# SEQUENTIAL ENTRY IN MANY-TO-ONE MATCHING MARKETS

ELETTE BOYLE AND FEDERICO ECHENIQUE

ABSTRACT. We study sequential bargaining in many-to-one matching markets. We show that there is an advantage to entering late in the market, and that the last agent to enter the market will receive his or her best partner in a stable matching, extending the results of Blum and Rothblum (2002) and Cechlárová (2002) for the marriage model. We also discuss the relation between sequential bargaining and a possible alternative formulation based on the NTU Shapley value.

## 1. INTRODUCTION

We study sequential bargaining in many-to-one matching markets. Consider a two-sided matching market, such as a labor market. Suppose that agents sequentially enter the market, and that each time an agent arrives, the market adjusts to produce a new stable matching. Following Blum and Rothblum (2002) and Cechlárová (2002), we study a simplified model of sequential entry, in which participants enter the market one at a time, in some arbitrary given order. When an agent enters the market, there is an adjustment from the existing provisional matching to a new one. The entrant proposes to match with her most preferred partner. This can leave a previously paired agent unmatched, in which case this agent would propose to her most preferred partner, and so on.

We show that sequential entry results in an advantage to entering late into the market. Other factors remaining equal, it is always better to enter later than earlier. Further, the last agent to enter receives his/her best possible outcome of all stable outcomes for the market. Our results generalize similar results obtained for the one-to-one (marriage) market by Blum and Rothblum (2002) and Cechlárová (2002). Biró, Cechlárová, and Fleiner (2007) present a different generalization, to a model which includes the roommate problem.

We also show that sequential entry violates a basic monotonicity axiom: one agent may rise in others' preferences but receive, on average, worse outcomes. Monotonicity is behind standard solutions for random-order bargaining in other environments (Young, 1985). So sequential entry in matching seems rather special.

We work with a separable version of the many-to-one model, the "college admissions" model. We reduce the college admissions model to a one-to-one model by treating each position in a college as an individual agent, each of which can match to at most one student. This isomorphism between models is nontrivial in our case. The ordering of identical college positions yields an additional degree of freedom in the market that is lost when entry is sequential. We show, however, that this difficulty can be overcome.

The sequential entry model can be thought of as a random-order bargaining solution, which suggests comparisons with the Shapley value. In Section 4 we discuss the possible definitions of a Shapley value for matching markets, and show that it is generally not well defined. Sequential entry is always well defined, but developing a theory of random sequential entry seems difficult.

We should mention that Ma (1996) and Klaus and Klijn (2006) also study sequential entry. They present examples to show that not all stable matchings can be obtained under sequential entry, and to discuss the validity of certain axioms. Other related papers are Cantala (2004) and Kojima and Ünver (2007).

In Section 2 we present the model. Section 3 contains our results on sequential entry for the many-to-one model. We discuss the Shapley value in Section 4, and coalitional monotonicity in Section 5.

## 2. The Model

We present the standard marriage and college-admissions model. See Roth and Sotomayor (1990) for a detailed exposition of matching markets. The marriage model plays a central role in this study since our results utilize a reduction of the college-admissions model to the marriage model, and because our results for the Shapley value hold for the marriage model.

2.1. **Marriage model.** We first describe the one-to-one marriage model. Let $M$ and $W$ be finite, disjoint sets. We will refer to the elements of $M$ as *men* and the elements of $W$ as *women*. A generic element of $M \cup W$ is an *agent*.

A *preference relation* is a linear, transitive, and antisymmetric binary relation. A preference relation for a man $m \in M$, denoted $>_m$,

is understood to be over the set $W \cup \{m\}$. Similarly, $>_w$ denotes a preference relation over $M \cup \{w\}$. A *preference profile* is a list $P$ of preference relations for men and women, i.e.

$$P = \left( (>_m)_{m \in M}, (>_w)_{w \in W} \right).$$

We will assume that no man or woman is indifferent over two different partners; i.e., if $w \neq w'$, then either $w >_m w'$ or $w' >_m w$. Preferences with this property are called *strict*. Denote by $\geq_m$ the weak relation associated to $>_m$. That is, $w' \geq_m w$ if either $w' = w$ or $w' >_m w$. The definition of $\geq_w$ is analogous.

The triple $(M, W, P)$ describes an instance of the marriage model. We shall often consider instances obtained by deleting some agents from $(M, W, P)$: we denote by $(M, W, P; M', W')$ the instance $(M', W', P')$, with $M' \subseteq M$, $W' \subseteq W$, and where $P'$ is obtained from $P$ by restricting agents' preferences to the relevant subsets of potential partners.

A *matching* is a function $\mu : M \cup W \to M \cup W$ such that, for all $m \in M$ and $w \in W$, $\mu(m) \in W \cup \{m\}$ and $\mu(w) \in M \cup \{w\}$, and $m = \mu(w)$ if and only if $w = \mu(m)$. If $w = \mu(m)$ we say that $w$ and $m$ are *matched* in $\mu$. If $a = \mu(a)$, for $a \in M \cup W$, we say that $a$ is *single*. Say that an agent is *acceptable* to $a$ if $a$ prefers to match with the agent over being single.

A matching $\mu$ is defined to be *individually rational* if, for any $m$ and $w$, $\mu(m) \geq_m m$ and $\mu(w) \geq_w w$; i.e. all matched agents have acceptable matches. Say that a pair $(w, m)$ *blocks* $\mu$ if $w \neq \mu(m)$, $w >_m \mu(m)$, and $m >_w \mu(w)$. A matching is *stable* if it is individually rational and there is no pair that blocks it. Denote by $S(M, W, P)$ the set of all stable matchings in the market $(M, W, P)$.

