Procedurally Fair and Stable Matching*

Bettina Klaus† Flip Klijn‡

November 2004

Abstract: We motivate procedural fairness for matching mechanisms and study two procedurally fair and stable mechanisms: *employment by lotto* (Aldershof *et al.*, 1999) and the *random order mechanism* (Roth and Vande Vate, 1990, Ma, 1996). For both mechanisms we give various examples of probability distributions on the set of stable matchings and discuss properties that differentiate employment by lotto and the random order mechanism. Finally, we consider an adjustment of the random order mechanism, *the equitable random order mechanism*, that combines aspects of procedural and "endstate" fairness.
Keywords: procedural fairness, random mechanism, stability, two-sided matching.
JEL classification: C78, D63

1 Introduction

The marriage model describes a two-sided, one-to-one matching market without money where the two sides of the market for instance are workers and firms (job matching) or medical students and hospitals (matching of students to internships). We use the common terminology in the literature and refer to one side of the market as "men" and to the other as "women." An outcome for a marriage market is called a matching, which can simply be described by a collection of single agents and "married" pairs (consisting of one man and one woman). Loosely speaking, a matching is stable if all agents have acceptable spouses and there is no couple whose members both like each other better than their current spouses. Gale and Shapley (1962) formalized this notion of stability for marriage markets and provided an algorithm to calculate stable matchings. These classical results (Gale and Shapley, 1962) inspired many researchers to study stability not only for the marriage model, but for more general models as well. We refer to Roth and Sotomayor (1990) for a comprehensive account on stability for two-sided matching models.

In this paper we study fairness and stability in marriage markets. Masarani and Gokturk (1989) showed several impossibilities to obtain a fair deterministic matching mechanism within the context of Rawlsian justice. In contrast to this cardinal approach we focus on the ordinal aspects of the model and opt for an approach of procedural fairness. Since for any deterministic matching mechanism we can detect an inherent favoritism either for one side of the

*We thank two referees for helpful comments and suggestions. B. Klaus's and F. Klijn's research is supported by Ramón y Cajal contracts of the Spanish *Ministerio de Ciencia y Tecnología.* The work of the authors is also partially supported by Research Grant BEC2002-02130 from the Spanish *Ministerio de Ciencia y Tecnología* and by the Barcelona Economics Program of CREA. This paper is part of the PAC Project CIT-2-CT-2004-506084 funded by the European Commission.

†Departament d'Economia i d'Història Econòmica, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain; e-mail: `bettina.klaus@uab.es`

‡Corresponding author. Institut d'Anàlisi Econòmica (CSIC), Campus UAB, 08193 Bellaterra (Barcelona), Spain; Tel. (34) 93 580 6612; Fax. (34) 93 580 1452; e-mail: `fklijn@pareto.uab.es`

market or for some agents over others, in order to at least recover *ex ante* fairness, we consider probabilistic matching mechanisms that assign to each marriage market a probability distribution on the set of stable matchings. We do not intend to judge the fairness of a probabilistic matching mechanism by judging the assigned probability distributions (the "endstate"), but by considering procedurally fair matching algorithms in which the sequence of moves for the agents is drawn from a uniform distribution. Hence, whenever an agent has the same probability to move at a certain point in the procedure that determines the final probability distribution, we consider the random matching mechanism to be *procedurally fair*. In other words, here we focus on "procedural justice" rather than on endstate justice (see Moulin, 1997,2003). After a discussion of procedural fairness, we explain our procedural fairness concept for matching markets, discuss two procedurally fair and stable matching mechanisms, and conclude with a stable mechanism that combines some aspects of procedural and endstate fairness.

The first procedurally fair and stable matching mechanism we consider, called *employment by lotto*, was proposed by Aldershof *et al.* (1999). Loosely speaking, employment by lotto can be considered to be a random serial dictatorship on the set of stable matchings. A first agent is drawn randomly and can discard all stable matchings in which he/she is not matched to his/her best partner (possibly him-/herself) in a stable matching. Exclude the first agent and his/her partner from the set of agents and randomly choose the next agent who can discard all stable matchings in which he/she is not matched to his/her best partner in the reduced set of stable matchings. Continue with this sequential reduction of the set of stable matchings until it is reduced to a singleton. Using all possible sequences of agents, this mechanism induces a probability distribution on the set of stable matchings. The associated probabilistic matching mechanism of this probabilistic sequential dictatorship equals employment by lotto. We give various examples of probability distributions on the set of stable matchings induced by employment by lotto and show certain limitations of this mechanism (*e.g.*, complete information of all agents' preferences is needed).

The second procedurally fair and stable matching mechanism we consider is a random matching mechanism based on Roth and Vande Vate's (1990) results. We follow Ma (1996) and refer to this rule as the *random order mechanism*. The basic idea is as follows. Imagine an empty room with one entrance. At the beginning, all agents are waiting outside. At each step of the algorithm one agent is chosen randomly and invited to enter. Before an agent enters, the matching in the room is stable. However, once an agent enters the room, the existing matching in the room may become unstable, meaning that the new agent can form a blocking pair with another agent that already is present in the room. By satisfying this (and possible subsequent) blocking pair(s) in a certain way a new stable matching including the entering agent is obtained for the marriage market in the room. After a finite number of steps a stable matching for the original marriage market is obtained. Using all possible sequences of agents, this mechanism induces a probability distribution on the set of stable matching. The associated probabilistic matching mechanism equals the random order mechanism. We give various examples of probability distributions on the set of stable matchings induced by the random order mechanism. Furthermore, we correct the probability distribution for the marriage market considered by Ma (1996). We detected that the small mistake in the calculations by Ma (1996) is due to the fact that even though the example looks very symmetric, some of the calculations are not as "symmetric" since the random order mechanism does not satisfy what we call *independence of dummy agents*; that is, the final probability distribution on the set of stable matchings may crucially depend on preferences of agents who are matched to the same partner in all stable matchings.

Third, following a suggestion by Romero-Medina (2002), we briefly discuss an adjustment of the random order mechanism, *the equitable random order mechanism*, that combines aspects of procedural and endstate fairness.

Our examples show that even for small markets the three mechanisms may give completely different outcomes. In all our examples, we implement the mechanisms in Matlab ©. In some examples the resulting probabilities are rounded.

The article is organized as follows. In Section 2 we introduce marriage markets and stability. In Section 3 we first introduce and discuss procedural fairness. In Section 3.1 we study employment by lotto, in Section 3.2 the random order mechanism, and in Section 3.3 its adjustment, the equitable random order mechanism. We conclude with Section 4.

