Random Paths to Stability in the Roommate Problem*

Effrosyni Diamantoudi
University of Aarhus

Eiichi Miyagawa
Columbia University

Licun Xue
University of Aarhus

June 23, 2002
Revised: April 16, 2003

Abstract

This paper studies whether a sequence of myopic blockings leads to a stable matching in the roommate problem. We prove that if a stable matching exists and preferences are strict, then for any unstable matching, there exists a finite sequence of successive myopic blockings leading to a stable matching. This implies that, starting from any unstable matching, the process of allowing a randomly chosen blocking pair to form converges to a stable matching with probability one. This result generalizes those of Roth and Vande Vate (1990) and Chung (2000) under strict preferences.

JEL Classification: C62; C71; C78; D71

Keywords: Roommates problem; Marriage problem; Core convergence; Indirect domination; Cycle; Odd ring

*We thank Alvin Roth for pointing out a false statement in the previous version, John Vande Vate for detailed suggestions, Hideo Konishi for helpful conversations, and anonymous referees for helpful comments and suggestions. Parts of this work were conducted while the authors were visiting the Department of Economics and Wallis Institute of Political Economy at the University of Rochester, followed by Diamantoudi and Xue's visit of the Department d'Economia i d'Història at the Universitat Autònoma de Barcelona (UAB) and Miyagawa's visit of the Department of Economics at Kobe University. We would like to acknowledge these institutions for their hospitality and the seminar participants at UAB for helpful comments.

1 Introduction

This paper studies whether a decentralized process of successive myopic blockings leads to the core in the roommate problem. Knuth (1976) addresses the issue for the marriage problem and provides an example in which a sequence of blockings generates a cycle. That is, he constructs a cycle of matchings such that each matching is generated from the previous one by letting a blocking pair form.

On the other hand, Roth and Vande Vate (1990) answer the question in the affirmative for the marriage problem by showing that the process does converge to a stable matching if the blocking pairs are chosen appropriately at each step of the process. That is, they show that for any unstable matching, there *exists* a finite sequence of successive blockings leading to a stable matching. This result is interesting since it implies that if a blocking pair is chosen randomly and every blocking pair is chosen with a positive probability, then the random process converges to a stable matching with probability one.

Chung (2000) generalizes the result of Roth and Vande Vate (1990) to the roommate problem. Chung identifies a condition, called the "no odd rings" condition, that is sufficient for the existence of a stable matching when preferences are not necessarily strict.[1] Moreover, he shows that, under the same condition, the core convergence of Roth and Vande Vate extends to the roommate problem. Chung's result generalizes Roth and Vande Vate's since the "no odd rings" condition holds always in the marriage problem.

When preferences are strict, the "no odd rings" condition says that there exists no ordered subset of agents $(i_1, \ldots, i_K)$ such that $K \geq 3$ is odd and (subscript modulo K) $i_{k+1} \succ_{i_k} i_{k-1} \succ_{i_k} i_k$ for all $k \in \{1, \ldots, K\}$. The following four-agent example, taken from Chung (2000), shows that the "no odd rings" condition is not necessary for either the non-emptiness of the core or

[1] For the roommate problem, Gale and Shapley (1962) show that there exists a preference profile for which a stable matching does not exist. Tan (1991) identifies a necessary and sufficient condition for the existence of a stable roommate matching when preferences are strict.

convergence to the core.

$$\begin{aligned}
&2 \succ_1 1 \succ_1 \cdots \\
&1 \succ_2 3 \succ_2 4 \succ_2 2 \\
&4 \succ_3 2 \succ_3 3 \succ_3 1 \\
&2 \succ_4 3 \succ_4 4 \succ_4 1.
\end{aligned}$$

While an odd ring exists, i.e., $(2, 3, 4)$, there exists a stable roommate matching, $\mu = [\{1, 2\}, \{3, 4\}]$. Moreover, it is easy to see that starting from any other matching, there exists a sequence of blocking pairs leading to μ. Indeed, 1 and 2 are each other's top choices and they would block any matching that has them apart. Once pair $\{1, 2\}$ is formed, 3 and 4 will also get together (if not already) since they prefer it to being alone.

We show that in the roommate problem, when a stable matching exists and preferences are strict, the process of myopic blockings leads to a stable matching whether or not the "no odd rings" condition is satisfied. This result generalizes that of Chung (2000) (and Roth and Vande Vate (1990)) under the assumption of strict preferences, since the "no odd rings" condition is sufficient but not necessary for the existence of a stable matching. On the other hand, while our result requires strict preferences, Chung's holds with weak preferences as long as the "no odd rings" condition is satisfied. It should be noted that our result does not generalize to the roommate problem with indifferences. Indeed, Chung (2000) shows that convergence does not necessarily occur in the roommate problem when preferences are not strict and there exists an odd ring.