2.2. **College admissions.** Let $\mathbf{C}$ and $S$ be finite, disjoint sets of agents. We call the elements of $\mathbf{C}$ *colleges* and the elements of $S$ *students*. Suppose that each college $C \in \mathbf{C}$ has a *quota* $q_C \geq 0$ of available *positions*. College $C$ can then accept at most $q_C$ students, and the remaining positions will remain unfilled. Thus, the potential student bodies for college $C$ are sets $A \cup B$, where $|A \cup B| = q_C$, $A \subseteq S$ and $B$ has $q_C - |A|$ copies of the element $C$ of $\mathbf{C}$ (corresponding to the unfilled positions). Formally speaking, $A \cup B$ is an unordered list; see Roth and Sotomayor (1990). Let $D_C$ be the set of all potential student bodies for $C$ and endow $C$ with a preference relation $>_C$ over $D_C$. We assume that $>_C$ satisfies the property that, for any $D \in D_C$, if $a \in D$ and $a' \notin D$, then

$$\left( (D \cup \{a'\}) \setminus \{a\} \right) >_C D \text{ if and only if } \{a'\} >_C \{a\}.$$

This property is called *responsiveness*.

Each student $s \in S$ is similarly endowed with a preference relation $\geq_s$ over $\mathbf{C} \cup \{s\}$. Let $P$ be the list of all colleges' and students' preferences. Then the data on a college-admissions market is collected in $(\mathbf{C}, S, P)$.

A *many-to-one matching* is a function $\mu$ on $\mathbf{C} \cup S$ such that, for all $C \in \mathbf{C}$ and $s \in S$, $\mu(C) \in D_C$, $\mu(s) \in \mathbf{C} \cup \{s\}$, and $s \in \mu(C)$ if and only if $\mu(s) = C$.

A matching $\mu$ is *individually rational* if $\mu(s) = C$ implies that $C >_s s$ and $\{s\} >_C \{C\}$. A matching is *stable* if it is individually rational and there is no pair $(C, s)$ such that $C >_s \mu(s)$ and $\{s\} >_C \{a\}$ for some $a \in \mu(C)$.

We use a (well known) reduction of college admissions to marriage markets. The reduction works by considering each available position in a college as a separate agent. In order to maintain strict preferences, the first position of a college is set to be preferred by all students over the second position of the college, over the third, and so on (Roth and Sotomayor, 1990).

Formally, given an instance of the college-admission model, $(\mathbf{C}, S, P)$, we construct an instance of the marriage model, $(M, W, P)$ by setting $M = S$; taking $W$ to be the set containing $q_C$ copies, $c_1, \ldots c_{q_C}$ of college $C$ for each $C \in \mathbf{C}$; and defining preferences of men and women such that each $c_j$ has the same preferences as $C$ over singleton students (now men), and each $s \in M$ has preferences $>_s$ over copies of different colleges, satisfying for each $C$, for all $s \in M$, $c_i >_s c_j$ if and only if $i < j$. We use the notation $c \in C$ to denote one of the copies of $C \in \mathbf{C}$.

2.3. **Sequential entry.** We present a model of sequential entry due to Blum and Rothblum (2002) (following ideas of Roth and Vande Vate (1990) and Blum, Roth, and Rothblum (1997)). The setup is a variant on the classic Gale-Shapley stable matching algorithm. We first present the model within the marriage model, and then its extension to the college-admissions model.

The Gale-Shapley algorithm (Gale and Shapley, 1962) takes as input an instance of the marriage model and gives a matching as output. It starts with all men being "active" and all women being provisionally single, and it iterates the following subroutine: Let each active man propose to match with his preferred woman, of those women he has not already proposed to. Each woman selects her preferred man between her provisional match and those who proposed to her during the round. At the conclusion of this round, take each woman to be provisionally matched to the man she selects. A man becomes active after a round if he was rejected by a woman during the round and there are still acceptable women to whom he has not yet proposed. Exit the

subroutine. The algorithm terminates when there are no more active men.

Gale and Shapley show that the outcome $\mu$ of the algorithm is a stable matching. In fact it is the *men-optimal stable matching*, in the sense that all men prefer their match in $\mu$ over their match in any other stable matching, and all women have the opposite preference. Similarly, the version of the algorithm with women proposing to men produces the *women-optimal stable matching*.

We study a straightforward modification of the Gale-Shapley algorithm: Take as input to the subroutine an instance $(M, W, P)$, and a matching $\mu$ in $(M, W, P; M \setminus \{m\}, W)$ with $m \in M$. The output will be a matching in $(M, W, P)$. Start with $m$ being the only "active" agent and all women provisionally matched to their matches in $\mu$. Repeat the subroutine in the Gale-Shapley algorithm (with only male agents proposing) until no men are active. Denote the resulting matching by $h(M, W, P; m, \mu)$. Analogously, $h(M, W, P; w, \mu)$ would be the matching that results from the above algorithm with women proposing, and agent $w$ starting as active.

An *ordering* of the agents is an ordered list of the elements of $M \cup W$. We denote by $R$ the set of all orderings, and by $\rho$ a generic ordering. Let $\rho^k$ denote the ordered list of the first $k$ elements of $\rho$, $\rho_M^k$ the set of men in $\rho^k$ and $\rho_W^k$ the set of women in $\rho^k$. Let $\rho_k$ denote the $k$th agent in $\rho$, equivalently, the last element of $\rho^k$.

Let $\rho \in R$ and consider the recursive procedure where we let $\mu^{k+1} = h(\rho_M^k, \rho_W^k, P; \rho_k, \mu^k)$, starting from the matching $\mu^1(\rho_1) = \rho_1$. Let $f(\rho)$ be the matching that results after the last agent has been added.