2 Matching Markets and Stability

First we introduce the model of a two-sided one-to-one matching market without money. For convenience we apply Gale and Shapley's (1962) interpretation of a "marriage market." For further details on the interpretation and standard results we refer to Roth and Sotomayor's (1990) comprehensive book on two-sided matching.

There are two finite and disjoint sets of agents: a set $M = \{m_1, \ldots, m_a\}$ of "men" and a set $W = \{w_1, \ldots, w_b\}$ of "women," where possibly $a \neq b$. The set of agents equals $N = M \cup W$. Let $n = |N|$. We denote a generic agent by i, a generic man by m, and a generic woman by w.

Each agent has a complete, transitive, and strict preference relation over the agents on the other side of the market and the prospect of being alone. Hence, man m's preferences $\succeq_m$ can be represented as a strict ordering $P(m)$ of the elements in $W \cup \{m\}$, for instance: $P(m) = w_3\, w_2\, m\, w_1\, \ldots\, w_4$ indicates that m prefers w_3 to w_2 and he prefers remaining single to any other woman. Similarly, woman w's preferences $\succeq_w$ can be represented as a strict ordering $P(w)$ of elements in $M \cup \{w\}$. Let P be the profile of all agents' preferences: $P = (P(i))_{i \in N}$.

We write $w \succ_m w'$ if m strictly prefers w to w' ($w \neq w'$), and $w \succeq_m w'$ if m likes w at least as well as w' ($w \succ_m w'$ or $w = w'$). Similarly we write $m \succ_w m'$ and $m \succeq_w m'$. A woman w is *acceptable* to a man m if $w \succ_m m$. Analogously, m is acceptable to w if $m \succ_w w$.

A *marriage market* is a triple (M, W, P). A *matching* for a marriage market (M, W, P) is a one-to-one function μ from N to itself, such that for each $m \in M$ and for each $w \in W$ we have $\mu(m) = w$ if and only if $\mu(w) = m$, $\mu(m) \notin W$ implies $\mu(m) = m$, and similarly $\mu(w) \notin M$ implies $\mu(w) = w$. If $\mu(m) = w$, then man m and woman w are matched to one another. If $\mu(i) = i$, then agent i is *single*. We call $\mu(i)$ the *match of agent* i at μ. When denoting a matching μ we list the women that are matched to men $m_1, m_2, \ldots$, *e.g.*, $\mu = w_3\, w_4\, m_3\, w_1$ denotes a matching where m_1 is matched to w_3, m_2 to w_4, m_3 to himself, and m_4 to w_1.

A key property of matchings is *stability*. First, since agents can always choose to be single, we require a voluntary participation condition. A matching μ is *individually rational* if only acceptable agents are matched to one another, *i.e.*, $\mu(i) \succeq_i i$ for all $i \in N$. Second, if an agent can improve upon his/her present match by switching to another agent such that this agent is better off as well, then we would expect this mutually beneficial "trade" to be carried out, rendering the given matching instable. For a given matching μ, a pair (m, w) is a *blocking pair* if they are not matched to one another but prefer one another to their matches at μ, *i.e.*, $w \succ_m \mu(m)$ and $m \succ_w \mu(w)$. A matching is *stable* if it is individually rational and if there are no blocking pairs. With a slight abuse of notation, we denote the set of stable matchings for marriage market

(M, W, P) by $S(P)$. Gale and Shapley (1962) proved that $S(P) \neq \emptyset$. Furthermore, any set of stable matchings has the structure of a (distributive) lattice, which we explain next.

For any two matchings μ and μ' we define the function $\lambda := \mu \vee_M \mu'$ on N that assigns to each man his more preferred match from μ and μ' and to each woman her less preferred match. Formally, let $\lambda = \mu \vee_M \mu'$ be defined for all $m \in M$ by $\lambda(m) := \mu(m)$ if $\mu(m) \succ_m \mu'(m)$ and $\lambda(m) := \mu'(m)$ otherwise, and for all $w \in W$ by $\lambda(w) := \mu(w)$ if $\mu'(w) \succ_w \mu(w)$ and $\lambda(w) := \mu'(w)$ otherwise. Similarly, we define the function $\mu \wedge_M \mu'$ that gives each man his less preferred match and each woman her more preferred match. The following theorem (published by Knuth, 1976, but attributed to John Conway) establishes the lattice structure of the set of stable matchings.

Theorem 2.1 [Lattice Theorem] *If $\mu, \mu' \in S(P)$, then also $\mu \vee_M \mu', \mu \wedge_M \mu' \in S(P)$.*

From Theorem 2.1 and the existence of a stable matching it follows easily that there is a stable matching μ_M that is optimal for all men in the sense that no other stable matching μ gives to any man m a match $\mu(m)$ that he prefers to $\mu_M(m)$. Similarly, there is a stable matching μ_W that is optimal for all women. In fact, Gale and Shapley (1962) already provided an algorithm, called the deferred acceptance procedure, to calculate μ_M and μ_W.

Since preferences are strict, the set of matched agents does not vary from one stable matching to another (Roth, 1982), *i.e.*, the set of single agents is the same for all stable matchings.

Theorem 2.2 *For all $i \in N$ and all $\mu, \mu' \in S(P)$, $\mu(i) = i$ implies $\mu'(i) = i$.*

3 Procedural Fairness

Typing the keyword "procedural fairness" into an internet search engine confirms that the issue of procedural fairness is central in many areas of our lives: a Google search (conducted on October 21st, 2004) resulted in many "webreferences" on procedural fairness in personnel and office management (*e.g.*, concerning staff dismissal or promotion), legal applications of procedural fairness (*e.g.*, in family and employment law), and procedural fairness guidelines for private schools and universities. Apart from administrative and legal applications, procedural fairness also plays an important role in economics. Two recent papers that empirically investigate procedural fairness in economic environments are Anand (2001) and Bolton *et al.* (2004). Anand (2001) argues for the relevance of procedural fairness in economics and social choice and uses survey data to analyze various hypotheses on different aspects of procedural fairness. Bolton *et al.* (2004) use experimental data to analyze the distinction and relation between procedural and allocation (endstate) fairness for ultimatum games and "battle-of-the-sexes" games. We refer the interested reader to these two papers for further references of procedural fairness in administrative and management sciences, economics, law, and psychology.