The convergence result does not easily extend to the case in which coalitions of any sizes can form, even when preferences are strict and satisfy reasonable restrictions. Counter-examples are given in Section 3.

There are a few papers that study the same issue in more abstract settings. Green (1974) and Feldman (1974) obtain convergence results for certain subclasses of NTU games, but their results do not apply to the roommate problem. Sengupta and Sengupta (1996) show that a similar convergence result holds for any TU game with non-empty core.

2 Main Result

We consider a *roommate problem* (Gale and Shapley, 1962), which is a list $(N, (\succcurlyeq_i)_{i \in N})$ where N is a finite set of agents and, for each $i \in N$, $\succcurlyeq_i$ is a complete and transitive preference relation defined over N. The strict preference associated with $\succcurlyeq_i$ is denoted by $\succ_i$. We limit ourselves to a roommate problem in which preferences are strict, i.e., $k \succcurlyeq_i j$ and $j \succcurlyeq_i k$ only if $k = j$. Thus, $k \succcurlyeq_i j$ means that either $k \succ_i j$ or $k = j$.

A *matching* is a function $\mu \colon N \to N$ such that for all $i, j \in N$, if $\mu(i) = j$, then $\mu(j) = i$. Here, $\mu(i)$ denotes the agent with whom agent i is matched. We allow $\mu(i) = i$, which means that agent i is alone. We sometimes write $\mu \succ_i \mu'$, which means $\mu(i) \succ_i \mu'(i)$. A *marriage problem* (Gale and Shapley, 1962) is a roommate problem $(N, (\succcurlyeq_i)_{i \in N})$ such that N is the union of two disjoint sets M and W, and each agent in M (respectively W) prefers being alone to being matched with any other agent in M (respectively W).

A matching μ is *blocked* by a pair $\{i, j\} \subseteq N$ (possibly $i = j$) if

$$j \succ_i \mu(i) \text{ and } i \succ_j \mu(j). \tag{1}$$

That is, i and j both prefer each other to their mates at μ. We allow $i = j$, in which case (1) means that $i \succ_i \mu(i)$, i.e., i prefers being alone to being matched with $\mu(i)$. When (1) holds, we call $\{i, j\}$ a *blocking pair* of μ. A matching is *stable* if there exists no blocking pair. A matching is *individually rational* if there exists no blocking pair $\{i, j\}$ with $i = j$.

Given a blocking pair $\{i, j\}$ of a matching μ, another matching μ' is obtained from μ by *satisfying* the pair if $\mu'(i) = j$ and for all $k \in N \setminus \{i, j\}$,

$$\mu'(k) = \begin{cases} k & \text{if } \mu(k) \in \{i, j\} \\ \mu(k) & \text{otherwise.} \end{cases}$$

That is, once i and j are matched, their mates (if any) at μ are alone in μ', and the other agents are matched as in μ.

The following is our main result.

Theorem 1. *Consider any roommate problem in which preferences are strict and a stable matching exists. Then for any unstable matching μ, there exists* *a finite sequence of matchings* ($\mu = \mu_1, \mu_2, \ldots, \mu_K$) *such that for any* $k \in \{1, 2, \ldots, K-1\}$, μ_{k+1} *is obtained from* μ_k *by satisfying a blocking pair of* μ_k *and* μ_K *is stable.*

Our proof differs significantly from those of Roth and Vande Vate (1990) and Chung (2000). Roth and Vande Vate's proof constructs a sequence of myopic blockings in such a way that it can be associated with an increasing (with respect to set inclusion) sequence of sets that contain no blocking pair. Each step of their construction uses the fact that there are two genders. Chung extends Roth and Vande Vate's results by adapting their process to the roommate problem under the "no odd rings" condition. The "no odd rings" condition enables him to construct a process where each agent can be labeled a certain gender.

Proof. Fix a stable matching μ'. Given any unstable matching μ, let $n(\mu)$ denote the number of pairs (including singletons) that are common to both μ and μ'. It suffices to show the following.

Claim. For any unstable matching μ, there exists a finite sequence of matchings ($\mu = \mu_1, \mu_2, \ldots, \mu_L$) such that for each $\ell \in \{1, 2, \ldots, L-1\}$, $\mu_{\ell+1}$ is obtained from μ_ℓ by satisfying a blocking pair of μ_ℓ and that $n(\mu_L) \geq n(\mu)+1$.

