

Stochastic Stability for Roommate Markets

Bettina Klaus, Flip Klijn, and Markus Walzl

May 2008

Barcelona Economics Working Paper Series

Working Paper nº 357

Stochastic Stability for Roommate Markets*

Bettina Klaus,[†] Flip Klijn,[‡] and Markus Walzl[§]

May 2008

Abstract

We show that for any roommate market the set of stochastically stable matchings coincides with the set of absorbing matchings. This implies that whenever the core is non-empty (e.g., for marriage markets), a matching is in the core if and only if it is stochastically stable, i.e., stochastic stability is a characteristic of the core. Several solution concepts have been proposed to extend the core to all roommate markets (including those with an empty core). An important implication of our results is that the set of absorbing matchings is the only solution concept that is core consistent *and* shares the stochastic stability characteristic with the core.

JEL classification: C62, C71, C78.

Keywords: core, (pairwise) stability, roommate markets, stochastic stability.

1 Introduction

In a *roommate market* (see Gale and Shapley, 1962), a finite set of agents has to be partitioned into pairs (roommates) and singletons. We will refer to such a partition as a *matching*. Each agent has strict preferences over each of the other agents (i.e., sharing a room with him) and staying alone (or relying on an outside option). Hence, a roommate market is a simple example of *hedonic coalition* as well as *network formation*. In hedonic coalition formation (see Bogomolnaia and Jackson, 2002), a set of agents has to be partitioned and agents have preferences

*We thank Elena Molis (for providing Example 3), Fuhito Kojima, Utku Ünver, and other participants of the SISL mini-conference on matching (CalTech) for useful comments and discussions. B. Klaus thanks the Netherlands Organisation for Scientific Research (NWO) for its support under grant VIDI-452-06-013. F. Klijn's research was supported through the Spanish *Plan Nacional I+D+I* (SEJ2005-01690), the *Generalitat de Catalunya* (Barcelona Economics Program of XREA), and the *Consolider-Ingenio* 2010 (CSD2006-00016) program.

[†]*Corresponding author:* Department of Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands; e-mail: b.klaus@algec.unimaas.nl

[‡]Institute for Economic Analysis (CSIC), Campus UAB, 08193 Bellaterra (Barcelona), Spain; e-mail: flip.klijn@iae.csic.es

[§]Department of Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands; e-mail: m.walzl@algec.unimaas.nl

over coalitions (i.e., all subsets of agents). Thus, for roommate markets coalition formation is restricted to coalitions of at most two agents. In network formation (see Jackson and Watts, 2002), links between agents can be established and agents have preferences over their links (or even the entire network structure). Thus, for roommate markets network formation is restricted to at most one link per agent (and agents have preferences over this direct link only). Moreover, a roommate market can be interpreted as an extension of one of the most famous and simplest type of (two-sided) matching markets, a so-called *marriage market*. In a marriage market, agents are either male or female, and a man (woman) only wants to be matched to a woman (man) or to him(her)self. This setting is equivalent to a roommate market where the set of agents consists of two disjoint subsets and every agent in a certain subset prefers staying alone to being matched to another agent in the same subset. Hence, roommate markets are a particularly interesting class of matching markets because they lie in the “intersection” of network and coalition formation models.

In all the matching, coalition, and network models mentioned above, stability is a central property. For roommate markets, a matching is stable if all roommates are mutually acceptable and no pair of agents would prefer to be roommates instead of having their current matches. It is well-known that for marriage and roommate markets this (pairwise) stability notion is equivalent to core stability. However, when extending the class of marriage markets to the class of roommate markets a problem occurs: while the core for a marriage market is always non-empty, the core of a roommate market can well be empty (see Gale and Shapley, 1962). As a consequence, roommate markets can be considered an important benchmark for the development of solution concepts for matching, network and coalition formation models that may exhibit an empty core or set of stable matchings, network or coalition structures.¹