The notion of procedural fairness we are interested in is equivalent to Rawls's (1971, p. 86) *pure procedural justice*: "By contrast, pure procedural justice obtains when there is no independent criterion for the right result: instead there is a correct or fair procedure such that the outcome is likewise correct or fair, whatever it is, provided that the procedure has been properly followed. This situation is illustrated by gambling. If a number of persons engage in a series of fair bets, the distribution of cash after the last bet is fair, or at least not unfair, whatever this distribution is." As in Rawls's (1971) pure procedural justice, we introduce procedural fairness in a situation in which there is no criterion for what constitutes a fair outcome other than the procedure itself. His classification of fair gambling as procedurally just already points towards the formalization of procedural fairness that we apply in our matching context: we will use matching mechanisms that are based on fair lotteries (uniform randomization) as a means to establish procedural fairness.

We are interested in matching mechanisms that produce stable matchings and that can be considered "fair." Before explaining the concept of procedural fairness that we apply here, we define stable matching mechanisms. A *stable matching mechanism* μ is a function that for any marriage market (M, W, P) assigns a stable matching $\mu(M, W, P)$.

Two well-known and widely applied stable matching mechanisms are the *man-optimal* and the *woman-optimal deferred acceptance (DA)* algorithm by Gale and Shapley (1962). As discussed in Section 2, for any marriage market (M, W, P), the man-optimal DA algorithm yields the (unique) stable matching preferred by all men and the woman-optimal DA algorithm yields the (unique) stable matching preferred by all women. However, although stable, for all marriage markets where the man-optimal matching differs from the woman-optimal matching, which is the rule rather than the exception, each of the matching mechanisms clearly favors one side of the market. If there is no obvious reason why one side of the market should be favored, this favoritism can be considered "unfair."

This inherent incompatibility between stability and fairness is not restricted to the man-optimal and the woman-optimal DA algorithm, but in fact extends to all deterministic matching rules. Given the lattice structure of the set of stable matchings, for some marriage markets any deterministic matching mechanism is bound to favor one side of the market; for instance whenever the set of stable matchings consist of a man-optimal and a woman-optimal matching. Even if the matching mechanism does not choose a man-optimal or woman-optimal matching whenever possible, depending on the lattice structure of stable matchings, some agents may have to be favored relative to other agents on both sides of the market. Therefore, in order to formulate fairness without sacrificing stability, we consider probabilistic stable matching mechanisms, that is, for each marriage problem (M, W, P) a *probabilistic stable matching mechanism* assigns a probability distribution $\mathcal{P}(M, W, P)$ over the set of stable matchings $S(P)$.

We do not intend to judge the fairness of a probabilistic stable matching mechanism by judging the *endstate*, that is, the assigned probability distributions, but by considering procedurally fair matching algorithms in which the sequence of moves for the agents is drawn from a uniform distribution. Loosely speaking, whenever each agent has the same probability to move at a certain point in the procedure that determines the final probability distribution, we consider the respective probabilistic stable matching mechanism to be *procedurally fair*.

3.1 Procedural Fairness: Employment by Lotto

Aldershof *et al.* (1999) proposed a probabilistic stable matching mechanism, called employment by lotto that can be considered to be a random serial dictatorship on the set of stable matchings. A first agent is drawn randomly and can discard all stable matchings in which he/she is not matched to his/her best partner (possibly him-/herself) in a stable matching. Note that now the first agent is matched to the same partner in all remaining stable matchings. Exclude the first agent and his/her partner from the set of agents and randomly choose the next agent who can discard all stable matchings in which he/she is not matched to his/her best partner in the reduced set of stable matchings. Continue with this sequential reduction of the set of stable matchings until it is reduced to a singleton. Using all possible sequences of agents, this mechanism induces a probability distribution on the set of stable matchings. The associated probabilistic matching mechanism of this probabilistic sequential dictatorship mechanism equals employment by lotto.

Aldershof *et al.*'s (1999) original definition of employment by lotto combines a proposal algorithm with a refining process of the set of linear inequalities that describe the set of stable matchings.

The Employment by Lotto Algorithm

As mentioned before, we opt for a different description of the procedure than Aldershof *et al.* (1999); see also Klaus and Klijn (2004a). The description of employment by lotto as a probabilistic sequential dictatorship mechanism on the set of stable matchings enables us to avoid introducing further notation and technicalities.

Employment by Lotto (EL) Algorithm

Input: A marriage market (M, W, P). Set $N_1 := N$, $S_1 := S(P)$, and $t := 1$.

Step t: Choose an agent i_t from N_t at random.
Match agent i_t to his/her most preferred match $ch(i_t)$ among $\{j : j = \mu(i_t) \text{ for some } \mu \in S_t\}$.
If $N_t \backslash \{i_t, ch(i_t)\} = \emptyset$, then stop and define $\{EL(P)\} := S_t$. Otherwise set
$N_{t+1} := N_t \backslash \{i_t, ch(i_t)\}$, $S_{t+1} := S_t \backslash \{\mu \in S_t : \mu(i_t) \neq ch(i_t)\}$, and go to Step $t := t + 1$.

Recall that $|M| = a$ and $|W| = b$. It is easy to see that the algorithm ends in a finite number r ($\max\{a, b\} \leq r \leq a + b$) of steps that only depends on the preferences (this follows from Theorem 2.2). The outcome is a random stable matching $EL(P) \in S(P)$, generated by a sequence of agents $(i_1, \ldots, i_r)$. Let Q be the set of such sequences and let $q = |Q|$. Moreover, for any $\mu \in S(P)$, let $Q_\mu \subseteq Q$ be the (possibly empty) set of sequences that lead to μ. Denote $q_\mu = |Q_\mu|$. Note that if $a = b$ and if all men and women are mutually acceptable, then $r = a$ and $q = 2a \cdot (2a - 2) \cdot \ldots \cdot 2$.

The employment by lotto algorithm induces in a natural way a probability distribution $\mathcal{P} = \{p_\mu\}_{\mu \in S(P)}$ over the set of stable matchings: for any $\mu \in S(P)$, the probability that $EL(P) = \mu$ equals $p_\mu = \frac{q_\mu}{q}$. The following example shows that a stable matching that constitutes an *endstate* compromise between contrary preferences on both sides of the market may never result from employment by lotto.[1]

Example 3.1 Employment by lotto may never find an endstate compromise
Let (M, W, P) with $a = b = 3$ and P listed below. The three stable matchings for this market are listed below as well.