To prove the claim, take an unstable matching μ.

Step 1. If μ is not individually rational for some agents, we first let these agents (and their mates) become single, which can be done by a finite sequence of myopic blockings. Thus in what follows, we assume, without loss of generality, that μ is actually individually rational.

Step 2. The claim is trivial if μ is blocked by a pair that is matched under μ', since then satisfying the pair induces a matching μ_2 for which $n(\mu_2) \geq n(\mu)+1$.[2] Thus in what follows, we assume otherwise. The following summarizes the two assumptions on μ.

D1. Matching μ is individually rational and there exists no pair $\{i, j\} \subseteq N$ such that $\mu'(i) = j$, $\mu' \succ_i \mu$, and $\mu' \succ_j \mu$.

[2]The inequality can be strict. For example, if $\mu' = [\{1, 2\}, \{3\}]$ and $\mu = [\{1\}, \{2, 3\}]$, then satisfying $\{1, 2\}$ directly generates μ' and $n(\mu') = 2 = n(\mu) + 2$.

Since μ' is stable, the symmetric condition holds for μ', i.e.,

D2. Matching μ' is individually rational and there exists no pair $\{i, j\} \subseteq N$ such that $\mu(i) = j$, $\mu \succ_i \mu'$, and $\mu \succ_j \mu'$.

The symmetry between D1 and D2 simplifies the argument that follows.

Step 3. Let us define a function $f\colon N \to N$ by

$$f(i) = \begin{cases} \mu(i) & \text{if } \mu \succ_i \mu' \\ \mu'(i) & \text{otherwise.} \end{cases}$$

That is, $f(i)$ is whomever agent i prefers between $\mu(i)$ and $\mu'(i)$.

We now let each agent i "point" to $f(i)$. Since the number of agents is finite, there exists at least one "cycle."[3] A cycle is an ordered set of distinct agents $c = (i_1, i_2, \ldots, i_m)$ such that i_1 points to i_2, i_2 points to i_3, $\ldots$, and i_m points to i_1. If $\mu(i) = \mu'(i) = i$, then i alone forms a cycle of size 1. If $\mu(i) = \mu'(i) = j \neq i$, then $\{i, j\}$ forms a cycle of size 2.

The following 9-agent example may be helpful to see the construction.

$$\begin{gathered}
2 \succ_1 7 \succ_1 6 \succ_1 3 \succ_1 4 \succ_1 8 \succ_1 5 \succ_1 1 \succ_1 9 \\
5 \succ_2 3 \succ_2 1 \succ_2 4 \succ_2 8 \succ_2 7 \succ_2 6 \succ_2 2 \succ_2 9 \\
4 \succ_3 2 \succ_3 7 \succ_3 5 \succ_3 6 \succ_3 1 \succ_3 8 \succ_3 3 \succ_3 9 \\
8 \succ_4 5 \succ_4 3 \succ_4 6 \succ_4 1 \succ_4 2 \succ_4 7 \succ_4 4 \succ_4 9 \\
6 \succ_5 4 \succ_5 2 \succ_5 8 \succ_5 7 \succ_5 3 \succ_5 1 \succ_5 5 \succ_5 9 \\
1 \succ_6 8 \succ_6 5 \succ_6 7 \succ_6 3 \succ_6 4 \succ_6 2 \succ_6 6 \succ_6 9 \\
3 \succ_7 1 \succ_7 8 \succ_7 2 \succ_7 5 \succ_7 6 \succ_7 4 \succ_7 7 \succ_7 9 \\
7 \succ_8 6 \succ_8 4 \succ_8 1 \succ_8 2 \succ_8 5 \succ_8 3 \succ_8 8 \succ_8 9 \\
9 \succ_9 1 \succ_9 2 \succ_9 3 \succ_9 4 \succ_9 5 \succ_9 6 \succ_9 7 \succ_9 8.
\end{gathered}$$

In this example, one stable matching is $\mu' = [\{1, 2\}, \{3, 4\}, \{5, 6\}, \{7, 8\}, \{9\}]$. Let the initial matching be $\mu = [\{2, 3\}, \{4, 5\}, \{6, 1\}, \{7, 8\}, \{9\}]$, which is unstable and is blocked by $\{1, 7\}$. It can be verified that $\{1, 7\}$ is actually the

[3]This argument resembles one used in David Gale's *top trading cycles algorithm*, which identifies the unique core allocation in house allocation problems (Shapley and Scarf, 1974).

unique blocking pair of μ. Since $\{1,7\}$ is not matched under μ', it follows that D1 is satisfied. Applying function f to μ and μ' generates 3 cycles: $(1 \to 2 \to 3 \to 4 \to 5 \to 6 \to 1)$, $(7 \to 8 \to 7)$, and $(9 \to 9)$.