One way to obtain a non-empty solution concept for all models mentioned above is to apply a (myopic) *blocking dynamics* as introduced by Roth and Vande Vate (1990). As these blocking dynamics are finite Markov processes, there exists always at least one absorbing set, i.e., a minimal set of states that once entered throughout the blocking dynamics is never left. Thus, the set of solutions (matchings or coalition/network structures) contained in all absorbing sets – the *set of absorbing solutions* – is always non-empty. However, if the core is non-empty, the set of absorbing solutions does not always coincide with the core, i.e., the set of absorbing solutions is not always core consistent. As an example consider blocking dynamics in markets with indivisible goods as analyzed by Serrano and Volij (2007) and Bochet, Klaus, and Walzl (2007) or hedonic coalition formation (see Example 2 or Bogomolnaia and Jackson, 2002).

For *solvable* roommate markets, i.e., roommate markets with a non-empty core, Diamantoudi, Miyagawa, and Xue (2004) prove the existence of “paths to stability,” i.e., the blocking dynamics always ends up in a stable matching. In a recent paper, Iñarra, Larrea, and Molis (2008a) offer an extension of this result to general roommate markets. They introduce so-called *P*-stable matchings as a solution concept and show that the blocking dynamics always ends up in a *P*-stable matching. Their analysis is based on Tan’s (1991) notion of stable partitions. In another paper, Iñarra, Larrea, and Molis (2008b) introduce the set of absorbing matchings as a solution concept for roommate markets. They relate the absorbing sets to *P*-stable matchings and find that all absorbing sets share a common structure.

¹For more general models of matching as well as coalition or network formation, various stability notions exist and pairwise stability is no longer equivalent to core stability.

In this paper, we are interested in stochastic stability for roommate markets, i.e., we analyze blocking dynamics where agents with a small probability make mistakes in their myopic blocking decisions. This approach has first been used for equilibrium selection in non-cooperative game theory (see Kandori, Mailath, and Rob, 1993; Young, 1993) and has been adapted to network formation (and marriage markets in particular) by Jackson and Watts (2002). As in Diamantoudi et al. (2004) and Iñarra et al. (2008a), we adapt the blocking dynamics introduced in Roth and Vande Vate (1990) for roommate markets. In each period, the process is at a matching μ and a pair of agents (or a single agent) is randomly selected. If each agent of the pair (or the singleton) is better off being matched to the other agent (himself) than staying with his partner under μ , the pair (the singleton) blocks μ and the pair (the singleton) is matched while former matching partners under μ become single. This process induces a sequence of matchings called a *blocking path*. As the blocking dynamics resembles a finite Markov process, there is always an absorbing set. It follows directly from the definition of the blocking dynamics that an absorbing set is a singleton if and only if the respective matching is stable. With this, the existence of paths to stability for solvable roommate markets as in Diamantoudi et al. (2004) implies that the set of stable matchings and the set of absorbing matchings coincide in a roommate market if and only if the roommate market is solvable. Likewise, the blocking dynamics has at least one non-singleton absorbing set if and only if the roommate market is not solvable.

The set of stochastically stable matchings in a roommate market, however, is not determined by the existence of paths to stability. Instead, the following result turns out to be crucial. Let the distance between two matchings be measured by the number of agents who have the same partner and the distance between a matching and an absorbing set be measured by the maximum number of agents who have the same partner. We show that for any non-absorbing matching μ and any absorbing set A , there exists a blocking path starting from μ that does not “wander off”, i.e., a blocking path that ends in an absorbing matching that is either closer or at equal distance to the absorbing set A (Theorem 1). Using Theorem 1, we establish our main result: for roommate markets the set of absorbing matchings equals the set of stochastically stable matchings (Theorem 2). As corollaries (Corollaries 1 and 2) we obtain that a matching in a marriage market or a solvable roommate market is in the core if and only if it is stochastically stable. That is, stochastic stability characterizes the core of a solvable roommate market and characterizes the set of absorbing matchings for a general roommate market. Theorem 2 also generalizes Diamantoudi et al.’s (2004, Theorem 1) “random paths to stability” result (Corollary 3).