Preferences					Stable Matchings			
$P(m_1) =$	$\underline{w_1}$	$\widetilde{w_2}$	$\overline{w_3}$	m_1	$\overline{\mu} =$	w_3	w_2	w_1
$P(m_2) =$	$\underline{w_3}$	$\widetilde{w_1}$	$\overline{w_2}$	m_2	$\widetilde{\mu} =$	w_2	w_1	w_3
$P(m_3) =$	$\underline{w_2}$	$\widetilde{w_3}$	$\overline{w_1}$	m_3	$\underline{\mu} =$	w_1	w_3	w_2
$P(w_1) =$	$\overline{m_3}$	$\widetilde{m_2}$	$\underline{m_1}$	w_1				
$P(w_2) =$	$\overline{m_2}$	$\widetilde{m_1}$	$\underline{m_3}$	w_2				
$P(w_3) =$	$\overline{m_1}$	$\widetilde{m_3}$	$\underline{m_2}$	w_3				

[1] Aldershof *et al.* (1999) observe that if a stable matching μ does not match any agent to his/her man/woman optimal match, then $p_\mu = 0$. More precisely, if for all $i \in N$ it holds that $\mu_M(i) \neq \mu(i) \neq \mu_W(i)$, then $p_\mu = 0$. Klaus and Klijn (2004a) use an extension of the marriage market in Example 3.1 to prove that the converse is not true, *i.e.*, $p_\mu = 0$ does not necessarily imply that for all $i \in N$, $\mu_M(i) \neq \mu(i) \neq \mu_W(i)$.

Note that in matching $\underline{\mu}$ all men are matched to their most preferred match and all women are matched to their least preferred match ($\underline{\mu} = \mu_M$ is underlined at preference profile P)[2]. Matching $\overline{\mu}$ establishes the other extreme: all women are matched to their most preferred match and all man are matched to their least preferred match ($\overline{\mu} = \mu_W$). At matching $\widetilde{\mu}$ agents are matched neither to their most, nor to their least preferred match. In fact, at $\widetilde{\mu}$ all agents are matched to their second choice, which is why we consider $\widetilde{\mu}$ to be an endstate compromise in this situation. We depict the corresponding lattice in Figure 1. The nodes denote the stable matchings and the first number in each series is the corresponding probability resulting from employment by lotto (the other two numbers are probabilities from other random matching mechanisms that we discuss later). The solid arcs denote comparability or unanimity on each side of the market. For instance $\overline{\mu} \to \widetilde{\mu}$ in Figure 1 means that all men weakly prefer their matches at $\widetilde{\mu}$ to their matches at $\overline{\mu}$ and all women weakly prefer their matches at $\overline{\mu}$ to their matches at $\widetilde{\mu}$ (*i.e.*, $\overline{\mu} \vee_M \widetilde{\mu} = \widetilde{\mu}$).

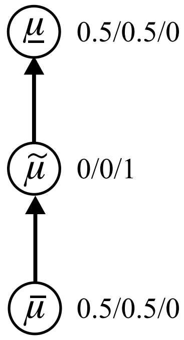

Figure 1: Lattice of Example 3.1

It is easy to check that whenever agent i_1 in the EL algorithm is a man, then $EL(P) = \mu_M$, and whenever agent i_1 in the EL algorithm is a woman, then $EL(P) = \mu_W$. So, $p_{\mu_M} = \frac{1}{2} = p_{\mu_W}$. Hence, for the endstate compromise matching $\widetilde{\mu}$, $p_{\widetilde{\mu}} = 0$. ◇

Finally, one might think that the employment by lotto algorithm is equivalent to the following procedure: first pick an agent i_1 at random, match i_1 to $ch(i_1)$, and remove i_1 and $ch(i_1)$ from the marriage market and the preference lists of the remaining agents. Repeat this procedure with the reduced marriage market, *etc.* Unfortunately, this procedure may not find a stable matching since, for instance, $ch(i_1)$ and $ch(i_2)$ thus obtained may form a blocking pair for the resulting matching. We demonstrate this using the marriage market introduced in Example 3.1. Suppose that m_1 first chooses w_1. In the reduced market there are two stable matchings: at μ_1 men m_2, m_3 are matched to w_3, w_2 and at μ_2 men m_2, m_3 are matched to w_2, w_3. Next, assume that w_2 can choose in the reduced market. Since w_2 prefers her match at μ_2 over her match at μ_1, the resulting matching for the original market matches men m_1, m_2, m_3 to w_1, w_2, w_3. However, this matching is not stable ((m_2, w_1) is a blocking pair).

Hence, in general it is necessary to calculate the complete set of stable matchings.

Properties of Employment by Lotto

We discuss two properties that set employment by lotto apart from the second procedurally fair matching mechanism that we consider in Section 3.2. First, we explain that employment

[2]Matchings $\overline{\mu}$ and $\widetilde{\mu}$ are marked in a similar way.

by lotto is based on a strong information requirement. Next, we point out that the probability distributions obtained by employment by lotto do not depend on agents that are matched to the same partner in all stable matchings.

Complete Information needed: As mentioned before, in order to apply employment by lotto it is necessary to calculate the set of stable matchings. From an informational point of view that means that a central planner or all agents need complete information of preferences.

We call an agent that is matched to the same partner (including being single) at all stable matchings a *dummy agent*. We call a probabilistic stable matching mechanism *independent of dummy agents* if dummy agents have no influence on the final probability distribution in the following sense. Delete all dummy agents from the original set of agents and apply the matching mechanism to the obtained reduced marriage market. Then, the probabilities for the remaining agents do not change. In order to formalize this property, we need some notation. Let (M, W, P) be a marriage market and let $D \subseteq N$ be the set of all dummy agents. Then $M\backslash D$ denotes all men that are not dummy agents, $W\backslash D$ denotes all women that are not dummy agents, and $P_{N\backslash D} = (P(i)_{N\backslash D})_{i\in N\backslash D}$ denotes the profile of reduced preferences induced by $(P(i))_{i\in N\backslash D}$. Formally, for all $i \in M\backslash D$ and all $j, k \in \{i\} \cup W\backslash D$, if $j \succeq_i k$ at $P(i)$, then $j \succeq_i k$ at $P(i)_{N\backslash D}$. (Similarly for $i \in W\backslash D$.) Then, after eliminating all dummy agents, we obtain the reduced marriage market $(M\backslash D, W\backslash D, P_{N\backslash D})$. Note that there exists a one-to-one mapping between matchings in $S(P)$ and $S(P_{N\backslash D})$: by eliminating dummy agents from a matching $\mu \in S(P)$ we obtain a matching $\mu_{N\backslash D} \in S(P_{N\backslash D})$, and *vice versa*, by adding dummy agents with their respective matches to a matching $\mu_{N\backslash D} \in S(P_{N\backslash D})$ we obtain a matching $\mu \in S(P)$.