We now proceed to prove a few facts on f.

Step 4. For all $i \in N$,

$$f(i) = i \iff \mu(i) = \mu'(i) = i.$$

The "$\Longleftarrow$" part is mentioned above. The converse follows from the individual rationality of μ and μ' and the assumption that preferences are strict.

Step 5. For all $i \in N$, if $f(i) = j \neq i$, then

$$\begin{aligned} f(i) = \mu(i) &\Longrightarrow f(j) = \mu'(j) \\ f(i) = \mu'(i) &\Longrightarrow f(j) = \mu(j). \end{aligned}$$

To show the first part, suppose, by way of contradiction, that $f(j) = \mu(j) \neq \mu'(j)$. This implies $\mu \succ_j \mu'$. Furthermore, since $\mu'(j) \neq \mu(j) = i$, we have $\mu(i) \neq \mu'(i)$. This and $f(i) = \mu(i)$ imply $\mu \succ_i \mu'$. But then D2 is violated since i and j are matched under μ. The second part follows from a symmetric argument that leads to a violation of D1.

This step trivially implies that for all $i \in N$,

$$\begin{aligned} \mu \succ_i \mu' &\Longrightarrow \mu' \succ_{f(i)} \mu \\ \mu' \succ_i \mu &\Longrightarrow \mu \succ_{f(i)} \mu'. \end{aligned}$$

Step 6. For all $i \in N$,

$$f(f(i)) = i \iff \mu(i) = \mu'(i).$$

The "$\Longleftarrow$" part is mentioned prior to Step 4. To see the converse, note first that if $f(i) = i$, then the conclusion follows from Step 4. So, suppose $f(i) = j \neq i$. Without loss of generality, assume $j = \mu(i)$. Then, Step 5 implies $f(j) = \mu'(j)$. Since $f(j) = i$, we have $\mu'(j) = i$, and this together with $\mu(i) = j$ implies that i and j are matched with each other in both μ and μ'.

Step 7. For all $i \in N$, there exists $t \in \{1, 2, \ldots\}$ such that $f^t(i) = i$.[4] To see this, take any $i \in N$ and consider the sequence of agents $\sigma_i = (i, f^1(i), f^2(i), \ldots)$. Since the number of agents is finite, some agents appear more than once in this sequence. Let t be the minimum number for which $f^t(i) = f^m(i)$ for some $m < t$. We show that $f^t(i) = i$. Suppose, by way of contradiction, that $f^t(i) \neq i$. Then the sequence looks like

$$\sigma_i = (i = i_0, i_1, i_2, \ldots, i_{m-1}, i_m, i_{m+1}, \ldots, i_{t-1}, i_t = i_m, i_{m+1}, \ldots),$$

where $i_m \neq i$. By Step 4, $i_h \neq i_{h+1}$ for all h. By the minimality of t, all agents in $\{i, i_1, \ldots, i_{t-1}\}$ and in particular, agents i_{m-1}, i_m, and i_{t-1} are distinct. Since $f(i_{m-1}) = i_m = f(i_{t-1})$, agent i_m is matched with i_{m-1} in one of the matchings and with i_{t-1} in the other. Thus, $f(i_m)$ is either i_{m-1} or i_{t-1}. But then, in either case, $f(f(i_m)) = i_m$ and Step 6 implies $\mu(i_m) = \mu'(i_m)$, which is not possible since i_{m-1} and i_{t-1} are distinct.

This step shows that each agent belongs to a unique cycle. Thus, we let c_i denote the cycle that i belongs to and let $S_i \subseteq N$ denote the set of agents who belong to the cycle. Since c_i is a cycle, it follows that for all $i, j \in N$, if $j \in S_i$, then $S_j = S_i$. Thus $\{S_i\}_{i \in N}$ generates a partition of N.

Step 8.[5] For all $i \in N$, if $\mu(i) \neq \mu'(i)$, then $|S_i| \geq 4$ and $|S_i|$ is even. Indeed, if $\mu(i) \neq \mu'(i)$, then Steps 4 and 6 imply that $|S_i|$ is at least 3. The alternation property proved in Step 5 then implies that $|S_i|$ is even.