An important implication of our results is that the set of absorbing matchings is the only solution concept that is core consistent (i.e., if the core is non-empty, the solution coincides with the core) *and* shares the stochastic stability characteristic with the core. Finally, we explain with two examples of many-to-many matching and hedonic coalition formation (Examples 1 and 2) why our main result is a very strong structural result for roommate markets that cannot easily be extended to more general markets.

Our paper is organized as follows. In Section 2 we present the roommate model. In Section 3, we introduce the blocking dynamics and the concept of stochastic stability. Section 4 contains the main results (Theorems 1 and 2). We conclude in Section 5.

2 A Roommate Market

In a roommate market, a finite set of agents N has to be partitioned in pairs (roommates) and singletons. Each agent $i \in N$ has preferences R_i over sharing a room with any of the agents in $N \setminus \{i\}$ and having a room for himself (or consuming an outside option such as living off-campus). We assume that agents' preferences are total orders² over N . In particular, for all $i \in N$, $k R_i j$ and $j R_i k$ if and only if $j = k$. The strict preference relation associated with R_i is denoted by P_i . If $i P_i j$ then j is *unacceptable* to i . Since the set of agents N remains fixed throughout this paper, we simply denote a *roommate market* (Gale and Shapley, 1962) by its preference profile $R = (R_i)_{i \in N}$.

A solution to a roommate market, a *matching* μ , is a partition of N in pairs and singletons. Alternatively, we describe a matching by a function $\mu : N \rightarrow N$ of order two, i.e., for all $i \in N$, $\mu(\mu(i)) = i$. We denote the *set of matchings* for all roommate markets defined for the set of agents N by M . Agent $\mu(i)$ is agent i 's *mate*, i.e., the agent with whom he is matched to share a room (possibly himself). If $\mu(i) = i$ then we call i a *single*. With some abuse of notation we write $\mu R_i \mu'$ if and only if $\mu(i) R_i \mu'(i)$.

A matching μ is *blocked by a pair* $\{i, j\} \subseteq N$ (possibly $i = j$) if $j P_i \mu(i)$ and $i P_j \mu(j)$. If $\{i, j\}$ blocks μ , then $\{i, j\}$ is called a *blocking pair for* μ . A pair $\{i, j\} \subseteq N$ (possibly $i = j$) that is not a blocking pair for μ is called a *non-blocking pair for* μ . For any matching μ and *any* pair $\{i, j\}$ (possibly $i = j$), the matching μ' that results from μ by matching $\{i, j\}$ is defined by

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

i.e., the pair is matched, if i (j) was not single then his previous mate becomes single, and all other agents have the same mates as before. We say that μ' *results from* μ *by matching* $\{i, j\}$.

A matching is *individually rational* if there is no blocking pair $\{i, j\}$ with $i = j$. We denote the *set of individually rational matchings* for roommate market R by $I(R)$. A matching is *stable* if there is no blocking pair. We denote the *set of stable matchings* for roommate market R by $S(R)$. A roommate market R is *solvable* if $S(R) \neq \emptyset$. Otherwise it is called *unsolvable*. Note that for any roommate market the set of stable matchings equals the core.

A *marriage market* (Gale and Shapley, 1962) is a roommate market such that N is the union of two disjoint sets N_M and N_W (men and women), and each agent in N_M (respectively N_W) prefers being alone to being matched with any other agent in N_M (respectively N_W). An individually rational matching for a marriage market respects the partition of agents into two types and never matches two men or two women.

Gale and Shapley (1962) showed that all marriage markets are solvable and provided an unsolvable roommate market (Gale and Shapley, 1962, Example 3).

²A total order is a binary relation that satisfies antisymmetry, transitivity, and totality (comparability).