Independence of Dummy Agents: Let (M, W, P) be a marriage market and $\tilde{\mathcal{P}}$ the probability distribution on the corresponding set of stable matchings induced by a probabilistic stable matching mechanism, that is, for all matchings $\mu \in S(P)$, $\tilde{\mathcal{P}}(\mu)$ denotes the probability that matching μ is chosen. Similarly, for the reduced marriage market $(M\backslash D, W\backslash D, P_{N\backslash D})$, $\tilde{\mathcal{P}}(\mu_{N\backslash D})$ denotes the probability that the reduced matching $\mu_{N\backslash D}$ is chosen.

Then, the matching mechanism satisfies *independence of dummy agents for* (M, W, P) if and only if for all matchings $\mu \in S(P)$, $\tilde{\mathcal{P}}(\mu) = \tilde{\mathcal{P}}(\mu_{N\backslash D})$. The matching mechanism satisfies *independence of dummy agents* if it satisfies independence of dummy agents for all marriage markets.

Since in the employment by lotto algorithm a dummy agent will never reduce the set of remaining stable matchings, it is easy to see that employment by lotto satisfies independence of dummy agents.

We finish the discussion of employment by lotto with two illustrative examples which we will also discuss in Section 3.3. For any stable matching lattice we depict, the first number labelling a stable matching μ equals p_μ.

Example 3.2 Let (M, W, P) with $a = b = 5$ and P the preferences given by Table 1 in the Appendix.[3] The set of stable matchings is depicted in Table 1 as well. Note that $\mu_M = \mu_6$ and $\mu_W = \mu_1$. We depict the corresponding lattice in Figure 2. The solid arcs denote again comparability or unanimity on each side of the market. Dotted edges denote incomparability or disagreement on each side of the market. For instance $\mu_4 \cdots \mu_5$ in Figure 2 means that there is disagreement among the men (women) about which matching is better (for instance, $\mu_5(m_2) \succ_{m_2} \mu_4(m_2)$, but $\mu_4(m_4) \succ_{m_4} \mu_5(m_4)$). ◇

[3] We complete the preferences of a marriage market taken from Blair (1984).

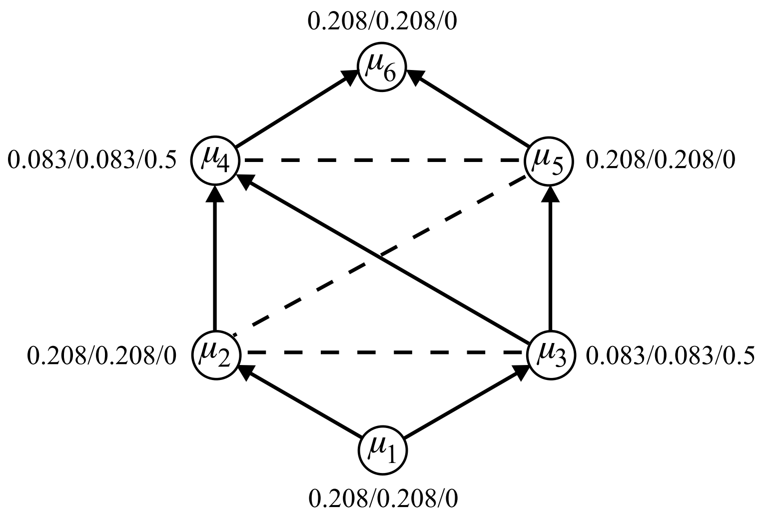

Figure 2: Lattice of Example 3.2

Example 3.3 Let (M, W, P) with $a = b = 4$ and P the preferences given by Table 2 in the Appendix. The set of stable matchings is depicted in Table 2 as well. We depict the corresponding lattice in Figure 3. $\diamond$

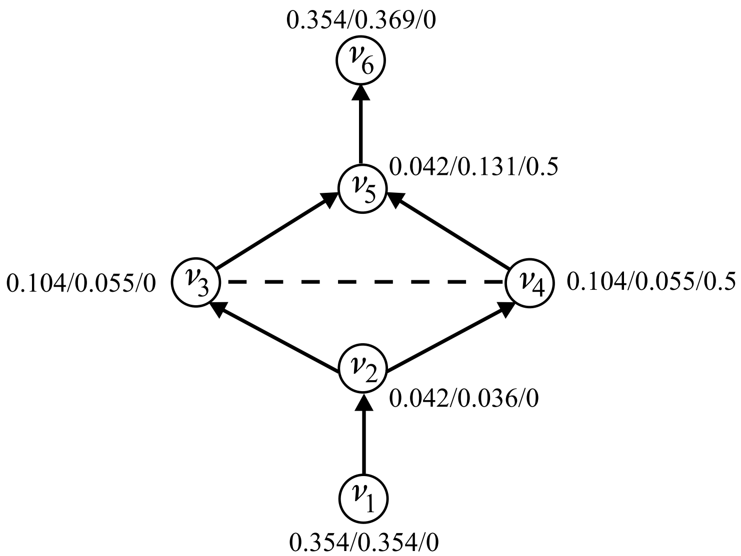

Figure 3: Lattice of Example 3.3

3.2 Procedural Fairness: the Random Order Mechanism

Ma (1996) described the random order mechanism, which is based on Roth and Vande Vate's (1990) random paths to stability. The basic idea is as follows. Imagine an empty room with one entrance. At the beginning, all agents are waiting outside. At each step of the algorithm, one agent is chosen randomly and invited to enter. Before an agent enters, the matching in the room is stable. However, once an agent enters the room, the existing matching in the room may become unstable, meaning that the new agent can form a blocking pair with another agent that already is present in the room. By satisfying this (and possible subsequent) blocking pair(s) in a certain way (described below in full detail) a new stable matching including the entering agent is obtained for the marriage market in the room. Since at each step a new agent enters the room and no agent leaves the room, the final outcome is a stable matching for the original marriage market. Using all possible sequences of agents, this mechanism induces a probability distribution on the set of stable matchings. The associated probabilistic matching mechanism equals the random order mechanism.

The Random Order Mechanism

We first give a formal description of the random order mechanism.

Random Order (RO) Mechanism

Input: A marriage market (M, W, P).
Set $R_0 := \emptyset$, μ_0 such that for all $i \in N$, $\mu_0(i) = i$, and $t := 1$.

Step *t*: Choose an agent i_t from $N \backslash R_{t-1}$ at random. Set $R_t := R_{t-1} \cup \{i_t\}$.
Suppose $i_t = w \in W$. (Otherwise replace w by m in Step t.)

