi[Q'](f)\} = 0$ and Q is not an ON equilibrium in the game induced by $\tilde{\varphi}$. ■

In Proposition 4.7, we establish a partially converse statement: the set of ordinal Nash equilibria in a random stable mechanism includes all the strategy profiles that are simultaneously equilibria in the mechanisms that yield the firm-optimal and the worker-optimal stable matchings.

Proposition 4.7 *Let Q be a Nash equilibrium in the games induced by φ^F and by φ^W. Then, Q is an ordinal Nash equilibrium in the game induced by a random stable mechanism $\tilde{\varphi}$.*

The following Lemma is useful in proving Proposition 4.7.

Lemma 4.2 *Let Q be a Nash equilibrium in the games induced by φ^F and by φ^W. Then, the set $S(Q)$ is a singleton.*

Proof. Assume that Q is a Nash equilibrium in the games induced by both φ^F and φ^W. Suppose, by contradiction, that $|S(Q)| \geq 2$. Clearly, this implies that $\varphi^F[Q] \neq \varphi^W[Q]$. Lemma 4.1 in Ma (2002) implies that, for any matching $\mu \in S(Q)$, we have $\varphi^F[Q](w) R_w \mu(w)$, for every $w \in W$. Since Q is an equilibrium in φ^W, the same lemma guarantees that $\varphi^W[Q](w) R_w \mu(w)$, for every $w \in W$ and for any $\mu \in S(Q)$. It follows that $\varphi^F[Q](w) = \varphi^W[Q](w)$, for every $w \in W$ and we contradict the initial assumption that $\varphi^F[Q] \neq \varphi^W[Q]$. ■

Proof. [Proof of Proposition 4.7] Suppose that Q is a Nash equilibrium in the games induced by both φ^F and φ^W. By Lemma 4.2, $|S(Q)| = 1$. Let us say that $S(Q) = \{\hat{\mu}\}$ and assume, by contradiction, that Q is not an ON equilibrium in the game induced by $\tilde{\varphi}$.

Suppose then that there exists a worker $w \in W$ and an alternative strategy Q'_w such that Q_w does not stochastically P_w-dominate Q'_w. Since $\tilde{\varphi}[Q] = \hat{\mu}$, this implies that $Pr\{\tilde{\varphi}[Q'_w, Q_{-w}]\ (w) P_w \hat{\mu}(w)\} > 0$. Note that, since Q is a Nash equilibrium in a stable mechanism, $\hat{\mu} \in IR(P)$. Hence, $\hat{\mu}(w) R_w w$ and it must be the case that w is matched to a firm under every matching in $S(Q'_w, Q_{-w})$. Let μ' be such that $\mu'(w) P_w \mu(w)$, for every $\mu \in supp\tilde{\varphi}[Q'_w, Q_{-w}]\}$ and define $Q''_w = \mu'(w)$. Since $\mu' \in S(Q''_w, Q_{-w})$ (it is still individually rational and no blocking pairs emerged), Theorem 5.12 in Roth and Sotomayor (1990) ensures that w is matched to f under every matching in $S(Q''_f, Q_{-f})$. Then, there exists no stable matching mechanism where Q is a Nash equilibrium, since for every stable matching mechanism φ, $\varphi[Q''](w) = f$ and $f P_w \hat{\mu}(w)$. It follows that no worker can profitably deviate in the game induced by $\tilde{\varphi}$.

Hence, there exists a firm f and a strategy Q'_f such that Q_f does not stochastically P_f-dominate Q'_f, *i.e.*, we have $Pr\{\tilde{\varphi}[Q'_f, Q_{-f}](f) \bar{P}_f \hat{\mu}(f)\} > 0$. Since $\hat{\mu} \in IR(P)$, $\hat{\mu}(f) \bar{R}_f \varnothing$ and under all matchings in $S(Q'_f, Q_{-f})$, f has, at least, one position filled. Let μ' be such that $\mu'(f) \bar{P}_f \mu(f)$, for every $\mu \in supp\tilde{\varphi}[Q'_f, Q_{-f}]\}$. Define Q''_f such that $A(Q''_f) = \mu'(f) \cap W$. Note that $\mu' \in IR(Q''_f, Q_{-f})$ and that no pair of agents blocks μ' under the preference profile (Q''_f, Q_{-f}). Therefore, $\mu' \in S(Q''_f, Q_{-f})$ and, since firms have the same positions filled under every stable matching (Theorem 5.12 in Roth and Sotomayor (1990)), the definition of Q''_f guarantees that f holds $\mu'(f)$ in every element of $S(Q''_f, Q_{-f})$. Finally, for every stable matching mechanism φ, $\varphi[Q''_f, Q_{-f}](f) = \mu'(f)$ and $\mu'(f) \bar{P}_f \hat{\mu}(f)$. It follows that there exists no stable mechanism where Q is a Nash equilibrium. ■

The proof of the above result reveals that a sufficient condition for an ordinal Nash equilibrium is in fact being a Nash equilibrium in every deterministic stable mechanism. This appears to be an extremely strong condition to fulfill. Nevertheless, we will now describe a class of random mechanisms for which such condition becomes necessary for an ordinal Nash equilibrium.

In the particular case that μ^I is the empty matching, Roth and Vande Vate (1990) have shown that, in the marriage model, every element of the stable set for the revealed preferences can be achieved with positive probability when the random mechanism they define is applied. In fact, starting from a situation in which all agents are unmatched, by successively satisfying all the pairs of a stable matching, we can guarantee that this matching is reached with positive probability. This random process is an instance of what we will name as *really* random stable mechanisms.

A *really random stable mechanism* $\tilde{\varphi}$ assigns positive probability to at least two different elements of the set of stable matchings, *i.e.*, $|supp\tilde{\varphi}[Q]| \geq 2$ for every Q such that $|S(Q)| \geq 2$. In Example 4.1, the mechanism that assigns probability 0.5 to the firm-optimal and to the worker-optimal stable matchings is clearly a really random stable mechanism. The following result is an implication of Propositions 4.6 and 4.7 in the particular case that $\tilde{\varphi}$ is really random.

Corollary 4.1 *Let $\tilde{\varphi}$ be a really random stable mechanism. Then, the profile of strategies Q is an ordinal Nash equilibrium in the game induced by $\tilde{\varphi}$ if and only if the set of stable matchings $S(Q)$ is a singleton and there exists a stable mechanism φ where Q is a Nash equilibrium.*

Proof. Follows directly from Propositions 4.6 and 4.7, and the fact that Proposition 4.1 implies $supp\tilde{\varphi}[Q] = S(Q)$ for a really random stable mechanism $\tilde{\varphi}$. ■

For illustration, consider once more Example 4.1 and note that the set of stable matchings for truthtelling is a singleton; further, it can easily be shown that it is an equilibrium in the mechanism that yields, say, the firm-optimal stable matching. Corollary 4.1 thus implies that straightforward behavior is an ordinal Nash equilibrium in the random stable mechanism described in the example.

4.5 Non-Preference Strategies

We have explored the game induced by a random mechanism, claiming that one of the main motivations of this paper is the study of some decentralized markets. This may be objected on the grounds that up to this point we have restricted our analysis to a one-period game where strategies are preference lists, which perfectly mirrors the functioning of a centralized market, but falls short of an illustration of a decentralized market. In particular, in matching processes of the kind described by Roth and Vande Vate (1990), at each moment in time, a pair of randomly chosen agents meets and (temporarily) matches if this is consistent with both agents' strategies.

This clearly fits the structure of a sequential game. In this context, restricting each agent to hold the potential partner that is higher on some fixed preference ordering sustains the validity of the results of the preceding section. However, in a sequential game, agents can be expected to use richer strategies, conditioning behavior on the history of the game, and not necessarily acting consistently with a unique preference ordering. The strategy of matching with the first partner one meets and rejecting every other agent is an example of such kind of strategies.