3 Stochastic Stability

3.1 The Unperturbed Blocking Dynamics and Absorbing Matchings

For each roommate market R , we model the blocking dynamics by a *Markov process* $(M, T(R))$.³ The *state space* is the set of matchings M and $T(R)$ is a *transition matrix* that induces the following *blocking dynamics*. In each period $t = 1, 2, \dots$, the process is at a matching $\mu(t)$ and a pair $\{i, j\} \subseteq N$ (possibly $i = j$) of agents is randomly selected with some positive (but not necessarily identical) period independent probability.⁴ If $\{i, j\}$ is a blocking pair for $\mu(t)$, then the process moves from $\mu(t)$ to the matching $\mu(t+1)$ that results from $\mu(t)$ by matching $\{i, j\}$. Otherwise, $\mu(t+1) = \mu(t)$. For $\mu, \mu' \in M$, let $T(R)(\mu, \mu')$ denote the probability that the dynamic process moves from matching μ to μ' from one period to the next.⁵ Note that for $\mu, \mu' \in M$, $\mu \neq \mu'$, $T(R)(\mu, \mu') > 0$ if and only if there is a blocking pair for μ that results in μ' . For any $\mu \in M$, $\sum_{\mu' \in M} T(R)(\mu, \mu') = 1$.

We call a sequence of matchings μ_1, \dots, μ_k a *path from μ_1 to μ_k* if for all $l \in \{1, \dots, k-1\}$, μ_{l+1} results from μ_l by matching a pair.⁶ A path μ_1, \dots, μ_k is a *blocking path from μ_1 to μ_k* if for all $l \in \{1, \dots, k-1\}$, μ_{l+1} results from μ_l by matching a blocking pair. For any path μ_1, \dots, μ_k we define $T(R)(\mu_1, \dots, \mu_k) := T(R)(\mu_1, \mu_2) \cdot T(R)(\mu_2, \mu_3) \cdot \dots \cdot T(R)(\mu_{k-1}, \mu_k)$. Thus, $T(R)(\mu_1, \dots, \mu_k)$ is the probability that the blocking dynamics moves from μ_1 to μ_k along the path μ_1, \dots, μ_k . Whenever for all $l \in \{1, \dots, k-1\}$, $\mu_{l+1} \neq \mu_l$, we have $T(R)(\mu_1, \dots, \mu_k) > 0$ if and only if μ_1, \dots, μ_k is a blocking path.

Note that blocking pairs are always better off in the short run (even though they might be worse off later), i.e., agents are myopic but they do not make mistakes. We will consider a dynamic process with a positive probability of mistakes (or perturbations) in Section 3.2. For further reference we therefore label the blocking process as defined in this section as the *unperturbed blocking dynamics*.

An important solution for any dynamic process is its collection of absorbing sets (or – here – matchings). A set $A \subseteq M$ is an absorbing set if it is a minimal set of matchings that once entered throughout the blocking dynamics is never left, i.e., for all $\mu, \mu' \in A$, there exists a blocking path μ, \dots, μ' , and for all $\mu \in A$ and $\mu'' \notin A$, there exists no blocking path μ, \dots, μ'' .⁷ The *collection of absorbing sets* is denoted by $\mathcal{A}(R)$. Note that $\mathcal{A}(R)$ is a collection of mutually disjoint sets of matchings. It is convenient to also denote the *set of absorbing matchings* by $A(R) = \{\mu \in M : \text{there exists } A \in \mathcal{A}(R) \text{ with } \mu \in A\}$. Illustrating examples of roommate markets, their absorbing sets, and absorbing matchings can be found in Iñarra et al. (2008b, Example 2) and Appendix D (Example 3).

The next lemma shows that the set of absorbing matchings deserves attention as a solution concept for roommate markets: the set of absorbing matchings is always non-empty and for solvable roommate markets the set of absorbing matchings coincides with the set of stable matchings. These results are straightforward implications of the definition of an unperturbed blocking dynamics and Diamantoudi et al. (2004, Theorem 1).

³We summarize basic notions and terminology of Markov processes in Appendix A.

⁴We make the assumption of time homogeneity for notational convenience. Our results hold as long as for every period the probability for any pair of agents to be selected is bounded away from zero.