Stable Room Procedure

Case (i) There exists no blocking pair (m, w) for μ_{t-1} with $m \in R_t$:
Stop if $t = n$ and define $RO(P) := \mu_{t-1}$. Otherwise set $\mu_t = \mu_{t-1}$ and go to Step $t := t + 1$.

Case (ii) There exists a blocking pair (m, w) for μ_{t-1} with $m \in R_t$:
Choose the blocking pair (m^*, w) for μ_{t-1} with $m^* \in R_t$ that w prefers most.
If $\mu_{t-1}(m^*) = m^*$, then define μ_t such that $\mu_t(w) := m^*$, $\mu_t(m^*) := w$, and for all $i \in N \backslash \{w, m^*\}$, $\mu_t(i) := \mu_{t-1}(i)$. Stop if $t = n$ and define $RO(P) := \mu_t$. Otherwise go to Step $t := t + 1$.
If $\mu_{t-1}(m^*) = w' \in W$, then redefine $\mu_{t-1}(w) := m^*$, $\mu_{t-1}(m^*) := w$, $\mu_{t-1}(w') := w'$, and for all $i \in N \backslash \{w, m^*, w'\}$, $\mu_{t-1}(i) := \mu_{t-1}(i)$. Set $w := w'$, and repeat the Stable Room Procedure.

It is not difficult to see that the algorithm ends in exactly n steps. The outcome is a random stable matching $RO(P) \in S(P)$, generated by a sequence of agents $(i_1, \ldots, i_n)$. The set of possible sequences of agents equals the set of permutations of all agents denoted by Q^*. Hence, $|Q^*| = n!$. Moreover, for any $\mu \in S(P)$, let $Q^*_\mu \subseteq Q^*$ be the (possibly empty) set of sequences that lead to μ. Denote $q^*_\mu = |Q^*_\mu|$.

The random order mechanism induces in a natural way a probability distribution $\mathcal{P}^*$ over the set of stable matchings: for any $\mu \in S(P)$, the probability that $RO(P) = \mu$ equals $p^*_\mu = \frac{q^*_\mu}{n!}$. For any stable matching lattice we depict, the second number labelling a stable matching μ equals p^*_μ.

Note that, similarly as employment by lotto, the random order mechanism never chooses the "endstate compromise" matching $\tilde{\mu}$ in Example 3.1.

Properties of the Random Order Mechanism

We compare the random order mechanism with employment by lotto, using the same properties as in Section 3.1.

No Complete Information needed: An important advantage of the random order mechanism over employment by lotto is that it is not necessary to calculate the set of stable matchings beforehand. In order to be a part in the random order mechanism, each agent only needs to know his/her own preferences.

The following example shows however that the random order mechanism fails to satisfy independence of dummy agents.

Example 3.4 The random order mechanism does not satisfy independence of dummy agents. Let (M, W, P) with $a = b = 3$ and P the preferences given below.

Preferences					Stable Matchings			
$P(m_1) =$	w_1	w_2	w_3	m_1	$\mu_W =$	w_2	w_1	w_3
$P(m_2) =$	w_2	w_1	w_3	m_2	$\mu_M =$	w_1	w_2	w_3
$P(m_3) =$	w_3	w_2	w_1	m_3				
$P(w_1) =$	m_2	m_1	m_3	w_1				
$P(w_2) =$	m_1	m_3	m_2	w_2				
$P(w_3) =$	m_3	m_2	m_1	w_3				

Some calculations give $(p^*_{\mu_M}, p^*_{\mu_W}) = (\frac{5}{12}, \frac{7}{12})$.
After elimination of the two dummy agents m_3 and w_3, we obtain the marriage market $(\hat{M}, \hat{W}, \hat{P})$ with $a = b = 2$ and $\hat{P}$ the preferences given below.

Preferences				Stable Matchings		
$\hat{P}(m_1) =$	w_1	w_2	m_1	$\hat{\mu}_W =$	w_2	w_1
$\hat{P}(m_2) =$	w_2	w_1	m_1	$\hat{\mu}_M =$	w_1	w_2
$\hat{P}(w_1) =$	m_2	m_1	w_1			
$\hat{P}(w_2) =$	m_1	m_2	w_2			

Some calculations give $(p^*_{\hat{\mu}_M}, p^*_{\hat{\mu}_W}) = (\frac{1}{2}, \frac{1}{2}) \neq (\frac{5}{12}, \frac{7}{12})$. Hence, the random order mechanism violates independence of dummy agents. $\diamond$

Ma (1996) showed that the random order mechanism may not reach all stable matchings. Although Ma's (1996) proof of this result is correct, we show in Example 3.5 that the probability distribution he obtained in addition was not correct. In fact, Ma's (1996) verification of probabilities would have been correct under independence of dummy agents.

Example 3.5 Let (M, W, P) with $a = b = 4$ and P the preferences given by Table 3 in the Appendix.[4] We depict the corresponding lattice of stable matchings in Figure 4. Ma (1996) claimed that $(p^*_{\mu_1}, p^*_{\mu_2}, p^*_{\mu_3}, p^*_{\mu_4}, p^*_{\mu_5}, p^*_{\mu_6}, p^*_{\mu_7}, p^*_{\mu_8}, p^*_{\mu_9}, p^*_{\mu_{10}}) = (\frac{1}{4}, \frac{1}{8}, \frac{1}{8}, 0, 0, 0, 0, \frac{1}{8}, \frac{1}{8}, \frac{1}{4})$, but it is clear from Figure 4 that this is not true. Note however that EL does give these probabilities. A proof that Ma's (1996) claim on the probabilities in this example is wrong (that is, $\mathcal{P}^* \neq \mathcal{P}$) that does not rely on our computational results can be found in Klaus and Klijn (2004b). $\diamond$

Cechlárová (2002) extended Ma's result showing that for any marriage market the only matchings that may be obtained are those that assign to at least one agent his/her best stable partner. One of the open problems Cechlárová (2002, p. 4) mentioned is that "... it is not clear whether for each of those not excluded it is possible to find a suitable order of players [agents]

[4]This is a marriage market taken from Knuth (1976).