For the college-admissions market $(\mathbf{C}, S, P)$, we let $R$ be the set of all orderings of students and positions in colleges. So the orderings in $R$ are the orderings of $M \cup W$ in the marriage market associated to $(\mathbf{C}, S, P)$. We define $f(\rho)$ as the matching that results from the procedure we have described, but applied to the associated marriage model.

## 3. Main Results

We show that there is a strong late-mover advantage in sequential bargaining over many-to-one matching outcomes. Indeed, in the extreme case, the final agent to enter the market will receive his or her best partner in a stable matching. The result in item 1 of Theorem 1 was shown for the marriage model by Blum and Rothblum (2002) and Cechlárová (2002), and generalized to a single-party roommate model by Biró, Cechlárová, and Fleiner (2007). A weaker version of the

statement in item 2 is shown Blum and Rothblum (2002) and Biró, Cechlárová, and Fleiner (2007) for the marriage model.

**Theorem 1.** *Let $(\mathbf{C}, S, P)$ be an instance of the college admissions model. Let $\rho \in R$ be an ordering of the students and college positions. Then*

- (1) *$f(\rho)$ is a stable matching and the last agent $a$ in $\rho$ receives his/her preferred partner in any stable matching; if $a$ is the position of a college $C$, then $C$ receives his/her preferred student body in any stable matching.*
- (2) *Let $\rho'$ be an ordering in which agent $b$ occurs later than in $\rho$, and all agents that follow $b$'s position in $\rho'$ coincide in both orderings. If $b = s \in S$ is a student, then $f(\rho')(s) \geq_s f(\rho)(s)$; if $b$ is a position of a college $C \in \mathbf{C}$, then $f(\rho')(C) \geq_C f(\rho)(C)$.*

The proof of Theorem 1 requires some preliminary results. The first are two lemmas from Blum and Rothblum (2002).[1] These hold for the marriage market. Lemma 4 is a new result, which holds on the college admissions model.

**Lemma 2.** *Consider an instance of the marriage market. For any stable matching $\mu$ in the market $(M, W, P; M \setminus \{m\}, W)$, the resulting matching $h(M, W, P; m, \mu)$ is stable in $(M, W, P)$, and is preferred by all the men to any other stable matching in which the women are at least as well off as in $\mu$. Similarly for $h(M, W, P; w, \mu)$.*

**Lemma 3.** *Consider an instance of the marriage market. Let $\mu = f(\rho)$ and $\mu' = f(\rho')$, for two orderings $\rho$ and $\rho'$. If $\mu'(b) \geq_b \mu(b)$ for some agent $b$, then this relation is preserved when an additional player is added onto the end of both orderings. That is, if $\hat{\rho}$ and $\hat{\rho}'$ are the orderings obtained by adding the same agent at the end of $\rho$ and $\rho'$, we have $f(\hat{\rho}')(b) \geq_b f(\hat{\rho})(b)$.*

We omit the proofs of lemmas 2 and 3.

**Lemma 4.** *Consider an instance of the college-admissions model $(\mathbf{C}, S, P)$ and fix a matching $\mu$. When a college position enters the market, the resulting matching is the college-optimal ($\mathbf{C}$-optimal) stable matching such that no student is worse off than in $\mu$; when a student enters the market, the resulting matching is the student-optimal (S-optimal) stable matching such that no college is worse off than in $\mu$.*

---

[1]Lemma 2 is Theorem A.6, and Lemma 3 is Lemma 2.4 in Blum and Rothblum (2002).

*Proof.* From Lemma 2, the claim holds for students and individual college positions in the one-to-one market model. The **C**-optimal matching of a set of stable matchings $\mathcal{S}$ (when it exists) is the matching $\mu_{\mathbf{C}}$ such that $\mu_{\mathbf{C}}(C) \geq_C \mu(C) \; \forall C \in \mathbf{C}, \forall \mu \in \mathcal{S}$. By Theorem 5.27 in Roth and Sotomayor (1990), this condition is equivalent to $\mu_{\mathbf{C}}(c_i) \geq_{c_i} \mu(c_i) \; \forall$ college positions $c_i$, $\forall \mu \in \mathcal{S}$. This implies the college position-optimal matching s.t. no student is worse off is precisely the **C**-optimal matching s.t. no student is worse off. Consider now the second assertion. No college worse off in matching $\mu_1$ than in $\mu_2$ means that $\mu_1(C) \geq_C \mu_2(C) \; \forall C \in \mathbf{C}$. Again from Theorem 5.27 in Roth and Sotomayor (1990), this occurs if and only if $\mu_1(c_i) \geq_{c_i} \mu_2(c_i) \; \forall$ college positions $c_i$, implying the set of stable matchings such that no college is worse off is precisely the set of stable matchings such that no college position is worse off. $\qquad\square$

**Lemma 5.** *Consider the one-to-one model of the many-to-one market with students and individual college positions. If two positions $c_i$ and $c_j$ of the same college $C$ are switched in an ordering $\rho$, then the resulting stable matching is unaffected.*

*Proof.* Suppose $c_i$ and $c_j$ appear in the $m$th and $n$th positions of the ordering $\rho_1$, respectively; we may assume without loss of generality that $m < n$. Let $\rho_2$ be the ordering with $c_i$ and $c_j$ switched. For every $i \geq 1$, let $(\mathbf{C}_1^{(i)}, S_1^{(i)}, P_1^{(i)})$ and $(\mathbf{C}_2^{(i)}, S_2^{(i)}, P_2^{(i)})$ be the $i$th restriction submarkets containing the first $i$ players appearing in $\rho_1$ and $\rho_2$, and maintaining all relevant preferences. Take $\pi : (C \cap \mathbf{C}_2^{(i)}) \to (C \cap \mathbf{C}_1^{(i)})$ to be the unique bijection such that $c_k <_s c_l \Leftrightarrow \pi(c_k) <_s \pi(c_l)$ for each student $s \in S$ (Note that, from the construction of the market model, $c_k <_{s_0} c_l$ for some $s_0 \in S$ and $c_k, c_l \in C$ iff $c_k <_s c_l \; \forall s \in S$). Extend $\pi$ to the map $\phi : \mathbf{C}_2^{(i)} \to \mathbf{C}_1^{(i)}$ by $\phi(c) = c \; \forall c \notin C$. By the construction of the market model, a student prefers all positions of a college over all positions of a less preferred college; thus, the relation $c_k <_s c_l \Leftrightarrow \phi(c_k) <_s \phi(c_l)$ holds for all students $s \in S$ and for all positions $c$.