One of the difficulties that arises in attempting to capture such complex forms of behavior concerns the very essence of the mechanism that, following Roth and Vande Vate (1990), we assume to be stable with respect to the revealed preferences. In fact, such definition is compromised when, for some play of the game, no list of preferences is compatible with the strategy of a player. Hence, the set of feasible strategies of the sequential game is simply too large and precludes analysis in the theoretical framework we have been using. One potential course of action is therefore to impose that under any play of the sequential game the choices actually made are consistent with some preference ordering, even though they may correspond to incompatible preference orderings when several plays are considered. We can then speak of preference orderings that are "revealed" in the course of the play. A worker w that entertains the described strategy in the example above, would match the first firm to tender an offer to him under any play of the game, and reveal that this firm is preferred to every other firm that he eventually meets in the course of that play. Since meeting is random, this worker would reveal distinct preference lists under different plays of the game.

Hence, consider a sequential game where, starting from an arbitrary matching, at each moment in time, a pair of randomly chosen agents, composed of a firm and a worker, meets. Agents match upon meeting if this is consistent with their strategies. We assume that strategies are restricted to those compatible with a preference ordering for each play of the game, the revealed preference ordering, even though the information gathered in the course of the play might allow for other forms of behavior.[22] According to Roth and Vande Vate (1990), once the probability that a given pair of agents meets is bounded away from zero, each play of the game yields a matching stable with respect to the revealed orderings in the course of that play.

In Proposition 4.8, we show that the ordinal Nash equilibria in preference strategies obtained for the one-period game, characterized in Section 4, are robust to the enlarged strategy space. In fact, given a profile of preference strategies, if by means of a strategy that is not consistent with a unique preference ordering, an

[22] The lack of precision in defining what each player knows along the game is deliberate. The result that follows is valid in a perfect information setting, as well as when agents are only partially aware of the history of the game.

agent may improve his position, he is certainly capable of doing so using a simple preference strategy.

Proposition 4.8 *In the sequential game induced by a random stable mechanism $\tilde{\varphi}$, for any collection of stated preferences Q_{-v} for agents other than an arbitrary agent v, agent v always has a best response that is consistent with a unique preference ordering.*

Proof. First, consider an arbitrary worker w and fix Q_{-w}. Let s_w denote an arbitrary strategy for w, revealing a preference ordering (not necessarily the same) under each play of the game. Denote by Q_w^i the preference ordering that is consistent with s_w under some play i. In general, we have $supp\tilde{\varphi}[s_w, Q_{-w}] = \{\mu_1, ..., \mu_k\}$, where $\mu_i \in S(Q_w^i, Q_{-w})$, for $i = 1, ..., k$. Now let $Q_w = \mu_j(w)$ where $\mu_j(w) R_w \mu_i(w)$, for all $\mu_i \in \{\mu_1, ...\mu_k\}$. Since $\mu_j \in S(Q_w^j, Q_{-w})$, we must have $\mu_j \in S(Q_w, Q_{-w})$ (it is still individually rational and there are fewer blocking pairs). Hence, given that the same agents are matched under any two elements of the stable set and the only firm w finds acceptable is $\mu_j(w)$, this worker is matched to $\mu_j(w)$ under every matching in $S(Q_w, Q_{-w})$. It follows that any lottery over $S(Q_w, Q_{-w})$ gives w a partner at least as good as any lottery over $S(s_w, Q_{-w})$. Since s_w and Q_{-w} are arbitrary, this completes the proof for a worker w.

Now take an arbitrary firm f. Let s_f denote a strategy for f with the same properties as the strategy for w above. Define Q_f^i as the preference ordering over individual workers that is consistent with s_f for some play i of the game. Let $supp\tilde{\varphi}[s_f, Q_{-f}] = \{\mu_1, ..., \mu_k\}$, where $\mu_i \in S(Q_f^i, Q_{-f})$, for $i = 1, ..., k$. Consider any alternative strategy Q_f for f such that $A(Q_f) = \mu_j(f) \cap W$ where $\mu_j(f) \bar{R}_f \mu_i(f)$, for all $\mu_i \in \{\mu_1, ..., \mu_k\}$ and for every responsive extension $\bar{R}_f$ of R_f. Then, $\mu_j \in IR(Q_f, Q_{-f})$ since $\mu_j \in IR(Q_f^j, Q_{-f})$. Moreover, $\mu_j \in S(Q_f, Q_{-f})$ since $\mu_j \in S(Q_f^j, Q_{-f})$ and no blocking pairs emerged. Given that the same positions of a firm are filled under any element of a stable set and by definition of Q_f, f is matched to $\mu_j(f)$ under every matching in $S(Q_f, Q_{-f})$. Since s_f and Q_{-f} are arbitrary, this completes the proof. ■

Nevertheless, this is far from being a characterization of equilibria in this new setting. In fact, the set of ordinal Nash equilibria is larger here, as the following example demonstrates.

Consider the matching market in Example 4.1. Let the strategy of each agent be defined as follows: s_f = "match only with w_i if f_1 is the first firm to meet a worker; match only with w_j otherwise" and s_{w_i} = "match only with f_i if f_1 is the first firm to meet a worker; match only with f_j otherwise," for $i = 1, 2$. This strategy profile leads to a non-degenerate probability distribution over matchings. Namely, both $\mu = \{(f_1, w_1), (f_2, w_2)\}$ and $\hat{\mu} = \{(f_1, w_2), (f_2, w_1)\}$ are obtained with a 50% probability. Hence, Proposition 4.1 rules out the possibility that s can be reproduced by an equilibrium in preference strategies. Still, s is an ordinal Nash equilibrium, since any unilateral deviation of a firm or worker may either leave the probability distribution unchanged or leave the deviator unmatched with positive probability. $\diamond$

4.6 Concluding Remarks

At the expense of using an ordinal equilibrium concept, we have provided a characterization of equilibria that arise in the game induced by a random stable mechanism. The analysis is set in the college admissions problem. First, we have proved that every ordinal Nash equilibrium yields a unique matching, while when agents act straightforwardly according to the true preferences several matchings may be obtained with positive probability. Hence, agents avoid uncertainty when behaving strategically. Furthermore, a matching can be reached at an ordinal Nash equilibrium if and only if it is individually rational for the true preferences. Ordinal equilibria where firms best reply by behaving straightforwardly always produce a matching stable for the true preferences. Conversely, every stable matching can be reached as the outcome of an equilibrium play of the game. In a different direction, we relate ordinal Nash equilibria induced by a random mechanism with Nash equilibria arising in the games induced by deterministic mechanisms. In particular, a preference profile is an ordinal equilibrium of the game induced by a mechanism that always assigns positive probability to two different matchings (if such matchings exists) if and only if the set of stable matchings is a singleton and it is a Nash equilibrium in some deterministic stable mechanism. In the last section of the paper we have tried to extend the above results, derived for a one-period game where the set of available strategies coincides with the set of all possible lists of preferences, to the sequential game that may arise in a decentralized market. Here we assume agents may use strategies that correspond to different preference orderings when different plays of the game are considered. We have shown that ordinal Nash equilibria in preference strategies are robust to the enlarged strategy space.

In what the above results are concerned, a couple of remarks is in order. The first observation concerns fairness and random matching mechanisms. In opposition to deterministic mechanisms, which are bound to favor one side of the market over the other, we have claimed that random matching mechanisms promote procedural fairness.[23] Nevertheless, “endstate” justice is a different issue. Indeed,

[23] For example, in the kind of process described in Roth and Vande Vate (1990), each pair of

the results that relate equilibria in the games induced by random and deterministic mechanisms imply that every equilibrium outcome in a random mechanism may be obtained by means of a deterministic mechanism. It follows that, based on these results and in what "endstate" justice is concerned, we should not expect random matching mechanisms to improve upon deterministic ones if equilibrium behavior is to be taken seriously.

Second, the aim of the last section is to shed some light on what happens once we move towards allowing for history-dependent strategies, preserving the stability of the mechanism. The purpose of this paper is to explore strategic behavior induced by random stable mechanisms, and not to provide a thorough analysis of the incentives agents face in decentralized markets. Therefore, relaxing the restriction we impose over the strategy sets would compromise our main goal.

To conclude, equilibrium behavior in random mechanisms has barely been treated in the matching literature. One of the difficulties that arises in attempting to apply the common game theoretical tools stems from the need to compare the probability distributions over matchings generated by a random mechanism when preferences are ordinal. By means of the concept of ordinal Nash equilibrium we have taken a step towards filling the gap in the literature, providing a fairly complete characterization of equilibrium behavior.