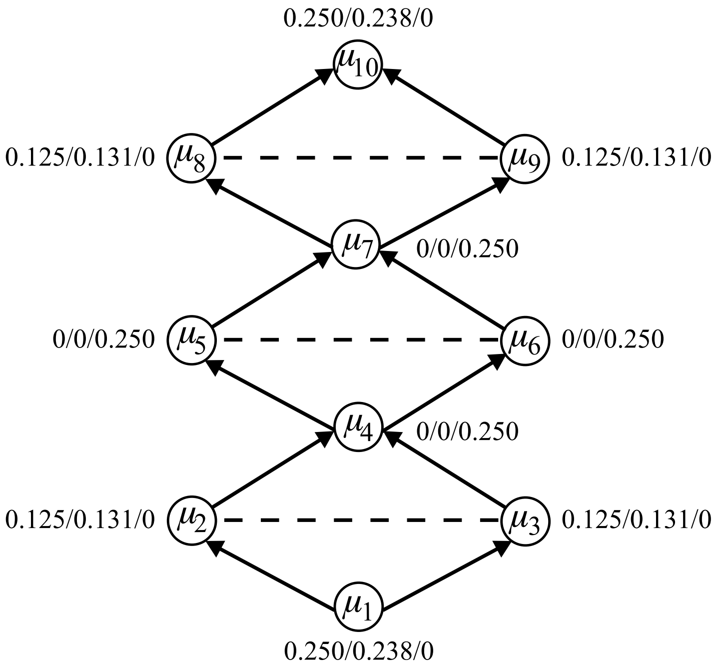

Figure 4: Lattice of Example 3.5

to get it." We answer this question by showing that a stable matching where some agents are matched to their best stable partner may never result from the random order mechanism. Consider the marriage market in Example 3.1 and add two agents m_4, w_4 such that for all $i = 1, 2, 3, 4$, $m_4 \succ_{m_4} w_i$, $w_4 \succ_{w_4} m_i$, and m_4, w_4 are placed anywhere in the preferences of the other agents. Now the previously stable matching $\widetilde{\mu} = w_2,\ w_1,\ w_3$ for the market in Example 3.1 extends to the stable matching $\widetilde{\mu} = w_2,\ w_1,\ w_3,\ w_4$ for the extended market. It is not difficult to see, however, that $p^*_{\widetilde{\mu}} = 0$, *i.e.*, it is not possible to find a suitable order of agents to reach $\widetilde{\mu}$ using the random order mechanism.

3.3 A Hybrid between Procedural Fairness and Endstate Justice: The Equitable Random Order Mechanism

Romero-Medina (2002) adapted the random order mechanism in order to limit the set of options available for each agent, trying to avoid in this way the inherent favoritism of optimal matchings. As a result, he mixes aspects of procedural and endstate fairness.

Since the description of his algorithm would be a bit tedious and we only discuss briefly the differences between the three mechanisms in a few examples, we refer the reader to Romero-Medina (2002) for its definition. In fact, Romero-Medina (2002) defined the algorithm for a fixed order of the agents and only in his final remarks suggested an extension by randomizing the order of the agents. Henceforth, we will call this extension the equitable random order mechanism.

For any marriage market (M, W, P) and any $\mu \in S(P)$, let $\bar{p}_\mu$ be the probability that μ is the outcome of the equitable random order (ERO) mechanism. For any stable matching lattice we depict, the third number labelling a stable matching μ equals $\bar{p}_\mu$.

In contrast to employment by lotto and the random order mechanism, the ERO mechanism chooses the "endstate compromise" matching $\tilde{\mu}$ in Example 3.1 not only with positive probability, but in fact with probability one. In the classical Example 3.5 the ERO mechanism demonstrates again nicely its avoidance of optimal matchings. The same occurs in Examples 3.2 and 3.3, although here probabilities seem to be split more arbitrarily.

In the example below we show that already for $a = b = 3$ the three mechanisms may give completely different and somewhat surprising outcomes. More specifically, it shows that the ERO mechanism may not always choose a probabilistic solution that "endstate compromises" between both sides of the market (as it did in Example 3.1): unlike the other two mechanisms, here the ERO mechanism always chooses the woman optimal matching μ_W.

Example 3.6 Recall that for the matching market in Example 3.4 there are two stable matchings $\mu_M = w_1, w_2, w_3$ and $\mu_W = w_2, w_1, w_3$. From Example 3.4, $(p^*_{\mu_M}, p^*_{\mu_W}) = (\frac{5}{12}, \frac{7}{12})$. Some calculations give $(p_{\mu_M}, p_{\mu_W}) = (\frac{1}{2}, \frac{1}{2})$ and $(\bar{p}_{\mu_M}, \bar{p}_{\mu_W}) = (0, 1)$. Note that the equitable random order mechanism fails to avoid the favoritism of one of the optimal matchings (μ_W). In contrast, the order two mechanisms, employment by lotto and the random order mechanism, spread probability over the two stable matchings, albeit in a slightly different way. $\diamond$

4 Concluding Remarks

In Sections 1 and 3 we have explained why in many two-sided matching markets it is not possible to implement a stable matching that at the same time does not favor one side of the market over the other. In a recent paper Klaus and Klijn (2004c) propose so-called median stable matchings as one way of choosing an endstate compromise. However, in order to guarantee a unique median matching, the number of stable matchings has to be odd.

In this article, we focus on another important aspect of fairness – procedural fairness – that has been known to affect many real life situations such as workplace regulations, family law, and general conflict resolution. We follow Rawls's (1971) notion of pure procedural justice and argue that in the absence of an objective criterion on the fairness of outcomes/end-states, the fairness of the procedure will lead to (procedurally) fair outcomes.

The most commonly known and applied stable matching mechanism is Gale and Shapley's (1962) deferred acceptance algorithm. In this algorithm, loosely speaking, only one side of the market can make offers while the other side can only accept or reject offers. Equivalently, the deferred acceptance algorithm can be formulated as an algorithm in which first all agents of one side of the market move and then all agents of the other side (Roth and van de Vate, 1990, Section 3). Since one side of the market has a "last mover advantage" that guarantees their best possible stable matching, clearly the deferred acceptance algorithm is neither procedurally nor endstate fair. By choosing the sequence of agents in a matching algorithm randomly such that agents' probabilities to move at a certain point in the algorithm are all the same, we introduce procedural fairness.

Apart from modelling procedural fairness for two-sided matching markets, we identify two known matching mechanisms as procedurally fair: employment by lotto and the random order mechanism (Sections 3.1 and 3.2). We try to complete the understanding of these stable matching mechanisms by giving various examples of matching markets and the probability distributions that are induced by employment by lotto and the random order mechanism. In order to understand better both mechanisms, we also identify two properties that differentiate them. Employment by lotto is based on complete information on the set of stable matchings, but it is independent of dummy agents. For the random order mechanism the reverse holds true: to participate in the random order algorithm each agent only needs to know his/her preferences, but dummy agents can influence the final outcome. Hence, these two properties allow for a clear distinction between the two mechanisms and a mechanism designer may use the complete information criterion or independence of dummy agents to choose between the two procedurally fair and stable matching mechanisms.