Let $\mu_1^{(i)}$ be a matching in $(\mathbf{C}_1^{(i)}, S_1^{(i)}, P_1^{(i)})$, and define the matching $\mu_2^{(i)}$ in $(\mathbf{C}_2^{(i)}, S_2^{(i)}, P_2^{(i)})$ by $\mu_2^{(i)}(c) = \mu_1^{(i)}(\phi(c)) \; \forall c \in \mathbf{C}_1^{(i)}$ and $\mu_2^{(i)}(s) = s \; \forall s$ s.t. $\mu_1^{(i)}(s) = s$. Show that $\mu_1^{(i)}$ is a stable matching iff $\mu_2^{(i)}$ is a stable matching. Suppose, to the contrary, there exists a blocking pair $(\hat{c}, \hat{s})$ in $\mu_2^{(i)}$. Then, $\hat{s} >_{\hat{c}} \mu_2^{(i)}(\hat{c})$ and $\hat{c} >_{\hat{s}} \mu_2^{(i)}(\hat{s})$. By properties of $\phi$, we have that $\phi(\hat{c}) >_{\hat{s}} \phi(\mu_2^{(i)}(\hat{s}))$. From the construction of $\mu_2^{(i)}$, $\mu_2^{(i)}(\mu_2^{(i)}(\hat{s})) = \mu_1^{(i)}(\phi(\mu_2^{(i)}(\hat{s})))$; since we also have trivially that $\mu_2^{(i)}(\mu_2^{(i)}(\hat{s})) = \hat{s} =$

$\mu_1^{(i)}(\mu_1^{(i)}(\hat{s}))$ and as $\mu_1^{(i)}$ is an injective mapping, this implies $\phi(\mu_2^{(i)}(\hat{s})) = \mu_1^{(i)}(\hat{s})$. So we have $\phi(\hat{c}) >_{\hat{s}} \mu_1^{(i)}(\hat{s})$. Now, $\hat{s} >_{\hat{c}} \mu_2^{(i)}(\hat{c}) = \mu_1^{(i)}(\phi(\hat{c}))$. If $\hat{c} \notin C$, then $\phi(c) = c$, implying $\hat{s} >_{\phi(\hat{c})} \mu_1^{(i)}(\phi(\hat{c}))$. If $\hat{c} \in C$, then $\phi(\hat{c}) \in C$ as well. Since all positions of college $C$ share the same preferences, we again have that $\hat{s} >_{\phi(\hat{c})} \mu_1^{(i)}(\phi(\hat{c}))$. Therefore, $(\hat{s}, \phi(\hat{c}))$ form a blocking pair in $\mu_1^{(i)}$, contradicting the stability of $\mu_1^{(i)}$. An identical argument, utilizing $\phi^{-1}$ in the place of $\phi$, can be used to show that $\mu_2^{(i)}$ stable implies $\mu_1^{(i)}$ is stable. Thus, one matching is stable iff the other matching is stable.

This result implies that for each $i$, the set of stable matchings in $(\mathbf{C}_1^{(i)}, S_1^{(i)}, P_1^{(i)})$ is in bijection with those in $(\mathbf{C}_2^{(i)}, S_2^{(i)}, P_2^{(i)})$, and are the same up to the labeling of the positions of college C. We wish to show that when a new agent enters both markets, $\mu_1^{(i)}$ is the resulting matching in $(\mathbf{C}_1^{(i)}, S_1^{(i)}, P_1^{(i)})$ iff $\mu_2^{(i)}$ is the resulting matching in $(\mathbf{C}_2^{(i)}, S_2^{(i)}, P_2^{(i)})$.

Suppose a new college position enters both markets. For each student $s \in S$, $\mu_2^{(i+1)}(s) \geq_s \mu_2^{(i)}(s) \Leftrightarrow \phi(\mu_2^{(i+1)}(s)) \geq_s \phi(\mu_2^{(i)}(s))$ (from properties of $\phi$) $\Leftrightarrow \mu_1^{(i+1)}(s) \geq_s \mu_1^{(i)}(s)$ (since $\mu_1^{(j)}(s) = \phi(\mu_2^{(j)}(s))$ from above). That is, no student is worse off in $\mu_1^{(i+1)}$ than the previous matching $(\mu_1^{(i)})$ iff no student is worse off in the analogue matching $\mu_2^{(i+1)}$ than in $\mu_2^{(i)}$. Now, suppose a new student enters both markets. $\mu_2^{(i+1)}(c) \geq_c \mu_2^{(i)}(c) \Leftrightarrow \mu_1^{(i+1)}(\phi(c)) \geq_c \mu_1^{(i)}(\phi(c))$ (by definition of $\mu_2^{(i)}(s)$) $\Leftrightarrow \mu_1^{(i+1)}(\phi(c)) \geq_{\phi(c)} \mu_1^{(i)}(\phi(c))$ (since $c$ and $\phi(c)$ are positions of the same college and all positions of a college share the same preferences). As $\phi$ is a bijection between $\mathbf{C}_1^{(i)}$ and $\mathbf{C}_2^{(i)}$, this holds for all $c \in \mathbf{C}_2^{(i)}$ iff $\mu_1^{(i+1)}(c') \geq_{c'} \mu_1^{(i)}(c')$ for all $c' \in \mathbf{C}_1^{(i)}$. That is, no college position is worse off in $\mu_1^{(i+1)}$ than in the previous matching iff the same holds for $\mu_2^{(i+1)}$.