⁵Note that there may be more than one blocking pair that takes μ to μ' . E.g., let $N = \{1, 2\}$, $1 P_1 2$, $2 P_2 1$, $\mu(1) = 2$, and $\mu'(1) = 1$. Then, $\{1\}$ and $\{2\}$ are both blocking pairs for μ .

⁶Such a pair can be a non-blocking pair (in particular, a pair of matched agents who remain matched).

⁷Iñarra et al. (2008b) use a different but logically equivalent definition of absorbing sets.

Lemma 1. Absorbing Matchings

For any unperturbed blocking dynamics $(M, T(R))$,

- (i) $A(R) \neq \emptyset$.
- (ii) $\{\mu\} \in \mathcal{A}(R) \Leftrightarrow \mu \in S(R)$.
- (iii) $S(R) \subseteq A(R) \subseteq I(R)$.
- (iv) $S(R) = A(R)$ if and only if R is a solvable roommate market.

Proof. Statement (i) follows from the finiteness of M . Statement (ii) is a direct implication of the definition of an unperturbed blocking dynamics and the fact that there are no blocking pairs for a stable matching. Statement (ii) implies the first inclusion of statement (iii). To prove the second inclusion of statement (iii), suppose there is $\mu \in A(R) \setminus I(R)$. Then, since $\mu \notin I(R)$, there is $i \in N$ with $i P_i \mu(i)$. Let μ' be the matching obtained from μ by (self-)matching $\{i\}$. Then, μ' and μ are in the same absorbing set. This, however, contradicts the absence of a blocking path from μ' to μ . Hence, $A(R) \subseteq I(R)$. To prove statement (iv), consider R with $S(R) \neq \emptyset$ and suppose there is $\mu \in A(R) \setminus S(R)$. Since $\mu \notin S(R)$, Diamantoudi et al. (2004, Theorem 1) implies that there exists a blocking path μ, \dots, μ' such that $\mu' \in S(R)$. Hence, μ and μ' belong to the same absorbing set. By statement (ii), $\{\mu'\} \in \mathcal{A}(R)$. Hence, $\mu = \mu'$; a contradiction. So, $A(R) \setminus S(R) = \emptyset$ and $S(R) = A(R)$. If R is not solvable, $S(R) = \emptyset$ and $S(R) \neq A(R)$ by (i). \square

Statement (iv) of Lemma 1 is also mentioned by Iñarra et al. (2008b, Corollary 2). However, our proof is different from Iñarra et al.'s (2008b, page 9) reasoning, which relies on the particular structure of so-called P -stable matchings (as introduced in Iñarra et al., 2008a).

3.2 The Perturbed Blocking Dynamics and Stochastic Stability

We now perturb the blocking dynamics $(M, T(R))$ by allowing for mistakes, i.e., if a non-blocking pair $\{i, j\}$ (possibly $i = j$) that is not yet matched is selected then it ends up being matched (by mistake) with a small probability $\epsilon > 0$. Then, for any two matchings μ and μ' , $\mu \neq \mu'$, the probability that the perturbed dynamics moves from μ to μ' from one period to the next is given by

$$T^\epsilon(R)(\mu, \mu') := T(R)(\mu, \mu') + \epsilon \sum_{\{i,j\} \in NB(\mu, \mu')} q(\mu, \{i, j\}),$$

where $NB(\mu, \mu')$ is the set of non-blocking pairs for μ that when matched result in μ' and $q(\mu, \{i, j\})$ is the probability that the pair $\{i, j\}$ is selected when the blocking dynamics is in state μ . For any $\mu \in M$, we define $T^\epsilon(R)(\mu, \mu) := 1 - \sum_{\mu' \neq \mu} T^\epsilon(R)(\mu, \mu')$.⁸ We call the resulting Markov process $(M, T^\epsilon(R))$ the *perturbed blocking dynamics*. Note that the perturbed blocking dynamics is ergodic: as mistakes induce (indirect) transitions between any two matchings its unique absorbing set is the entire state space M . Together with its aperiodicity this implies that the perturbed blocking dynamics exhibits a unique invariant distribution p^ϵ with support M that displays the long run probability distribution over matchings, i.e., for any $\mu \in M$, $p^\epsilon(\mu)$ is the probability that the perturbed blocking dynamics will be at matching μ in the long run.