---

agents has the same probability of meeting at a certain point in the procedure, and this determines procedural fairness.

Chapter 5
Giving Advice and Perfect Equilibria in Matching Markets

5.1 Introduction

There is a vast literature on two-sided matching markets. Theoretical investigations in matching exhaust issues on the existence of stable matchings, the structure of the set of such outcomes, and computational algorithms designed to reach them. The strategic decisions that confront individuals under matching mechanisms have also been broadly inspected, focusing particularly on incentives in stable matching mechanisms. That every individually rational matching can be reached as the outcome of an equilibrium play in the game induced by a stable mechanism is a well-known fact (Alcalde, 1996). Nevertheless, agents are, in general, poorly informed and this casts some doubts on the significance of the statement. Indeed, a great deal of information about the preferences of the other agents may be needed to compute an equilibrium; furthermore, the multiplicity of equilibria entails a lot of coordination among agents. Attention is then devoted to a more reasonable class of equilibria, narrowing the set of probable outcomes. In the mechanism that yields the optimal stable matching for one side of the market, Roth (1984b) showed that, although agents may have an incentive to misrepresent their preferences, every equilibrium in undominated strategies produces a matching that is stable with respect to the true preferences.

The purpose of this paper is to take this analysis further aiming at a characterization of perfect equilibria in markets organized to produce stable outcomes. Ordinal preferences entail the use of a perfect equilibrium concept with an ordinal flavor. In fact, in an ordinal perfect equilibrium agents play best replies to particular profiles of completely mixed strategies. A best reply, in this context, first-order stochastically dominates every alternative strategy against the mixed strategy profile being considered. Surprisingly, in the mechanism that induces the optimal stable matching for one side of the market, truth telling emerges as the unique ordinal perfect equilibrium. Hence, if acting straightforwardly is, in fact, an ordinal perfect equilibrium, we may postulate that the unique stable matching for the true preferences is the outcome of the game.

Nevertheless, only seldom is truth a Nash equilibrium in the game induced by the optimal stable mechanism for one side of the market. We can thus anticipate that the existence of ordinal perfect equilibria is exceptional. In reality, a necessary requirement for honesty to be an ordinal perfect equilibrium is being dominant for every agent. Hence, the set of ordinal perfect equilibria and Nash equilibria in dominant strategies coincide in these markets.

Still, the described results may be seen from a brighter perspective. Provided agents are poorly informed, truth telling may be prescribed as a very prudent form of behavior. In the complete information framework, Gale and Sotomayor (1985) have proved that, when any stable mechanism is in use, at least one agent can profitably misrepresent its preferences, except when there is a unique stable outcome. Yet, in order for participants to identify some strategies that perform better than truth telling, a lot of information about others' revealed preferences is needed. In the game induced by the mechanism that yields the firm-optimal stable matching, when each agent has certain beliefs about others' strategies, it is still a dominant strategy for each firm to act straightforwardly (Roth, 1989). On the other hand, Roth and Rothblum (1999) have shown that if workers do not have detailed information about the preferences revealed by other agents in the course of play, the scope of potentially profitable strategic behavior is significantly reduced, if we compare it with the complete information case. If such information exhibits a certain kind of symmetry, reversing the true order of two acceptable firms is to be considered imprudent behavior, but submitting a truncation of the true preferences may be beneficial. Informally, a truncation is a preference ordering that is order-consistent with the true preferences, but under which the worker restricts the number of firms he applies to. Ehlers (2004) takes a further step in the search of advice for workers in a matching market, providing a weaker condition on a worker's beliefs to obtain the conclusions of Roth and Rothblum (1999). Loosely speaking, a worker should not reverse the true ranking of two acceptable firms whenever he is not able to anticipate which new proposals he is going to receive after having rejected others. Moreover, Ehlers (2004) gives advice to workers who can distinguish between three sets of firms: the firms that will certainly propose to him, the firms that may propose, and those from which he does not expect a proposal.

Hence, there seems to be a clear consensus about how harmful altering the true order of firms may be in a low information environment. The results in this paper suggest that, when a worker contemplates obtaining a proposal from any acceptable firm, he should reveal his whole true preference ordering if he wants to minimize the probability of being unmatched. In fact, truncations may lead to more favorable outcomes, but at the expense of increasing the chances of being alone. Regardless of the incentives to act strategically, honesty thus remains a fundamental form of behavior.

These conclusions already stem from Barberà and Dutta (1995). Barberà and Dutta (1995) show that acting straightforwardly is the unique *protective strategy* for every agent. Loosely speaking, this means that when an agent compares truth telling with any misrepresentation of its preferences, there exists a potential partner with whom, by manipulating, it ends up matched for a larger set of actions of the other players, while less preferred potential partners are obtained, by either acting straightforwardly or strategically, against the same profiles for the rest of society. The concept of protective behavior is based on a refinement of a maxmin criterion and is particularly appropriate for games where agents are poorly informed and sufficiently risk averse.

We proceed as follows. In Section 5.2, we formally present the marriage model and introduce notation. We define the concept of ordinal perfect equilibrium in Section 5.3. In Section 5.4 we develop the main results. We conclude in Section 5.5 on further research.

5.2 The Marriage Model

Consider two finite and disjoint sets $F = \{f_1, ..., f_n\}$ and $W = \{w_1, ..., w_p\}$, where F is the set of firms and W is the set of workers. We let $V = W \cup F$ and sometimes refer to a generic agent by v, while w and f represent a generic worker and firm, respectively. Each agent has a strict, complete, and transitive preference relation over the agents on the other side of the market and the perspective of being unmatched. The preferences of a firm f, for example, can be represented by $P_f = w_3, w_1, f, w_2, ..., w_4$, indicating that f's first choice is to be matched to w_3, its second choice is w_1 and it prefers remaining unmatched to being assigned to any other worker. Equivalently, we may say that w_3 is the lowest ranked worker in P_f, with rank 1 ($r_{P_f}(w_3) = 1$), w_1 is ranked second ($r_{P_f}(w_1) = 2$), being unmatched is ranked third ($r_{P_f}(f) = 3$), and every other worker has a higher ranking in P_w. A worker is *acceptable* if the firm ranks him lower than having its position unfilled; in the above example, the set of acceptable workers is $A(P_f) = \{w_1, w_3\}$. Similarly, given P_w we may define an acceptable firm and $A(P_w)$. It is sufficient to describe only the ordering of acceptable partners, so that the in the above example preferences can be abbreviated as $P_f = w_3, w_1$. Let $P = (P_{f_1}, ..., P_{f_n}, P_{w_1}, ..., P_{w_p})$ denote the profile of all agents' preferences; we sometimes write it as $P = (P_v, P_{-v})$ where P_{-v} is the set of preferences of all agents other than v. Further, we may use P_U, where $U \subseteq V$, to denote the profile of preferences $(P_v)_{v \in U}$. We write $v' P_v v''$ when v' is preferred to v'' under preferences P_v and we say that v *prefers* v' to v''. We write $v' R_v v''$, when v likes v' *at least as well as* v'' (it may be the case that v' and v'' are the same agent).

Formally, a *marriage market* is a triple (F, W, P). An outcome for a marriage market, a *matching*, is a function $\mu : V \longrightarrow V$ satisfying the following: *(i)* for each f in F and for each w in W, $\mu(f) = w$ if and only if $\mu(w) = f$; *(ii)* if $\mu(f) \neq f$ then $\mu(f) \in W$; *(iii)* if $\mu(w) \neq w$ then $\mu(w) \in F$. If $\mu(v) = v$, then v is *unmatched* under μ, while if $\mu(w) = f$, we say that f and w are *matched* to one another. A description of a matching is given by $\mu = \{(f_1, w_2), (f_2, w_3)\}$, indicating that f_1 is matched to w_2, f_2 is matched to w_3 and the remaining agents in the market are unmatched. A matching μ is *individually rational* if each agent is acceptable to its partner, *i.e.*, $\mu(v) R_v v$, for all $v \in V$. We denote the set of all individually rational matchings by $IR(P)$. Two agents f and w form a *blocking pair* for μ if they prefer each other to the agents they are actually assigned to under μ, *i.e.*, $f P_w \mu(w)$ and $w P_f \mu(f)$. A matching μ is *stable* if it is individually rational and it is not blocked by any pair of agents. We denote the set of all stable matchings by $S(P)$. A firm f and a worker w are *achievable* for each other if f and w are matched under some stable matching.