Step 9. To complete the proof, let $\{i, j\} \subseteq N$ (with $i \neq j$) be a blocking pair of μ. We first note that $\mu' \succ_h \mu$ for some $h \in \{i, j\}$. Indeed, if $\mu \succcurlyeq_h \mu'$ for all $h \in \{i, j\}$, then $\{i, j\}$ also blocks μ', in contradiction with the stability of μ'. Thus we assume, without loss of generality, that agent $i = 1$ prefers μ' to

[4] Here, $f^{m+1}(i) = f(f^m(i))$ for all $m \in \{1, 2, \ldots\}$ and $f^1(i) = f(i)$.

[5] Suppose that, in Steps 3–8, μ is also a stable matching. Then both D1 and D2 are satisfied; hence, Steps 3–8 apply to any pair μ, μ' of stable matchings. An implication of these steps is that if μ and μ' are stable and agent i prefers μ' to μ, then both $\mu'(i)$ and $\mu(i)$ prefer μ to μ'. This is a generalization of the *decomposition lemma* of Knuth (1976) (see also Roth and Sotomayor (1990)) to the roommate problem. In fact, there are similarities between our proof of Step 7 and Knuth's proof of his decomposition lemma for the marriage problem. A corollary of the lemma is that the set of agents who are single is the same in all stable matchings.

μ, and that the cycle that agent 1 belongs to is given by $c_1 = (1, 2, 3, \ldots, m)$. By Step 8, $m \geq 4$ and m is even. Since agent 1 prefers μ' to μ, Step 5 implies that for all $i \in \{1, \ldots, m\}$, $i+1 = \mu'(i)$ if i is odd, and $i+1 = \mu(i)$ if i is even. Matchings μ and μ' then look like

$$\begin{aligned} \mu' &= [\{1,2\}, \{3,4\}, \ldots, \{m-3, m-2\}, \{m-1, m\}, \ldots] \\ \mu &= [\{2,3\}, \{4,5\}, \ldots, \{m-2, m-1\}, \{m, 1\}, \ldots]. \end{aligned}$$

Let μ_2 denote the matching that is obtained from $\mu_1 \equiv \mu$ by satisfying $\{1, j\}$. Note that $j \neq 2$ since $f(2) = \mu(2) = 3$ and so 2 prefers 3 to 1. Thus, satisfying $\{1, j\}$ does not create a pair in μ'. Specifically,

$$n(\mu_2) = \begin{cases} n(\mu) & \text{if } \mu(j) \neq \mu'(j) \\ n(\mu) - 1 & \text{if } \mu(j) = \mu'(j). \end{cases}$$

The second case follows from the fact that if $\mu(j) = \mu'(j)$, then satisfying $\{1, j\}$ breaks pair $\{j, \mu'(j)\}$.

Under μ_2, agent m is alone. Since μ' is stable, it is individually rational, which implies that m prefers being matched with $m-1$ to being alone. Moveover, since $f(m-1) = m$, agent $m-1$ prefers m to $m-2$. Hence, $\{m-1, m\}$ blocks μ_2 provided that $m-1 \neq j$. Thus we distinguish two cases.

Case 1. $m - 1 \neq j$. Then, $\{m-1, m\}$ blocks μ_2 as we just noted. Let μ_3 denote the matching obtained from μ_2 by satisfying this blocking pair. Then $n(\mu_3) = n(\mu_2) + 1$. If $\mu(j) \neq \mu'(j)$, then $n(\mu_3) = n(\mu) + 1$ as desired.

So, suppose $\mu(j) = \mu'(j)$, which implies $j \notin \{1, \ldots, m\}$. Then $n(\mu_3) = n(\mu)$ and μ_3 looks like

$$\mu_3 = [\{2,3\}, \{4,5\}, \ldots, \{m-2\}, \{m-1, m\}, \{1, j\}, \ldots, \{\mu'(j)\}, \ldots].$$

In this matching, $m-2$ is alone.

It suffices to show that μ_3 is blocked by $\{m-3, m-2\}$. This is easy if $m \geq 6$. Indeed, $m-2$ prefers being matched with $m-3$ to being alone since $\mu'(m-2) = m-3$, and $m-3$ prefers $m-2$ to $m-4$ since $f(m-3) = m-2$.