We now show that $\mu_2$ is $S$-optimal (resp, $\mathbf{C}$-optimal) of a set of stable matchings in $(\mathbf{C}_2^{(i)}, S_2^{(i)}, P_2^{(i)})$ iff $\mu_1$ is $S$-optimal (resp, $\mathbf{C}$-optimal) of the corresponding set of stable matchings in $(\mathbf{C}_1^{(i)}, S_1^{(i)}, P_1^{(i)})$. Consider two matchings $\mu_2$ and $\tilde{\mu}_2$ in $(\mathbf{C}_2^{(i)}, S_2^{(i)}, P_2^{(i)})$. $\mu_2(s) \geq_s \tilde{\mu}_2(s) \Leftrightarrow \phi(\mu_2(s)) \geq_s \phi(\tilde{\mu}_2(s)) \Leftrightarrow \mu_1(s) \geq_s \tilde{\mu}_1(s)$ in $(\mathbf{C}_1^{(i)}, S_1^{(i)}, P_1^{(i)})$. Similarly, $\mu_2(c) \geq_c \tilde{\mu}_2(c) \ \forall c \in \mathbf{C}_2^{(i)} \Leftrightarrow \mu_1(\phi(c)) \geq_c \tilde{\mu}_1(\phi(c)) \ \forall c \in \mathbf{C}_2^{(i)} \Leftrightarrow \mu_1(c') \geq_{c'} \tilde{\mu}_1(c') \ \forall c' \in \mathbf{C}_1^{(i)}$.

Therefore, by Lemma 4, whenever a student or a college position enters the market in $(\mathbf{C}_1^{(i)}, S_1^{(i)}, P_1^{(i)})$ and $(\mathbf{C}_2^{(i)}, S_2^{(i)}, P_2^{(i)})$, the resultant

matching in the former is $\mu_1^{(i+1)}$ if and only if the resulting matching in the latter is $\mu_2^{(i+1)}$. Continuing inductively, we see that the final matchings resulting from the two orderings $\rho_1$ and $\rho_2$ must be the same up to permutation of the positions of college $C$. However, in order for a matching $\mu$ to be stable, it must be that the most preferred student in $\mu(C)$ is paired with the first position in $C$, the second most preferred student with the second position in $C$, and so on. Hence, since both $\mu_1$ and $\mu_2$ are stable matchings, it must be that $\mu_1 = \mu_2$. $\qquad\square$

*Proof of Theorem 1.* We first prove item 1 using that the result is true for the marriage market associated to $(\mathbf{C}, S, P)$.

The result is immediate when a student is the last to enter. Suppose that the last agent in $\rho$ is a position of a college $C$. Let $\mu$ be the college-optimal stable matching in $(\mathbf{C}, S, P)$. By Lemma 4, $f(\rho)$ is stable. If $\mu(C) >_C f(\rho)(C)$ there must be a position $c \in C$ such that $s$ is $c$'s partner in $\mu$, $s'$ is $c$'s partner in $f(\rho)$, and $s >_c s'$ (see Roth and Sotomayor (1990)). Now consider the order $\rho'$ which only differs from $\rho$ in that the last position of $\rho$ switches places with $c$. By Lemma 5, $f(\rho) = f(\rho')$, the fact that $c$ does not obtain its best partner in a stable matching contradicts the statement of Theorem 1 for the marriage market.

We now prove item 1. For the marriage market, our statement is more general than the one in Blum and Rothblum (2002), so we present a separate proof first for the marriage market.

Consider an instance of the marriage market, $(M, W, P)$. Let $n$ be the number of agents. Suppose agent $b$ is in the $k$th position in the ordering $\rho'$, so $b = \rho'_k$. By Lemma 2, then, $b$'s partner in the matching $f(\rho'^k)$ is his/her best stable partner in any stable matching in $(M, W, P; \rho'_M, \rho'_W)$. But since the agents coming after $b$ coincide for both $\rho$ and $\rho'$, the agents coming before $b$ also coincide. So $\rho_M = \rho'_M$ and $\rho_W = \rho'_W$ and hence $b$ prefers his/her partner in $f(\rho'^k)$ over his/her partner in $f(\rho^k)$.

Now proceed by induction, using Lemma 3: For each $i \geq k$, $f(\rho'^i)(b) \geq_b f(\rho^i)(b)$ implies that $f(\rho'^{i+1})(b) \geq_b f(\rho^{i+1})(b)$ implies that, as we are adding the same agent to both orderings.

We now prove the result for the college-admissions model. For $a \in S$, the result follows from the argument for the marriage market. Consider then $a = c_0$, a position in college $C$. By Theorem 1, $\mu_1(c_0) \geq_{c_0} \mu_2(c_0)$. By Theorem 5.27 in Roth and Sotomayor (1990), it remains to show that $\mu_1(c_j) \geq_{c_j} \mu_2(c_j) \ \forall j \neq 0$ such that $c_j \in C$. Suppose $c_0$ appears $n$-th in the ordering $\rho_1$. For every $i \geq n$, let $(\mathbf{C}^{(i)}, S^{(i)}, P^{(i)})$ be the $i$th restriction submarket containing only the first $i$ players appearing in

$\rho_1$ (and $\rho_2$), and maintaining all relevant preferences. Let $\mu_1^{(i)}$ and $\mu_2^{(i)}$ be the stable matchings resulting from the first $i$ players in $\rho_1$ and $\rho_2$ entering the market.