The proof of existence of stable matchings in Gale and Shapley (1962) is constructed by means of the deferred-acceptance algorithm. At each step of the algorithm, proposals are issued by one side of the market according to its preferences, while the other side merely reacts to such offers by rejecting all but the best. Hence, in the case that firms make job offers, the algorithm starts with each firm proposing to the first worker on its list and each worker rejecting all proposals but the best. This yields the first tentative matching. Next, every rejected firm makes an offer to its second favorite worker and again workers only hold the one they prefer among those just received and the one held from the previous step. The algorithm proceeds by creating, at each step, a tentative matching and terminates when each firm is either held by a worker or has been rejected by every worker on its list of preferences. This algorithm arrives at the firm-optimal stable matching, with the property that all firms are in agreement that it is the best stable matching. The deferred-acceptance algorithm with workers proposing produces the worker-optimal stable matching with corresponding properties. Further, the optimal stable matching for one side of the market is the worst stable matching for every agent on the other side of the market, a result presented in Knuth (1976) but attributed to John Conway. Still, there is a set of agents who are indifferent between any stable matching. The first statement of this result appears in McVitie and Wilson (1970); later, it was proved in Roth (1984b) and Gale and Sotomayor (1985). We state it formally in the next Proposition for further reference.

Proposition 5.1 *In a matching market (F, W, P), the set of unmatched agents is the same for all stable matchings.*

Finally, a *matching mechanism* $\tilde{\varphi}$ maps preference profiles into lotteries over matchings. In what follows, in a matching market (F, W, P), we consider the revelation game induced by $\tilde{\varphi}$ in which agents are each faced with the decision of what strategies to act on. The strategy space of a player in the game is the set of all possible preference lists: given the true preference ordering P_v, each player v may eventually reveal a different order Q_v over the players on the other side of the market. A matching mechanism $\tilde{\varphi}$ and a preference profile Q induce a random matching $\tilde{\varphi}[Q]$. Throughout the paper, we only consider *stable* matching mechanisms. Hence, $\tilde{\varphi}[Q]$ denotes the probability distribution induced over the set of stable matchings $S(Q)$ and $\tilde{\varphi}[Q](v)$ is the probability distribution induced over agent v's achievable matches. We use, for example, $\Pr\{\tilde{\varphi}[Q](v)R_v\hat{v}\}$ to denote the probability that v obtains a partner at least as good as $\hat{v}$ according to v's true preferences R_v when the profile Q is used in the mechanism $\tilde{\varphi}$. In the particular case that the mechanism is deterministic, we let $\tilde{\varphi}[Q]$ denote the unique outcome matching. The mechanism that yields the firm-optimal stable matching with certainty is an example of a deterministic stable matching mechanism and will be denoted by φ^F. We let φ^W represent the mechanism that leads to the worker-optimal stable matching.

5.3 Ordinal Perfect Equilibria

In this section we define ordinal perfect equilibria. We present all definitions for stable mechanisms in general, even though many results refer to the particular case of deterministic mechanisms, namely the mechanisms that yield the optimal stable matching for one side of the market.

Consider $w \in W$ (what follows also holds for a representative firm, with obvious modifications) with true preferences P_w and let Q_{-w} be a strategy profile for all the agents other than w. Given a stable mechanism $\tilde{\varphi}$ and given Q_{-w}, we say that the strategy Q_w *stochastically* P_w*-dominates* Q'_w if, for all $v \in F \cup \{w\}$, $\Pr\{\tilde{\varphi}[Q_w, Q_{-w}](w)R_w v\} \geq \Pr\{\tilde{\varphi}[Q'_w, Q_{-w}](w)R_w v\}$. Thus, for all $v \in F \cup \{w\}$, the probability of w being assigned to v or to a strictly preferred agent is higher under $\tilde{\varphi}[Q_w, Q_{-w}](w)$ than under $\tilde{\varphi}[Q'_w, Q_{-w}](w)$. Hence, if we consider the problem that player w faces given the strategy choices Q_{-w} of the other players, a particular strategy choice Q_w may be preferred if it stochastically dominates every other alternative strategy. In this case we say that Q_w is a *best reply* to Q_{-w}.

Definition 5.1 *Given the profile of preferences P, the profile of strategies Q is an* ordinal Nash equilibrium *(*ON equilibrium*) in the game induced by $\tilde{\varphi}$ if, for each agent v in V, Q_v is a best reply to Q_{-v}.*

The concept of ordinal Nash equilibrium deserves a couple of remarks. First, it was introduced in d'Aspremont and Peleg (1988) and its use is required given the very nature of random matching.[24] In fact, agents' preferences are ordinal in

[24] This concept was also used in the context of voting theory in Majumdar and Sen (2004) and in matching markets in Ehlers and Massó (2003), and Majumdar (2003).

nature. Since no natural utility representation of these preferences exists (and no expected utilities can be computed), this ordinal criterion provides a means for comparing probability distributions over potential partners. Second, it is quite a strong equilibrium concept. Under an ordinal Nash equilibrium, each agent plays its best response to the others' strategies for every utility representation of the preferences. However, in the particular case that $\tilde{\varphi}$ is a deterministic mechanism, the concept boils down to plain Nash equilibrium.

For our purposes, some of the above definitions have to be extended to mixed strategies. We let σ denote a mixed strategy and we let $\sigma(Q) = \prod_{v \in V} \sigma_v(Q_v)$ be the probability of profile Q under the mixed strategy σ. Given a stable mechanism $\tilde{\varphi}$ and a mixed strategy profile σ, we let $\tilde{\varphi}[\sigma]$ denote the probability distribution induced over the whole set of matchings that satisfies the following: $\Pr\{\tilde{\varphi}[\sigma] = \mu\} = \sum_{Q \in supp\sigma} \sigma(Q) \cdot \Pr\{\tilde{\varphi}[Q] = \mu\}$. As before, given a mixed strategy profile σ_{-w}, the pure strategy Q_w *stochastically* P_w*-dominates* Q'_w if, for all $v \in F \cup \{w\}$, $\Pr\{\tilde{\varphi}[Q_w, \sigma_{-w}](w) R_w v\} \geq \Pr\{\tilde{\varphi}[Q'_w, \sigma_{-w}](w) R_w v\}$. The strategy Q_w is a *best reply* to σ_{-w} if it stochastically P_w-dominates every alternative pure strategy. We are now in condition to define ordinal perfect equilibria.

Definition 5.2 *Given the profile of preferences P, the profile of strategies Q is an* ordinal perfect equilibrium *in pure strategies (*OP equilibrium*) in the game induced by $\tilde{\varphi}$ if there exists a sequence of completely mixed strategies σ^k, $\{\sigma^k\} \xrightarrow[k \to \infty]{} Q$, with the property that, for every $k \geq 1$, Q_v is a best reply to σ^k_{-v}, for every agent v in V.*

Hence, we require that the profile Q be a limit of a sequence of totally mixed profiles σ^k and that Q_v stochastically P_v-dominates every alternative pure strategy when the opponents use the perturbed strategies σ^k_{-v}.

5.4 Results

The first couple of results apply to any stable matching mechanism. In Theorem 5.1, we take a prescriptive point of view and establish that no unacceptable partners should be included in one's list if not matching unacceptable partners is the major concern. Moreover, if an agent wishes to minimize the probability of being unmatched, it should submit a comprehensive preference ordering. In fact, the existence of even the slightest chance of being matched to an acceptable partner should not be neglected.

Theorem 5.1 *Let $\tilde{\varphi}$ be a stable mechanism. If Q_v is agent v's best reply to a completely mixed strategy profile σ_{-v}, then Q_v lists all the partners that are acceptable according to v's true preferences P_v (i.e., $A(Q_v) = A(P_v)$).*

Proof. Let v be an arbitrary worker. Since the model is symmetric between firms and workers, what follows also holds for an arbitrary firm.