Before we proceed to the case when $m < 6$, we note that the argument in this step up to this point applies to the example given before Step 4. In the example, the unique blocking pair of μ is $\{1,7\}$. The argument in the preceding paragraph applies to the example because agents 1 and 7 belong to different cycles (thus $m-1 \neq j$), the blocking of $\{1,7\}$ breaks a pair in μ' (i.e., $\mu'(7)=\mu(7)$), and agent 1 belongs to a cycle of size 6 (i.e., $m=6$). The associated sequence of blockings is as follows. Starting with μ, satisfying $\{1,7\}$ generates

$$\mu_2 = [\{2,3\},\{4,5\},\{6\},\{1,7\},\{8\},\{9\}].$$

Since $\{7,8\}$ is broken, the number of common pairs decreases by 1, i.e., $n(\mu_2) = n(\mu)-1 = 1$. Now, under μ_2, agent 6 is alone and prefers being matched with 5. On the other hand, agent 5 is "pointing" to 6 and prefers 6 to 4. Thus $\{5,6\}$ blocks μ_2 and we are in Case 1 of Step 9. Satisfying $\{5,6\}$ generates

$$\mu_3 = [\{2,3\},\{4\},\{5,6\},\{1,7\},\{8\},\{9\}].$$

Since 4 is alone and 3 is pointing to 4, $\{3,4\}$ blocks μ_3, which generates

$$\mu_4 = [\{2\},\{3,4\},\{5,6\},\{1,7\},\{8\},\{9\}].$$

For this matching, the number of common pairs is $n(\mu_4) = 3 = n(\mu)+1$, as our claim implies. In fact, it is straightforward to complete the sequence that leads to μ': μ_4 is blocked by $\{1,2\}$, which generates

$$\mu_5 = [\{1,2\},\{3,4\},\{5,6\},\{7\},\{8\},\{9\}].$$

This matching is in turn blocked by $\{7,8\}$, which generates μ'.

We now return to the proof and proceed to the case when $m < 6$, i.e., $c_1 = (1,2,3,4)$. Then

$$\begin{aligned}
\mu' &= [\{1,2\},\{3,4\},\ldots,\{j,\mu'(j)\},\ldots]\\
\mu &= [\{2,3\},\{4,1\},\ldots,\{j,\mu'(j)\},\ldots]\\
\mu_2 &= [\{2,3\},\{4\},\{1,j\},\ldots,\{\mu'(j)\},\ldots]\\
\mu_3 &= [\{2\},\{3,4\},\{1,j\},\ldots,\{\mu'(j)\},\ldots].
\end{aligned}$$

Since $\{1,j\}$ blocks μ, agent j prefers 1 to $\mu(j)=\mu'(j)$. On the other hand,

$\{1,j\}$ does not block μ' since μ' is stable. It follows that agent 1 prefers $\mu'(1)=2$ to j, which implies that $\{1,2\}$ blocks μ_3. This blocking generates a matching $\mu_4=[\{1,2\},\{3,4\},\ldots,\{j\},\{\mu'(j)\},\ldots]$ and $n(\mu_4)=n(\mu)+1$, as desired.

Case 2. $m-1=j$. Since c_i is a cycle, we can use the above argument letting agent $m-1$ play the role of agent 1.[6] We then conclude that the desired result follows if $m-3\neq 1$ or, equivalently, $m\geq 6$. Indeed, if $m\geq 6$, then $\{m-2,m-3\}$ blocks μ_2 inducing a matching μ_3 such that $n(\mu_3)=n(\mu)+1$.

Thus, we are left with the case in which $m=4$. That is, $c_1=(1,2,3,4)$ and $\{i,j\}=\{1,3\}$. Then

$$\begin{aligned}\mu' &= [\{1,2\},\{3,4\},\ldots]\\ \mu &= [\{2,3\},\{4,1\},\ldots]\\ \mu_2 &= [\{1,3\},\{2\},\{4\},\ldots].\end{aligned}$$

Since μ' is stable, it is not blocked by $\{1,3\}$, which implies that either agent 1 prefers $\mu'(1)=2$ to 3 or agent 3 prefers $\mu'(3)=4$ to 1. Without loss of generality, assume that agent 1 prefers 2 to 3. Since 2 is alone in μ_2, it follows that $\{1,2\}$ blocks μ_2. This blocking generates a matching $\mu_3=[\{1,2\},\{3\},\{4\},\ldots]$ and $n(\mu_3)=n(\mu)+1$. □

Theorem 1 differs from the result of Chung (2000, Lemma 1) in two respects. First, when preferences are strict, Chung's result holds under the "no odd rings" condition, while our result holds as long as a stable matching exists. As mentioned in the introduction, the "no odd rings" condition is sufficient but not necessary for the existence of a stable matching.