Suppose $c_j \in C$ appears $m$th in $\rho_1$, with $m \geq n$. From item 1 in Theorem 1, since $c_j$ appears last of the $m$ players in $\rho_1$, $c_j$ will receive the best possible stable student in the matching $\mu_1^{(m)}$. $\Rightarrow$ $\mu_1^{(m)}(c_j) \geq_{c_j} \mu_2^{(m)}(c_j)$. By Lemma 3 this relation is preserved as each of the following players enters the market. Continuing inductively, we have $\mu_1(c_j) \geq_{c_j} \mu_2(c_j)$. (Note that for $m > n$, this argument holds also in the opposite direction, implying that $\mu_1(c_j) = \mu_2(c_j)$.)

Suppose, then, $c_j \in C$ appears $l$th in $\rho_1$, with $l < n$. From Lemma 5, $\rho_1$ will yield the same stable matching as the ordering $\hat{\rho}_1$ with the positions of $c_0$ and $c_j$ switched. As $c_j$ appears later in $\hat{\rho}_1$ than $\rho_2$, we have by item 2 in Theorem 1, $\rho_2(c_j) \leq_{c_j} \hat{\rho}_1(c_j) = \rho_1(c_j)$.

It has been shown that $\mu_1(c_j) \geq_{c_j} \mu_2(c_j)$ $\forall j$ such that $c_j \in C$. Therefore, $\mu_1(C) \geq_C \mu_2(C)$. $\qquad\square$

## 4. The Shapley Value in Matching Markets

The discussion of sequential entry may suggest ideas of sequential random-order bargaining. It is natural to consider an alternative approach, using the Shapley value. In this section, we focus on the marriage model, and show that the Shapley value will typically not be well defined.

The Shapley value is the best-known prediction in random-order bargaining, and its extension to non-transferable utility (NTU) games (Shapley, 1969) is in principle applicable to matching problems. While the Shapley value for transferable-utility (TU) games is the result of a simple calculation, in NTU games it requires a fixed-point property, and is not guaranteed to exist for arbitrary NTU games (it can also have problematic properties, as illustrated in Roth (1980)).

As we show below, in the NTU games that result from a matching market, the Shapley value will not generally be well defined. Indeed, it does not seem to exist beyond a set of very symmetric situations. We view the problems with the Shapley value as one reason to be interested in the $f$ outcomes discussed in Section 3.

4.1. **Preliminaries.** A *transferable utility (TU) game* is a pair $(N, v)$ composed of a set of players $N$ and a real-valued function $v : 2^N \to \mathbb{R}$ on the power set of $N$, satisfying $v(\emptyset) = 0$. A *coalition* is a nonempty subset of $N$. For a coalition $S$ of players within $N$, $v(S)$ associates a value of worth to that particular coalition.

**Definition 6.** The *Shapley value* of a TU game $(N, v)$ is the vector whose $i$th component is given by

$$\phi_i(v) = \sum_{S \subseteq N} \frac{|S|!(n - |S| - 1)!}{n!} (v(S \cup \{i\}) - v(S)),$$

for each player $i \in N$.

A *non-transferable utility (NTU) game* is a pair $(N, V)$ composed of a set $N$ and a function $V$ taking each subset $S$ of $N$ to a subset of $\mathbb{R}^s$ (where $s = |S|$) satisfying the following conditions:

(i) $V(S) = \emptyset$ iff $S = \emptyset$
(ii) If $x = (x_1, \ldots, x_s) \in V(S)$ and $y = (y_1, \ldots, y_s) \in \mathbb{R}^s$ with $y_i \leq x_i$ for each $i$, then $y \in V(S)$
(iii) $V(S)$ is a closed subset of $\mathbb{R}^s$
(iv) $V(S) \cap (x + \mathbb{R}_+)$ is bounded for all $x \in \mathbb{R}^s$

The $s$-dimensional vectors in $V(S)$ represent possible ways of distributing the worth of the subset $S$ among its $s$ members. An example of an NTU game is a two-sided matching market, where players assign quantitative utility values to each possible partner.

We now define the Shapley value of a NTU game $(N, V)$ (Shapley, 1969).

**Definition 7.** For $\lambda \in \mathbb{R}^n$, let $v_\lambda(S) = sup\{\lambda^S \cdot x | x \in V(S)\}$, where $\lambda^S$ is the projection of $\lambda$ to the $s$-dimensional subspace corresponding to $S$. If this supremum is finite for all $S$, then $v_\lambda$ defines a TU game. A *Shapley value* of $(N, V)$ is a point $y$ in $\bar{V}(N)$ (the closure of the image of $N$ under $V$) such that there exists a positive vector $\lambda$ with $v_\lambda$ defined and $\lambda y = \phi(v_\lambda)$, where $\phi$ yields the Shapley value of the TU game $v_\lambda$.

See Aumann (1985) or Peleg and Sudhölter (2003) for an exposition.

4.2. **Matching Markets as NTU Games.** We define a NTU game to model the matching market as follows. Fix an instance of a matching market $(M, W, P)$ and fix *utilities* representing the preferences in $P$; so, for example, for a man $m$ we have $u_m : W \cup \{m\} \to \mathbb{R}$ such that $w \geq_m w'$ if and only if $u_m(w) \geq u_m(m')$. Without loss of generality, we choose utilities so that $u_m(m) = 0$ and $u_w(w) = 0$; that is, being single gives zero utility for all $m \in M$ and $w \in W$. Note that, in this setup with numerical utilities, we have the potential for quantitative preference effects instead of the outcome relying solely on relative preferences. For instance, an agent now has the capacity to rank two players nearly the same, or to really like or dislike a potential match.