First, we will show that ranking an unacceptable firm f' as acceptable in Q_v is not a best reply to a completely mixed strategy profile for agents other than v, σ_{-v}.

(i) Take any Q_{-v} and note that under any matching $\mu \in S(P_v, Q_{-v})$, v is either always unmatched or always matched to an acceptable firm. Hence, $\Pr\{\tilde{\varphi}[P_v, Q_{-v}](v) R_v v\} = 1$, for all Q_{-v}. It follows that $\Pr\{\tilde{\varphi}[P_v, \sigma_{-v}](v) R_v v\} = \sum_{Q_{-v} \in supp \sigma_{-v}} \sigma(Q_{-v}) \cdot \Pr\{\tilde{\varphi}[P_v, Q_{-v}](v)\ R_v v\} = 1$.

(ii) Now consider $\hat{Q}_{-v}$ such that $\hat{Q}_{f'} = v$ and no other firm ranks v as acceptable. Then, v will be matched to f' in every matching that is stable for $(Q_v, \hat{Q}_{-v})$. Consequently, $\Pr\{\tilde{\varphi}[Q_v, \hat{Q}_{-v}](v) R_v v\} = 0$.

(iii) Given that $\hat{Q}_{-v}$ has positive probability under σ_{-v}, (ii) implies $\Pr\{\tilde{\varphi}[Q_v, \sigma_{-v}](v)\ R_v v\} = \sum_{Q_{-v} \in supp \sigma_{-v}} \sigma(Q_{-v}) \cdot \Pr\{\tilde{\varphi}[Q_v, Q_{-v}](v) R_v v\} < 1$. Hence, $\Pr\{\tilde{\varphi}[P_v, \sigma_{-v}](v) R_v v\} > \Pr\{\tilde{\varphi}[Q_v, \sigma_{-v}](v) R_v v\}$ and Q_v is not a best reply to σ_{-v}.

So, let Q_v only rank acceptable firms. We will now prove that deleting an acceptable firm from P_v cannot be a best reply to the completely mixed strategy profile σ_{-v}. So let $f \in A(P_v)$, but $f \notin A(Q_v)$. Let Q'_v be such that the restriction of Q_v and of Q'_v to $A(Q_v)$ coincide, but $f \in A(Q'_v)$ and $f' Q'_v f$, for all $f' \in A(Q_v)$. Note that $A(Q'_v) \subseteq A(P_v)$. We will show that Q_v does not stochastically P_v-dominate Q'_v when the other players choose σ_{-v}.

(i) If, for every Q_{-v}, v is unmatched under $\mu \in S(Q_v, Q_{-v})$, we have $\Pr\{\tilde{\varphi}[Q_v, Q_{-v}](v)\ P_v v\} = 0$, for all Q_{-v}. Since $A(Q'_v) \subseteq A(P_v)$, we also have $\Pr\{\tilde{\varphi}[Q'_v, Q_{-v}](v) P_v v\} \geq 0$. Hence, $\Pr\{\tilde{\varphi}[Q'_v, Q_{-v}](v) P_v v\} \geq \Pr\{\tilde{\varphi}[Q_v, Q_{-v}](v) P_v v\}$, for every Q_{-v}.

(ii) Otherwise, take any Q_{-v} such that $\mu(v) \in F$, with $\mu \in S(Q_v, Q_{-v})$. Let $Q' = (Q'_v, Q_{-v})$. We will prove that $\mu \in S(Q')$. Clearly, $\mu \in IR(Q')$, by definition of Q'. Now assume, by contradiction, that (f', w) block μ, *i.e.*, $f' Q'_w \mu(w)$ and $w Q'_{f'} \mu(f')$. Since $Q'_{f'} = Q_{f'}$, for every $f' \in F$, we have $w Q_{f'} \mu(f')$. Also, given that $Q'_w = Q_w$, for every $w \neq v$, the stability of μ for Q implies that $w = v$. Hence, $f' Q'_v \mu(v)$ and $v Q_{f'} \mu(f')$. It follows from the definition of Q'_v and the stability of μ for Q that $f' = f$ and that $\mu(v) = v$. This contradicts the initial assumption $\mu(v) \in F$.

We proved that, if $\mu(v) \in F$, for some $\mu \in S(Q_v, Q_{-v})$, we have $\mu \in S(Q')$. Since the set of unmatched agents is the same for all stable matchings (the first statement of this result appears in McVitie and Wilson, 1970; it also appears in Gale and Sotomayor, 1985, and Roth, 1984), v is matched under every matching that is stable for Q'. It follows that, for all Q_{-v}, $\Pr\{\tilde{\varphi}[Q'_v, Q_{-v}](v)P_v v\} \geq \Pr\{\tilde{\varphi}[Q_v, Q_{-v}](v)P_v v\}$.

(iii) To see that there exists some $\hat{Q}_{-v}$ for which v ends up alone when stating Q_v, but matched when using Q'_v, suppose $\hat{Q}_f = v$ and no other firm ranks v as acceptable. Then, v is matched to f with certainty at $(Q'_v, \hat{Q}_{-v})$, whereas he stays alone if using Q_v. As a consequence, there exists a $\hat{Q}_{-v}$ such that $1 = \Pr\{\tilde{\varphi}[Q'_v, \hat{Q}_{-v}](v)P_v v\} > \Pr\{\tilde{\varphi}[Q_v, \hat{Q}_{-v}](v)P_v v\} = 0$.

(iv) Since all profiles of preferences Q_{-v} have positive probability in the completely mixed strategy profile σ_{-v}, v will be unmatched with higher probability when using Q_v than when using Q'_v, we have $\Pr\{\tilde{\varphi}[Q'_v, \sigma_{-v}](v)P_v v\} > \Pr\{\tilde{\varphi}[Q_v, \sigma_{-v}](v)P_v v\}$ and Q_v is not a best reply to σ_{-v}. ■

Since ordinal perfect equilibrium strategies are best replies to completely mixed strategy profiles it immediately follows from the above result that ordinal perfect equilibrium strategies have to be exhaustive, listing all the acceptable partners, but leaving out those considered unacceptable. We state this formally in the following corollary.

Corollary 5.1 *Let $\tilde{\varphi}$ be a stable mechanism. If Q is an ordinal perfect equilibrium in the game induced by $\tilde{\varphi}$, every agent v ranks in Q_v all the partners that are acceptable according to its true preferences P_v (i.e., $A(Q_v) = A(P_v)$, for all $v \in V$).*

Two further implications of the above theorem are worth noticing. The first and most immediate goes against the celebrated properties of strict truncations. Formally, a strict truncation of an agent v's true preferences P_v containing p acceptable partners is a strategy that lists the first p', $p' < p$, elements of P_v as acceptable, preserving their order in P_v. Revealing a strict truncation of the true preferences may not be wise when one highly esteems being matched; furthermore, strict truncation strategies cannot be part of an ordinal perfect equilibrium in the game induced by a stable mechanism.

Second, it allows us to restrict the set of potential outcomes. As it will be readily understood, not every individually rational matching is sustainable as the outcome of an equilibrium play where agents fully reveal whom they are willing to match. In what follows, we describe those matchings that are beyond reach and state the result.

Definition 5.3 *Let $U(P)$ be the set of all individually rational matchings μ such that at least one of its blocking pairs either includes one agent that is unmatched under μ or consists of a pair of unmatched agents under μ.*

Proposition 5.2 *Let $\tilde{\varphi}$ be a stable mechanism. Let μ be a matching in $U(P)$. In the game induced by $\tilde{\varphi}$, μ is not sustainable in an ordinal perfect equilibrium.*

Proof. By Theorem 5.1, listing all the acceptable partners is a necessary requirement for an ordinal perfect equilibrium strategy. So, let Q be an ordinal Nash equilibrium such that $A(Q_v) = A(P_v)$ for all v. We will show that Q cannot support $\mu \in U(P)$.