Second, Chung's result holds with weak preferences provided that the "no odd rings" condition is satisfied, while we consider only strict preferences. In fact, our result cannot be generalized to the roommate problem with indifferences. Indeed, if preferences are not strict and an odd ring exists, then convergence does not hold necessarily. This is shown by Chung (2000) through

[6]Note that, by Step 5, agent $m-1$ also prefers μ' to μ.

the following four-agent example:

$$
\begin{aligned}
&2 \sim_1 1 \succ_1 \cdots \\
&1 \succ_2 3 \succ_2 4 \succ_2 2 \\
&4 \succ_3 2 \succ_3 3 \succ_3 1 \\
&2 \succ_4 3 \succ_4 4 \succ_4 1.
\end{aligned}
$$

Note that agent 1 is indifferent between being matched with 2 and being alone. In this example, $(2, 3, 4)$ is an odd ring and there exists a unique stable matching, $\mu = [\{1, 2\}, \{3, 4\}]$. It can be easily checked that, starting with any matching where 1 is alone, no sequence of myopic blockings leads to μ since being alone is a top choice for 1.

Recall that the class of roommate problems subsumes marriage problems and that a stable matching exists for any marriage problem (Gale and Shapley, 1962). Thus we obtain the following corollary.

Corollary 1. *Consider any marriage problem with strict preferences. Then for any unstable matching μ, there exists a finite sequence of matchings ($\mu = \mu_1, \mu_2, \ldots, \mu_K$) such that for any $k \in \{1, 2, \ldots, K-1\}$, μ_{k+1} is obtained from μ_k by satisfying a blocking pair of μ_k and μ_K is stable.*

This result has been obtained by Roth and Vande Vate (1990). Their result holds even when preferences are not strict. On the other hand, they consider the marriage problem only.[7]

It should also be noted that our result as well as those of Roth and Vande Vate (1990) and Chung (2000) say that myopic blockings can lead to *some* stable matching. It is not the case that myopic blockings can lead to

[7]The result of Roth and Vande Vate (1990) is used by Jackson and Watts (2002) to show that, for a random process of myopic blockings with trembles, the support of the long-run stationary distribution coincides with the set of stable marriage matchings. It would be interesting to study whether the result of Jackson and Watts (2002) extends to the roommate problem.

any stable matching, as the following 3×3 marriage example shows.

$$\begin{array}{ll} w_1 \succ_{m_1} w_2 \succ_{m_1} m_1 \succ_{m_1} w_3 & \qquad m_2 \succ_{w_1} m_1 \succ_{w_1} w_1 \succ_{w_1} m_3 \\ w_2 \succ_{m_2} w_1 \succ_{m_2} m_2 \succ_{m_2} w_3 & \qquad m_1 \succ_{w_2} m_2 \succ_{w_2} w_2 \succ_{w_2} m_3 \\ w_3 \succ_{m_3} m_3 \succ_{m_3} \cdots & \qquad m_3 \succ_{w_3} w_3 \succ_{w_3} \cdots \end{array}$$

There exist only two stable matchings: $\mu_1 = [\{m_1, w_1\}, \{m_2, w_2\}, \{m_3, w_3\}]$ and $\mu_2 = [\{m_1, w_2\}, \{m_2, w_1\}, \{m_3, w_3\}]$. It is easy to see that, starting from $\mu = [\{m_1, w_1\}, \{m_2, w_2\}, \{m_3\}, \{w_3\}]$, there exists no sequence of myopic blockings leading to μ_2. The only blocking pair of μ is $\{m_3, w_3\}$ and satisfying this pair leads to μ_1.[8]

3 General Coalition Formation

Our result does not easily extend to the case in which coalitions of any sizes can form. To see this, we consider "hedonic games" (Banerjee *et al.*, 2001; Bogomolnaia and Jackson, 2002), where arbitrary coalitions can form and each agent has preferences over coalitions he belongs to. Consider the following example with $N = \{1, 2, 3\}$, taken from Bogomolnaia and Jackson (2002):

$$\begin{aligned} &\{1,2\} \succ_1 N \succ_1 \{1,3\} \succ_1 \{1\} \\ &\{2,3\} \succ_2 N \succ_2 \{1,2\} \succ_2 \{2\} \\ &\{1,3\} \succ_3 N \succ_3 \{2,3\} \succ_3 \{3\}. \end{aligned}$$

This preference profile satisfies a condition of ordinal balancedness in Bogomolnaia and Jackson (2002). The core is a singleton and consists of the partition in which the grand coalition forms. Myopic blockings generate the following cycle: $[\{1, 2\}, \{3\}] \to [\{2, 3\}, \{1\}] \to [\{1, 3\}, \{2\}] \to [\{1, 2\}, \{3\}]$. Moreover, for each partition in the cycle, there exists only one blocking coalition. Hence there exists no path coming out of the cycle.