Let $N = M \cup W$ be the set of all players in the matching market. For each coalition $S \subseteq N$ of players, consider all the possible matchings

between the players involved. Each player has a utility associated with being paired to the other players, so each matching yields a vector $y$ of utilities corresponding to the involved players. Take $V(S)$ to be the set of all vectors $x \in \mathbb{R}^s$ such that $x_i \leq y_i$ for all components $i$ of the vectors for at least one utility vector $y$. Note that $V(\emptyset) = \emptyset$ and for all $S \neq \emptyset$, there exists at least one matching (specifically, the matching corresponding to the zero vector, in which every player remains single), implying $V(S) \neq \emptyset$. By construction, the defined function $V$ also satisfies conditions (ii), (iii), and (iv) above.

A matching market is *uneven* if the number of men differs from the number of women. We also say that *all agents are acceptable* if, for every agent, being single receives lower utility than being with anyone of the opposite sex.

**Proposition 8.** *If $(M, W, P)$ is an uneven matching market in which all agents are acceptable, a Shapley value does not exist for the NTU game induced by $(M, W, P)$.*

*Proof.* Let $(N, V)$ be the NTU game associated to $(M, W, P)$. Fix one $\lambda \in \mathbb{R}^n$. From the construction of (N,V), $v_\lambda(S \cup \{p\}) - v_\lambda(S) \geq 0$ for all players $p \in N$. That is, no player can decrease the value of a coalition when joining it. As all players are acceptable to at least one agent of the opposite party, $\forall p \in N, \exists q \in N$ such that $v_\lambda(\{p, q\}) - v_\lambda(\{q\}) > 0$. This implies $\phi(v_\lambda)_p > 0$ for each player $p$. However, since the parties are uneven, for any matching $\mu$ of the players of $N$, at least one player must remain single. Thus, $y \in V(N)$ implies $y_i \leq 0$ for some $i \in N$. This means that $\phi(v_\lambda) \notin V(N)$. Hence, since $\lambda$ was arbitrary, there cannot exist an NTU Shapley value in $V(N)$. $\square$

The proof of Proposition 8 suggests that an NTU Shapley value may not exist in any but the most symmetric matching situations. We proceed to show that, in the case that a Shapley value does exist, it must correspond to a valid matching of the players in the market. Further, in any market where there exists a single matching $\mu$ preferred over all other matchings by every player, the NTU Shapley value of the market is the utility vector corresponding to $\mu$.

**Proposition 9.** *If an NTU Shapley value exists in a one-to-one matching market, then it must correspond to the utility vector of a matching.*

*Proof.* Fix a matching market $(M, W, P)$ and let $(N, V)$ be the associated NTU game. Suppose $y \in \mathbb{R}^n$ is an NTU Shapley value of $(N, V)$, obtained as the TU Shapley value of $v_\lambda$. By efficiency of the TU Shapley value (e.g. see Peleg and Sudhölter (2003)), $\sum_i \phi(v_\lambda)_i = v_\lambda(N)$.

Now, by definition, $v_\lambda(N) = sup\{x \cdot \lambda | x \in V(N)\}$. Since $\phi(v_\lambda)_i = y_i\lambda_i$, we have $sup\{x \cdot \lambda | x \in V(N)\} = \sum_i y_i\lambda_i$. $\sum_i y_i\lambda_i = y \cdot \lambda$ by definition. Now, $V(N)$ is defined as the smallest comprehensive set in $\mathbb{R}^n$ that contains the utility vectors of all possible matchings. Thus, $y$ must itself be the utility vector corresponding to a valid matching. $\square$

**Proposition 10.** *Suppose there is a matching $\mu$ in $(M, W, P)$ such that, for all matchings $\nu$ and agents $a$, $\mu(a) \geq_a \nu(a)$. Then there is exactly one Shapley value of the associated NTU game, and it is the vector of utilities corresponding to $\mu$.*

*Proof.* Let $y_0$ be the utility vector corresponding to the matching $\mu$. As $\mu(a) \geq_a \nu(a)$ for all possible matchings $\nu$, the NTU game $(N, V)$ associated with the matching market $(M, W, P)$ is convex. As a consequence, the set of NTU Shapley values must be nonempty (see Theorem 6.3.1 in Ichiishi (1983)). From the proof of Proposition 9, if $y$ is a Shapley value, then $y \cdot \lambda = sup\{x \cdot \lambda | x \in V(N)\}$. In this case, as $V(N) = \{x \in \mathbb{R}^N | x \leq y_0\}$, this implies the only possible Shapley value is $y_0$ itself. Therefore, $y_0$ must be the unique NTU Shapley value of the market. $\square$

## 5. Coalitional monotonicity

The matching outcomes $f(\rho)$ can, by analogy with the Shapley value, be used to calculate "average" utilities when we think of each ordering $\rho$ as equally likely. For instance, in the case where players have different partners in each of the stable matchings, the average utilities will simply be the weighted average of the men-optimal and women-optimal solutions, with weighting ratio equal to the ratio of men to women in the game.

While in general the average utility values may not correspond to an actual matching, they could serve as a measure of bargaining power each player brings to the market. We now show, however, that the resulting average utilities are not monotonic with respect to increased utility opportunities. This finding is probably surprising, as the source of increased utility comes from agents rising in the preferences of the opposite sex. It is also pessimistic about the possibility of axiomatizing $f(\rho)$ using a variation of the systems of axioms used for the Shapley value (Young, 1985).

We note that the failure of coalitional monotonicity is in a sense inherent to matching problems: increasing in the preferences of an undesired partner is typically bad for an agent.

A *value* is a function $\phi$ mapping NTU games $(N, V)$ into vectors in $\mathbb{R}^n$. Say that a value satisfies *coalitional monotonicity* if, for all

players $p$ and for all functions $V$ and $W$, if it holds that both $V(S) \supseteq W(S) \ \forall S \subseteq N$ with $p \in S$ and $V(T) = W(T) \ \forall T \subseteq N$ with $p \notin T$, then it implies that $\phi_p(N, V) \geq \phi_p(N, W)$ (Young, 1985).