By contradiction, assume that it does. In the game induced by a stable mechanism, every Nash equilibrium yields a single matching with probability one (Pais, 2004). Hence, Q leads to μ with probability one. By definition of μ, there exists at least one blocking pair for μ consisting of a firm f and a worker w such that either f or w or both are unmatched under μ. If both are unmatched under μ, since $f \in A(Q_w)$ and $w \in A(Q_f)$, we have $fQ_w\mu(w)$ and $wQ_f\mu(f)$. Hence, μ is not stable for Q and it cannot be reached as the outcome of a random stable mechanism where agents use Q.

So, let μ be blocked by (f, w) such that one of its members is unmatched, while the other one is matched under μ. Since the model is symmetric between firms and workers, it is sufficient to prove the proposition for, say, f unmatched and w matched under μ. Now let Q'_w be identical to Q_w, but such that f is listed first in Q'_w even if it occupies a worse position in Q_w. Define $Q' = (Q'_w, Q_{-w})$. We will show that Q'_w is different from Q_w (*i.e.*, Q_w does not list f first); in addition, Q'_w is a profitable deviation to Q_w, since f and w are matched with certainty under any matching in $S(Q')$.

To prove that f is matched to w under the firm-optimal stable matching at Q', we will use the deferred-acceptance algorithm with firms proposing. Since $Q'_v = Q_v$, for every agent $v \neq w$, all proposals, acceptances, and rejections take place exactly the same way as when Q was being used, up to the point where w holds f's proposal. This moment comes since $w \in A(Q_f)$ and f is the first firm in Q'_w. Also, w will not reject f until the final matching is reached, so that f and w are together under the firm-optimal stable matching.

It follows from the definition of worker-optimal stable matching that w holds f or a firm ranked higher than f at Q'_w under any stable matching for Q'. Since f is the first firm in Q'_w, w is matched to f under all stable matchings. This implies that $Q'_w \neq Q_w$; otherwise, (f, w) could not block μ.

To conclude, we have $1 = \Pr\{\tilde{\varphi}[Q'_w, Q_{-w}](w) = f\}$. Since Q leads to μ with certainty, $\Pr\{\tilde{\varphi}[Q](w) = f\} = 0$. Hence, $\Pr\{\tilde{\varphi}[Q'_w, Q_{-w}](w) R_w f\} > \Pr\{\tilde{\varphi}[Q_w, Q_{-w}](w) R_w f\}$ and Q_w is not a best reply to Q_{-w}, contradicting that Q is a Nash equilibrium. Since no Nash equilibrium where agents list all the acceptable partners can sustain a matching in $U(P)$, no ordinal perfect equilibrium will. ■

In the other direction, it may be shown that every matching in $IR(P)\backslash U(P)$ may be sustained as the unique outcome of an equilibrium play of the game where all agents reveal the full set of acceptable partners.

Proposition 5.3 *Let $\tilde{\varphi}$ be a stable mechanism. Let μ be a matching in $IR(P)\backslash U(P)$. Then, there exists an ordinal Nash equilibrium Q in the game induced by $\tilde{\varphi}$ with the following properties:*

1. *every agent v ranks as acceptable in Q_v all of its acceptable partners under P_v*
2. *Q sustains μ with probability one.*

Proof. Let μ be a matching in $IR(P)\backslash U(P)$ and consider agent v. Let Q_v be such that (i) $A(Q_v) = A(P_v)$ and (ii) *if* v is matched under μ, $\mu(v) Q_v v'$, for all $v' \in A(P_v)$, *i.e.*:

$$Q_v = \begin{cases} \overbrace{\mu(v), \ldots\ldots\ldots\ldots\ldots\ldots}^{\text{All elements of } A(P_v) \text{ in any order}}, v & \text{if } \mu(v) \neq v \\ \underbrace{\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots}_{\text{All elements of } A(P_v) \text{ in any order}}, v & \text{if } \mu(v) = v \end{cases} .$$

We will show that Q has all the described properties.

First, we will show, by contradiction, that $\mu \in S(Q)$. By definition of Q, it is clear that $\mu \in IR(Q)$. So assume (f, w) blocks μ when the profile Q is considered. By (ii), this implies that f and w are unmatched under μ. We thus have $w Q_f f$ and $f Q_w w$. Since $A(Q_v) = A(P_v)$ for every agent v, it follows that $w P_f f$ and $f P_w w$. Hence, (f, w) are unmatched under μ and block μ for preferences P. This contradicts that $\mu \in IR(P)\backslash U(P)$.

Now we will prove that μ is the unique matching in $S(Q)$. Assume not and take any matching $\hat{\mu}$, $\hat{\mu} \neq \mu$, in $S(Q)$. Let v be such that $\mu(v) = v$. Then, $\hat{\mu}(v) = v$ since, by Proposition 5.1, the same set of agents is unmatched under every matching that belongs to $S(Q)$. On the other hand, for $\hat{v}$ such that $\mu(\hat{v}) \neq \hat{v}$, we must have $\hat{\mu}(\hat{v}) = \mu(\hat{v})$ by (ii). Otherwise, $(\hat{v}, \mu(\hat{v}))$ blocks $\hat{\mu}$. Hence, $\hat{\mu} = \mu$

and μ is the only matching in $S(Q)$. As a consequence, every random stable mechanism leads to μ with probability one.

To complete the proof, we must show that Q is an ordinal Nash equilibrium. By contradiction, suppose firm f can profitably deviate by matching worker w (the same argument holds for an arbitrary worker). This implies that there exists a worker w willing to accept f, *i.e.*, such that $fQ_w\mu(w)$. By (ii), we must have $\mu(w) = w$ and, by (i), $fP_w\mu(w)$. Since $\mu \in IR(P)\backslash U(P)$, it follows that $\mu(f)P_f w$ and we contradict the initial assumption: matching w is not a profitable deviation for f. ■

We now state an important result for deterministic stable mechanisms that follows from Barberà and Dutta (1995). In this paper, revealing the true preferences is most convenient for agents who are extremely risk averse. In fact, when an agent compares straightforward behavior with any misrepresentation of its preferences, there exists a potential partner with whom, by manipulating, it ends up matched for a larger set of actions of the other players; further, less preferred potential partners are obtained, by either acting straightforwardly or strategically, against the same profiles for the rest of society. It thus follows that when an agent's beliefs are such that all preference profiles for the other agents may be revealed with positive probability, behaving strategically is never a best reply. We state this formally in the next Theorem. Even though the result applies to the mechanism producing the firm-optimal stable matching, it is straightforward to extend it to a market using the worker-optimal stable mechanism.

Theorem 5.2 *[Barberà and Dutta (1995)] In the game induced by φ^F, if Q_v is a best reply to a completely mixed strategy profile σ_{-v}, then Q_v are agent v's true preferences (*i.e.*, $Q_v = P_v$).*

It clearly follows that only truth telling may be an ordinal perfect equilibrium in the game induced by the mechanism that yields an optimal stable matching. We state this as a corollary to Theorem 5.2.

Corollary 5.2 *In an ordinal perfect equilibrium of the game induced by φ^F every agent states its true preferences.*

This result anticipates that ordinal perfect equilibria only seldom exist in matching markets. The following example supports this observation.

Example 5.1 *A market where there are no ordinal perfect equilibria in pure strategies.*

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

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

Consider the game induced by the mechanism that produces the firm-optimal stable matching. Corollary 5.2 establishes that, under an ordinal perfect equilibrium in this game, every agent chooses the honest announcement of preferences. In this case, the mechanism would yield the matching $\mu = \{(f_1, w_1), (f_2, w_2)\}$. Nevertheless, truth telling fails to meet the basic requirement of being a Nash equilibrium of the game, since both workers can profitably deviate. For example, submitting $Q_{w_1} = f_2$ is a deviation for worker w_1, conveying the position in f_2. $\diamond$

We can extend the above result to random stable mechanisms that only assign positive probability to the firm-optimal and to the worker-optimal stable matchings.

Proposition 5.4 *Let $\tilde{\varphi}$ be a stable mechanism that yields the firm-optimal stable matching with probability α and the worker-optimal stable matching with probability $1 - \alpha$, $0 < \alpha < 1$. In an ordinal perfect equilibrium in the game induced by $\tilde{\varphi}$ every agent states its true preferences.*

Proof. Let w be an arbitrary worker. We will show that P_w is the only strategy that can be part of an OP equilibrium in the game induced by the random stable mechanism $\tilde{\varphi}$ as defined above. We omit the proof for an arbitrary firm f, since the same arguments, with obvious modifications, can be applied.