[8] Ma (1996) considers a variant of the random matching process used in the proof of Roth and Vande Vate (1990) and shows that the random process does not necessarily reach every stable matching, even though the process starts from singletons.

Separable preferences do not guarantee convergence either.[9] Consider the following example with $N = \{1, 2, 3, 4\}$.

$$\{1,2,3\} \succ_1 N \succ_1 \{1,2\} \succ_1 \{1,2,4\} \succ_1 \{1,3\} \succ_1 \{1,3,4\} \succ_1 \{1\} \succ_1 \{1,4\}$$
$$\{2,3,4\} \succ_2 N \succ_2 \{2,3\} \succ_2 \{1,2,3\} \succ_2 \{2,4\} \succ_2 \{1,2,4\} \succ_2 \{2\} \succ_2 \{1,2\}$$
$$\{1,3,4\} \succ_3 N \succ_3 \{3,4\} \succ_3 \{2,3,4\} \succ_3 \{1,3\} \succ_3 \{1,2,3\} \succ_3 \{3\} \succ_3 \{2,3\}$$
$$\{1,2,4\} \succ_4 N \succ_4 \{1,4\} \succ_4 \{1,3,4\} \succ_4 \{2,4\} \succ_4 \{2,3,4\} \succ_4 \{4\} \succ_4 \{3,4\}.$$

The core is a singleton consisting of the partition in which the grand coalition forms. Myopic blockings generate the following cycle: $[\{1,2,3\},\{4\}] \rightarrow [\{2,3,4\},\{1\}] \rightarrow [\{1,3,4\},\{2\}] \rightarrow [\{1,2,4\},\{3\}] \rightarrow [\{1,2,3\},\{4\}]$. Again, for every partition in the cycle, there exists only one blocking coalition.

It is also easy to construct examples that satisfy the weak top coalition property of Banerjee *et al.* (2001) where myopic blockings do not lead to the core.

[9] Agent i's strict preferences over $\{S \subseteq N : i \in S\}$ are *separable* if for all $S \subseteq N$ such that $i \in S$ and for all $j \in N \setminus S$, $S \cup \{j\} \succ_i S$ if and only if $\{i,j\} \succ_i \{i\}$.

References

Banerjee, S., Konishi, H., and Sönmez, T. (2001). "Core in a simple coalition formation game," *Social Choice and Welfare* **18**, 135–153.

Bogomolnaia, A. and Jackson, M. O. (2002). "The stability of hedonic coalition structures," *Games and Economic Behavior* **38**, 201–230.

Chung, K.-S. (2000). "On the existence of stable roommate matchings," *Games and Economic Behavior* **33**, 206–230.

Feldman, A. M. (1974). "Recontracting stability," *Econometrica* **42**, 35–44.

Gale, D. and Shapley, L. (1962). "College admissions and the stability of marriage," *American Methematical Monthly* **69**, 9–15.

Green, J. R. (1974). "The stability of Edgeworth's recontracting process," *Econometrica* **42**, 21–34.

Jackson, M. O. and Watts, A. (2002). "The evolution of social and economic networks," *Journal of Economic Theory* **106**, 265–295.

Knuth, D. E. (1976). *Mariages Stables et leurs relations avec d'autres problèmes combinatoires*, Montréal: Les Presses de l'Université de Montréal.

Ma, J. (1996). "On randomized matching mechanisms," *Economic Theory* **8**, 377–381.

Roth, A. E. and Sotomayor, M. A. O. (1990). *Two-Sided Matching: A Study in Game-Theoretic Modelling and Analysis*, Cambridge University Press.

Roth, A. E. and Vande Vate, J. H. (1990). "Random paths to stability in two-sided matching," *Econometrica* **58**, 1475–80.

Sengupta, A. and Sengupta, K. (1996). "A property of the core," *Games and Economic Behavior* **12**, 266–73.

Shapley, L. and Scarf, H. (1974). "On cores and indivisibility," *Journal of Mathematical Economics* **1**, 23–28.

Tan, J. J. M. (1991). "A necessary and sufficient condition for the existence of a complete stable matching," *Journal of Algorithms* **12**, 154–178.