A Appendix

Preferences							Stable Matchings					
$P(m_1) =$	w_1	w_3	w_2	w_4	w_5	m_1	$\mu_1 =$	w_3	w_1	w_2	w_5	w_4
$P(m_2) =$	w_2	w_3	w_1	w_4	w_5	m_2	$\mu_2 =$	w_3	w_1	w_2	w_4	w_5
$P(m_3) =$	w_3	w_2	w_1	w_4	w_5	m_3	$\mu_3 =$	w_1	w_3	w_2	w_5	w_4
$P(m_4) =$	w_4	w_5	w_1	w_2	w_3	m_4	$\mu_4 =$	w_1	w_3	w_2	w_4	w_5
$P(m_5) =$	w_5	w_4	w_1	w_2	w_3	m_5	$\mu_5 =$	w_1	w_2	w_3	w_5	w_4
$P(w_1) =$	m_2	m_1	m_3	m_4	m_5	w_1	$\mu_6 =$	w_1	w_2	w_3	w_4	w_5
$P(w_2) =$	m_3	m_2	m_1	m_4	m_5	w_2						
$P(w_3) =$	m_1	m_2	m_3	m_4	m_5	w_3						
$P(w_4) =$	m_5	m_4	m_1	m_2	m_3	w_4						
$P(w_5) =$	m_4	m_5	m_1	m_2	m_3	w_5						

Table 1: Preferences and stable matchings of Example 3.2

Preferences						Stable Matchings				
$P(m_1) =$	w_1	w_2	w_4	w_3	m_1	$\nu_1 =$	w_3	w_4	w_1	w_2
$P(m_2) =$	w_2	w_1	w_3	w_4	m_2	$\nu_2 =$	w_4	w_3	w_1	w_2
$P(m_3) =$	w_3	w_4	w_1	w_2	m_3	$\nu_3 =$	w_4	w_1	w_3	w_2
$P(m_4) =$	w_4	w_3	w_1	w_2	m_4	$\nu_4 =$	w_2	w_3	w_1	w_4
$P(w_1) =$	m_3	m_2	m_1	m_4	w_1	$\nu_5 =$	w_2	w_1	w_3	w_4
$P(w_2) =$	m_4	m_1	m_2	m_3	w_2	$\nu_6 =$	w_1	w_2	w_3	w_4
$P(w_3) =$	m_1	m_2	m_3	m_4	w_3					
$P(w_4) =$	m_2	m_1	m_4	m_3	w_4					

Table 2: Preferences and stable matchings of Example 3.3

Preferences						Stable Matchings				
$P(m_1) =$	w_1	w_2	w_3	w_4	m_1	$\mu_1 =$	w_4	w_3	w_2	w_1
$P(m_2) =$	w_2	w_1	w_4	w_3	m_2	$\mu_2 =$	w_4	w_3	w_1	w_2
$P(m_3) =$	w_3	w_4	w_1	w_2	m_3	$\mu_3 =$	w_3	w_4	w_2	w_1
$P(m_4) =$	w_4	w_3	w_2	w_1	m_4	$\mu_4 =$	w_3	w_4	w_1	w_2
$P(w_1) =$	m_4	m_3	m_2	m_1	w_1	$\mu_5 =$	w_3	w_1	w_4	w_2
$P(w_2) =$	m_3	m_4	m_1	m_2	w_2	$\mu_6 =$	w_2	w_4	w_1	w_3
$P(w_3) =$	m_2	m_1	m_4	m_3	w_3	$\mu_7 =$	w_2	w_1	w_4	w_3
$P(w_4) =$	m_1	m_2	m_3	m_4	w_4	$\mu_8 =$	w_2	w_1	w_3	w_4
						$\mu_9 =$	w_1	w_2	w_4	w_3
						$\mu_{10} =$	w_1	w_2	w_3	w_4

Table 3: Preferences and stable matchings of Example 3.5

References

Aldershof, B., Carducci, O.M., and Lorenc, D.C. (1999) "Refined Inequalities for Stable Marriage," *Constraints* **4**, 281-292.

Anand, P. (2001) "Procedural Fairness in Economic and Social Choice: Evidence from a Survey of Voters," *Journal of Economic Psychology* **22**, 247-270.

Blair, C. (1984) "Every Finite Distributive Lattice is a Set of Stable Matchings," *Journal of Combinatorial Theory* (Series A) **37**, 353-356.

Bolton, G.E., Brandts, J., Ockenfels, A. (2004) "Fair Procedures: Evidence from Games Involving Lotteries," *The Economic Journal*, forthcoming.

Cechlárová, K. (2002) "Randomized Matching Mechanism Revisited," mimeo, Institute of Mathematics, P.J. Šafárik University, Slovakia.

Gale, D. and Shapley, L.S. (1962) "College Admissions and the Stability of Marriage," *American Mathematical Monthly* **69**, 9-15.

Klaus, B. and Klijn, F. (2004a) "Employment by Lotto Revisited," UFAE and IAE Working Paper 627-04, Universitat Autònoma de Barcelona.

Klaus, B. and Klijn, F. (2004b) "Corrigendum to "On Randomized Matching Mechanisms" [Economic Theory 8(1996)377-381]," UFAE and IAE Working Paper 628-04, Universitat Autònoma de Barcelona.

Klaus, B. and Klijn, F. (2004c) "Median Stable Matching for College Admission," mimeo, Universitat Autònoma de Barcelona.

Knuth, D.E. (1976) Marriages Stables. Montreal: Les Presses de l'Université Montreal.

Ma, J. (1996) "On Randomized Matching Mechanisms," *Economic Theory* **8**, 377-381.

Masarani, F. and Gokturk, S.S. (1989) "On the Existence of Fair Matching Algorithms," *Theory and Decision* **26**, 305-322.

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

Moulin, H. (2003) Fair Division and Collective Welfare. Cambridge, Massachusetts: MIT Press.

Rawls, J. (1971) A Theory of Justice. Cambridge, Massachusetts: Harvard University Press.

Romero-Medina, A. (2002) "Equitable Selection in Bilateral Matching Markets," mimeo, Universidad Carlos III de Madrid.

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

Roth, A.E. and Sotomayor, M.A.O. (1990) Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis. Econometric Society Monograph Series. New York: Cambridge University Press.

Roth, A.E. and Vande Vate, J.H. (1990) "Random Paths to Stability in Two-Sided Matching," *Econometrica* **58**, 1475-1480.