Let Q_w be a strategy for w such that $Q_w \neq P_w$. Assume that Q_w is a best reply to a completely mixed strategy profile σ_{-w}. This has two implications. First, by Corollary 5.1, $A(Q_w) = A(P_w)$. Second, $\Pr\{\tilde{\varphi}[Q_w, \sigma_{-w}](w) R_w v\} \geq \Pr\{\tilde{\varphi}[P_w, \sigma_{-w}](w) R_w v\}$, for every v, a potential partner of w. By definition of $\tilde{\varphi}$, we have $\alpha \Pr\{\varphi^F[Q_w, \sigma_{-w}](w) R_w f\} + (1-\alpha) \Pr\{\varphi^W[Q_w, \sigma_{-w}](w) R_w f\} \geq \alpha \Pr \{\varphi^F [P_w, \sigma_{-w}](w) R_w f\} + (1-\alpha) \Pr\{\varphi^W[P_w, \sigma_{-w}](w) R_w f\}$, for every v. Nevertheless, since truth telling is a dominant strategy for workers in the game induced by the worker-optimal stable mechanism, it is a best reply to any mixed strategy profile and $\Pr\{\varphi^W[P_w, \sigma_{-w}](w) R_w f\} \geq \Pr\{\varphi^W[Q_w, \sigma_{-w}](w) R_w f\}$. Moreover, Theorem 5.2 states that honestly revealing the true preferences is a best reply in the game induced by the firm-optimal stable mechanism, so that $\Pr\{\varphi^F[P_w, \sigma_{-w}] (w) R_w f\} \geq \Pr\{\varphi^F[Q_w, \sigma_{-w}](w) R_w f\}$. It follows that $\varphi^W[P_w, \sigma_{-w}](w) = \varphi^W[Q_w, \sigma_{-w}](w)$ and $\varphi^F[P_w, \sigma_{-w}](w) = \varphi^F[Q_w, \sigma_{-w}](w)$. Since σ_{-w} is a completely mixed strategy profile, these distributions have full support, *i.e.*, every firm in $A(P_w)$ is obtained as a partner with positive probability. This implies that $Q_w = P_w$, contradicting the initial assumption.

As a consequence, only P_w can be a best reply to a mixed strategy; thus only P_w can be part of an OP equilibrium. ■

Finally, confirming our conjecture on the existence of ordinal perfect equilibria, in the next result we show that truth telling can only be an ordinal perfect equilibrium if it is a dominant strategy for every agent. Hence, the concept of ordinal perfect equilibrium and the apparently stronger concept of Nash equilibria in dominant strategies coincide.

Theorem 5.3 *In the game induced by φ^F, the sets of ordinal perfect equilibria and of Nash equilibria in dominant strategies coincide.*

Proof. It is clear that every Nash equilibrium in dominant strategies is an OP equilibrium. In fact, a dominant strategy is a best reply to all profiles of preferences stated by the other players; hence, it is also a best reply to any completely mixed strategy profile. The converse statement will be shown in what follows.

Theorem 5.2 imposes as a necessary requirement for an OP equilibrium that every agent states its true preferences. Hence, let P be an OP equilibrium in φ^F, but assume that stating the true preferences is not a dominant strategy for some worker w. Then, there exists at least one instance, *i.e.*, a strategy profile for the other players, under which playing strategically pays for worker w. Denote by Q_{-w} such a strategy profile and let Q_w be the best reply to Q_{-w}. Formally,

$$\varphi^F[Q_w, Q_{-w}](w) P_w \varphi^F[P_w, Q_{-w}](w) \text{ and} \tag{5.1}$$

$$\varphi^F[Q_w, Q_{-w}](w) R_w \varphi^F[\bar{Q}_w, Q_{-w}](w), \text{ for every } \bar{Q}_w. \tag{5.2}$$

Let, without loss of generality, $P_w = f_1, f_2, ..., f_m$ and $f_j = \varphi^F[Q_w, Q_{-w}](w)$, with $1 \leq j \leq m$.

Now define $Q'_w = f_j, f_{j-1}, ..., f_1$. Observe that $\varphi^F[Q_w, Q_{-w}] \in S(Q'_w, Q_{-w})$, since it remains individually rational once w uses Q'_w and there are potentially fewer blocking pairs for $\varphi^F[Q_w, Q_{-w}]$. Hence, Proposition 5.1 implies that w is matched under every matching in $S(Q'_w, Q_{-w})$; in addition, the definition of Q'_w implies that he is matched to a firm at least as good as f_j according to P_w. By (5.2), $\varphi^F[Q_w, Q_{-w}](w) R_w \varphi^F[Q'_w, Q_{-w}](w)$, so that we must have $\varphi^F[Q'_w, Q_{-w}](w) = f_j$. Then, if Q_w gives w matched to f_j against Q_{-w}, Q'_w also matches w to f_j. Hence, for the profile Q_{-w}, condition (5.1) yields $f_j = \varphi^F[Q'_w, Q_{-w}](w) P_w \varphi^F[P_w, Q_{-w}](w)$.

Now let us prove that there is no instance $\hat{Q}_{-w}$ under which P_w matches w to a firm at least as good as f_j, while Q'_w leaves w unmatched. By contradiction, assume that, by playing truthfully, w is matched to f_i, $i \leq j$, but unmatched when using Q'_w against $\hat{Q}_{-w}$ in the game induced by φ^F. If this is so, by Proposition 5.1, w is unmatched under every matching that is stable for $(Q'_w,\hat{Q}_{-w})$ and, in particular, we have $\varphi^W[Q'_w, \hat{Q}_{-w}](w) = w$. On the other hand, by definition of worker-optimal stable matching, $\varphi^W[P_w, \hat{Q}_{-w}](w) R_w \varphi^F[P_w, \hat{Q}_{-w}](w) = f_i$. Since $f_i R_w f_j$, we have $\varphi^W[P_w, \hat{Q}_{-w}](w) \in A(Q'_w)$. Now imagine Q'_w are w's true preferences. By acting strategically and using P_w, w is better off than by straightforwardly revealing Q'_w. This contradicts the fact that truth is a dominant strategy for workers in the game induced by φ^W. Hence, w is matched to a firm at least as good as f_j with P_w, by manipulating and using Q'_w, w will also be matched to a firm at least as good as f_j. We thus have, for every $\bar{Q}_{-w}$ that yields $\varphi^F[P_w, \bar{Q}_{-w}](w) R_w f_j$, that $\varphi^F[Q'_w, \bar{Q}_{-w}](w) R_w f_j$.

Consider a completely mixed strategy profile σ_{-w}. In the game induced by φ^F, when playing against σ_{-w}, Q'_w yields a higher probability of being matched to a firm at least as good as f_j than P_w. Clearly, P_w cannot be part of an OP equilibrium, contradicting the initial assumption. ■

The whole picture changes when we depart from the ordinal framework. As shown in the following example, if agents are able to go beyond an ordering of the possible matches and provide a measure of their preferences, strategic behavior may be held in a perfect equilibrium.

Example 5.2 *(Example 5.1 (revisited)) Acting strategically may be a perfect equilibrium when agents can give a cardinal meaning to their preferences.*

Consider the game induced by the mechanism that yields the firm-optimal stable matching in the matching market described above. Consider the profile of strategies $Q = (P_F, Q_W)$, such that each worker only finds his first choice acceptable in Q (*i.e.*, $Q_{w_1} = f_2$ and $Q_{w_2} = f_1$). We will show that, depending on the utility representation of the workers' preferences, Q may be a perfect equilibrium of the game.