**Proposition 11.** *The value obtained by averaging $f(\rho)$ over all $\rho \in R$ does not satisfy coalitional monotonicity.*

*Proof.* Show by counterexample. Consider the six-player market $(\{m_1, m_2, m_3\}, \{w_1, w_2, w_3\}, P)$ where $P$ is given by the following utility representations:

| Man | $\emptyset$ | $w_1$ | $w_2$ | $w_3$ |
|-----|---|---|---|---|
| $m_1$ | 0 | 3 | 2 | 1 |
| $m_2$ | 0 | 2 | 3 | 1 |
| $m_3$ | 0 | 1 | 2 | 3 |

| Woman | $\emptyset$ | $m_1$ | $m_2$ | $m_3$ |
|-------|---|---|---|---|
| $w_1$ | 0 | 2 | 3 | 1 |
| $w_2$ | 0 | 2 | 1 | 3 |
| $w_3$ | 0 | 1 | 2 | 3 |

By averaging $f(\rho)$ over all $\rho \in R$ we obtain approximately the vector of utilities

$$(2.42, 2.42, 3.00, 2.58, 1.58, 3.00).\text{[2]}$$

Consider now if $w_3$ changes her preferences to prefer $m_1$ as depicted below:

| Man | $\emptyset$ | $w_1$ | $w_2$ | $w_3$ |
|-----|---|---|---|---|
| $m_1$ | 0 | 3 | 2 | 1 |
| $m_2$ | 0 | 2 | 3 | 1 |
| $m_3$ | 0 | 1 | 2 | 3 |

| Woman | $\emptyset$ | $m_1$ | $m_2$ | $m_3$ |
|-------|---|---|---|---|
| $w_1$ | 0 | 2 | 3 | 1 |
| $w_2$ | 0 | 2 | 1 | 3 |
| $w_3$ | 0 | **4** | 2 | 3 |

If modeled as a NTU game with value function $V$ (as described above), we see that $V(\{m_1, w_3\})$ will increase from $\{x \in \mathbb{R}^2 | x \leq (1, 1)\}$ to $\{x \in \mathbb{R}^2 | x \leq (1, 4)\}$ after this change. Also, for every coalition $S$ s.t. $m_1 \in S$, $V(S)$ will remain constant or increase, and every coalition $T$ with $m_1 \notin T$ will remain unaffected.

However, the new vector obtained by averaging $f(\rho)$ is

$$(2.00, 2.42, 2.58, 2.58, 2.00, 3.42).$$

Note that the component corresponding to $m_1$ decreased from 2.42 to 2.00. □

<center>REFERENCES</center>

AUMANN, R. J. (1985): "An Axiomatization of the Non-Transferable Utility Value," *Econometrica*, 53(3), 599–612.

---

[2]The `C++` code, written by Elette Boyle, for making these calculations can be downloaded from `http://www.hss.caltech.edu/~fede/wp/`

BIRÓ, P., K. CECHLÁROVÁ, AND T. FLEINER (2007): "The dynamics of stable matchings and half-matchings for the stable marriage and roommates problems," mimeo, Safarik University.

BLUM, Y., A. E. ROTH, AND U. G. ROTHBLUM (1997): "Vacancy Chains and Equilibration in Senior-Level Labor Markets," *Journal of Economic Theory*, 76(2), 362–411.

BLUM, Y., AND U. G. ROTHBLUM (2002): "Timing Is Everything and Marital Bliss," *Journal of Economic Theory*, 103(2), 429–443.

CANTALA, D. (2004): "Restabilizing matching markets at senior level," *Games and Economic Behavior*, 48(1), 1–17.

CECHLÁROVÁ, K. (2002): "Randomized matching mechanism revisited," mimeo, Safarik University.

GALE, D., AND L. S. SHAPLEY (1962): "College Admissions and the Stability of Marriage," *The American Mathematical Monthly*, 69(1), 9–15.

ICHIISHI, T. (1983): *Game Theory for Economic Analysis*. Academic Press, New York, NY.

KLAUS, B., AND F. KLIJN (2006): "Procedurally fair and stable matching," *Economic Theory*, 27, 431–447.

KOJIMA, F., AND U. ÜNVER (2007): "Random Paths to Pairwise Stability in Many-to-Many Matching Problems: A Study on Market Equilibration," Forthcoming in International Journal of Game Theory.

MA, J. (1996): "On Randomized Matching Mechanisms," *Economic Theory*, 8, 377–381.

PELEG, B., AND P. SUDHÖLTER (2003): *Introduction to the Theory of Cooperative Games*. Kluwer Academic Publishers, Boston.

ROTH, A., AND M. SOTOMAYOR (1990): *Two-sided Matching: A Study in Game-Theoretic Modelling and Analysis*, vol. 18 of *Econometric Society Monographs*. Cambridge University Press, Cambridge England.

ROTH, A. E. (1980): "Values for Games without Sidepayments: Some Difficulties with Current Concepts," *Econometrica*, 48(2), 457–465.

ROTH, A. E., AND J. H. VANDE VATE (1990): "Random Paths to Stability in Two-sided Matching," *Econometrica*, 58(6), 1475–1480.

SHAPLEY, L. S. (1969): "Utility Comparison and the Theory of Games," *La Décision*, pp. 251–263.

YOUNG, H. P. (1985): "Monotonic Solutions of Cooperative Games," *International Journal of Game Theory*, 14(2), 65–72.

Massachusetts Institute of Technology
  *E-mail address*: eboyle@mit.edu

Division of the Humanities and Social Sciences, California Institute
of Technology, Pasadena CA 91125
  *E-mail address*: fede@caltech.edu