Each agent has five different strategies at its disposal (two of them stating two acceptable matches, other two naming only one, and the strategy where all potential partners are unacceptable). Let σ^k be a sequence of completely mixed strategy profiles such that, for $k \geq 1$ and for every agent v, $\sigma^k(\hat{Q}_v) = \frac{1}{k+4}$, for all $\hat{Q}_v \neq Q_v$, and $\sigma^k(Q_v) = 1 - \frac{4}{k+4}$. Note that $\{\sigma^k\}\underset{k\to\infty}{\longrightarrow} Q$. Revealing the true preferences is a dominant strategy for each firm f in this game (Dubins and Freedman, 1981, and Roth, 1982), outperforming every alternative strategy for every profile chosen by the other agents, namely σ^k_{-f}, for every $k \geq 1$. So, consider worker w_1 (by symmetry, what follows also holds for w_2); we will prove that Q_{w_1} is a best reply to $\sigma^k_{-w_1}$.

(i) Consider $Q'_{w_1} = w_1$; note that w_1 is always unmatched when using this strategy against any profile of strategies of the other players. Hence, w_1 is unmatched with certainty when playing Q'_{w_1} against $\sigma^k_{-w_1}$. It follows that Q'_{w_1} is stochastically P_{w_1}-dominated by every other strategy that w_1 may use, in particular by P_{w_1} when playing against $\sigma^k_{-w_1}$.

(ii) The strategy $Q''_{w_1} = f_1$ is also stochastically P_{w_1}-dominated by $P_{w_1} = f_2, f_1$ against $\sigma^k_{-w_1}$. In fact, if w_1 is unmatched when using P_{w_1} against Q_{-w_1}, he will certainly be unmatched with Q''_{w_1}. So, when playing against $\sigma^k_{-w_1}$, w_1 is matched with higher probability if he uses P_{w_1}. Moreover, there exist profiles of strategies for the other players such that w_1 is matched to f_2 under the outcome of the deferred-acceptance algorithm when revealing Q'''_{w_1}, but not with Q''_{w_1}, where f_2 is considered unacceptable. The conclusion follows.

(iii) Now consider $Q'''_{w_1} = f_1, f_2$. Note that w_1 is unmatched when using this strategy if and only if he is unmatched with $P_{w_1} = f_2, f_1$. Furthermore, for every profile of the other players such that w_1 is assigned to f_2 with Q'''_{w_1}, he is also assigned to f_2 when using P_{w_1}; and there are profiles of strategies for the other players such that w_1 is matched to f_2 when revealing P_{w_1}, but not with Q'''_{w_1}. It follows that P_{w_1} stochastically P_{w_1}-dominates Q'''_{w_1} against any completely mixed strategy profile $\sigma^k_{-w_1}$.

(iv) Since P_{w_1} outperforms Q'_{w_1}, Q''_{w_1}, and Q'''_{w_1}, it is sufficient to find under which conditions Q_{w_1} may be preferred to P_{w_1}. There is an instance under which submitting Q_{w_1} provides w_1 with a better partner. In fact, f_2 is w_1's partner under the firm-optimal stable matching with $(Q_{w_1}, \hat{Q}_{w_2}, P_{f_1}, P_{f_2})$, where $\hat{Q}_{w_2} = f_1, f_2$; by using P_{w_1} against the same profile for the others, w_1 ends up matched to f_1. Nevertheless, w_1 is unmatched when revealing Q_{w_1} against a larger set of profiles of the other players, than when using P_{w_1}. It turns out that Q_{w_1} is a best reply to $\sigma^k_{-w_1}$, if the following condition holds: $(k)^2[u(f_2) - u(f_1)] \geq (4(k+4)^2 - 17k - 51)[u(f_1) - u(w_1)]$.[25] In particular, when $u(f_2) - u(f_1)$ is larger than $u(f_1) - u(w_1)$, w_1 benefits from listing only his first choice. ◇

5.5 Further Research

As mentioned in Chapter 1, the aim of this paper is to narrow the set of potential equilibrium outcomes by imposing stronger rationality constraints than those underlying the concept of Nash equilibrium. Nevertheless, the analysis performed

[25] This expression results from considering the outcomes of the deferred-acceptance algorithm when w_1 uses P_{w_1} and Q_{w_1} for all possible combinations of the other agents' preferences. This calls for performing a total of 5^3 tedious comparisons that we leave out for obvious reasons.

here should be considered very preliminary. We have shown that only truth telling may be a best reply to a completely mixed strategy profile and, thus, part of an ordinal perfect equilibrium. Such negative result on the existence of ordinal perfect equilibria calls for a weaker concept. One possible course of action lies in considering that agents never submit strategies that are stochastically dominated against a completely mixed strategy profile, while those that are not stochastically dominated should be regarded as potential choices. Such concept is closer in spirit to the notion of perfect equilibrium in expected utilities. In fact, each of the latter strategies should be a best reply to a completely mixed strategy profile for some utility representation of the true preferences. We can then show that strategies that do not list the best partner first are stochastically dominated and conclude that a strict subset of the individually rational matchings can be sustained in weak ordinal perfect equilibria.

Bibliography

[1] ABDULKADIROGLU, A., AND T. SÖMNEZ (1999): "House Allocation with Existing Tenants," *Journal of Economic Theory*, 88, 233-260.

[2] ABDULKADIROGLU, A., AND T. SÖMNEZ (2003): "School Choice: A Mechanism Design Approach," *American Economic Review*, 93(3), 729-747.

[3] BARBERÀ, S., AND B. DUTTA (1995): "Protective Behavior in Matching Models," *Games and Economic Behavior*, 8, 281-296.

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

[5] D'ASPREMONT, C., AND B. PELEG (1988): "Ordinal Bayesian Incentive Compatible Representations of Committees," *Social Choice and Welfare*, 5, 261-279.

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

[7] EHLERS, L. (2003): "In Search of Advice for Physicians in Entry-Level Medical Labor Markets," mimeo.

[8] EHLERS, L. (2004): "In Search of Advice for Participants in Matching Markets which Use the Deferred-Acceptance Algorithm," *Games and Economic Behavior*, 48, 249-270.

[9] EHLERS, L., AND J. MASSÓ (2003): "Incomplete Information and Small Cores in Matching Markets," mimeo.

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

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

[12] HAERINGER, G., AND M. WOODERS (2004): "Decentralized Job Matching," mimeo.

[13] KNUTH, D. E. (1976): *Marriages Stables*, Les Presses de l'Université de Montréal, Montréal.

[14] MA, J. (2002): "Stable Matchings and the Small Core in Nash Equilibrium in the College Admissions Problem," *Review of Economic Design*, 7, 117-134.

[15] MAJUMDAR, D. (2003): "Ordinally Baeyesian Incentive Compatible Stable Matchings", mimeo.

[16] MAJUMDAR, D., AND A. SEN (2004): "Ordinally Bayesian Incentive Compatible Voting Rules," *Econometrica*, 72, 523-540.

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

[18] MOULIN, H. (1997): "Procedural cum End State Justice: An Implementation Viewpoint, " Justice, Political Liberalism and Utilitarism, editors, Maurice Salles and John Weymark. Proceedings of a conference in honor of John Harsanyi and John Rawls, Cambridge University Press, forthcoming.

[19] MOULIN, H. (2003): *Fair Division and Collective Welfare*, MIT Press, Cambridge, Massachussets.

[20] NIEDERLE, M., AND A. E. ROTH (2003): "Unraveling Reduces Mobility in a Labor Market: Gastroentrology With and Without a Centralized Match," *Journal of Political Economy*, forthcoming.

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

[22] ROTH, A. E. (1984a): "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory," *Journal of Political Economy*, 92, 991-1016.

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

[24] ROTH, A. E. (1989): "Two-Sided Matching with Incomplete Information about Others' Preferences," *Games and Economic Behavior*, 1, 191-209.

[25] ROTH, A. E. (1990): "New Physicians: A Natural Experiment in Market Organization," *Science*, 250, 1524-1528.

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

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

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

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

[30] ROTH, A. E., AND J. H. V. VATE (1991): "Incentives in Two-Sided Matching with Random Stable Mechanisms," *Economic Theory*, 1, 31-44.

[31] ROTH, A. E., AND X. XING (1994): "Jumping the Gun: Imperfections and Institutions Related to the Timing of Market Transactions," *American Economic Review*, 84, 992-1044.

[32] ROTH, A. E., AND X. XING (1997): "Turnaround Time and Bottlenecks in Market

Clearing: Decentralized Matching in the Market for Clinical Psychologists," *Journal of Political Economy*, 105, 284-329.

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