The *limit invariant distribution* p^* of blocking dynamics $(M, T(R))$ is the (unique) invariant distribution p^ϵ of a perturbed blocking dynamics $(M, T^\epsilon(R))$ in the limit of $\epsilon \rightarrow 0$, i.e., $p^* := \lim_{\epsilon \rightarrow 0} p^\epsilon$. Hence, $p^*(\mu)$ is the probability that the process will be at matching μ in the long run and in the limit of vanishing mistakes. Young (1993, Theorem 4 (i)) applied to our model

⁸Note that $T^\epsilon(R)(\mu, \cdot)$ is a well-defined probability distribution over M .

implies that the limit invariant distribution exists and that it is an invariant distribution of the blocking dynamics $(M, T(R))$. The support of every invariant distribution of the unperturbed blocking dynamics is a (non-empty) collection of its absorbing sets. Matchings in the support of p^* are called *stochastically stable*. Hence, if a matching μ is stochastically stable, the perturbed blocking dynamics will be at μ in the long run and in the limit of vanishing mistakes with strictly positive probability (i.e., $p^*(\mu) > 0$). We denote the set of stochastically stable matchings by $SS(R)$.

Lemma 2 (Young, 1993, Theorem 4(i)).

For any unperturbed blocking dynamics $(M, T(R))$, $\emptyset \neq SS(R) \subseteq A(R)$.

Young (1993) provides a general methodology to determine the set of stochastically stable matchings, which boils down to the following procedure for our model.

Consider the set of directed graphs that have vertex set M . Then, any graph is defined by its set of directed edges. We allow an edge from μ' to μ'' , denoted by $[\mu', \mu'']$, if μ'' results from μ' by matching a pair. A μ -tree T_μ (or a *spanning tree*) is a graph such that for every $\mu' \in M$ with $\mu' \neq \mu$ there is exactly one (cycle-free) sequence of edges (a *path*) from μ' to μ . Denote by \mathcal{T}_μ the set of all μ -trees. The irreducibility of the perturbed blocking dynamics guarantees that \mathcal{T}_μ is non-empty.

Let $[\mu', \mu'']$ be an edge in a μ -tree $T_\mu \in \mathcal{T}_\mu$. The *edge-resistance* $r(\mu', \mu'')$ indicates whether or not a mistake is needed to move from μ' to μ'' ,

$$r(\mu', \mu'') := \begin{cases} 0 & \text{if } T(R)(\mu', \mu'') > 0; \\ 1 & \text{otherwise.} \end{cases}$$

The *tree-resistance* $r(T_\mu)$ of a μ -tree equals the sum of corresponding edge-resistances, i.e., $r(T_\mu) = \sum_{[\mu', \mu''] \in T_\mu} r(\mu', \mu'')$. Finally, the *stochastic potential* of $\mu \in M$, denoted by $\gamma(\mu)$, is the minimal tree-resistance over all μ -trees, i.e., $\gamma(\mu) = \min_{T_\mu \in \mathcal{T}_\mu} r(T_\mu)$. A μ -tree T_μ with $r(T_\mu) = \gamma(\mu)$ is a *least resistance μ -tree*. The set of stochastically stable matchings $SS(R)$ can now conveniently be characterized as the set of matchings with minimal stochastic potential.

Lemma 3 (Young, 1993, Theorem 4(ii)).

Consider any unperturbed blocking dynamics $(M, T(R))$. Let $\mu \in M$. Then, $\mu \in SS(R)$ if and only if for all $\mu' \in M$, $\gamma(\mu) \leq \gamma(\mu')$.

4 Stability and Stochastic Stability

In Section 4.1 we define a distance measure between matchings and absorbing sets and establish various preliminary results. The key result of this section is that for any non-absorbing matching μ and any absorbing set A , there exists a blocking path starting from μ that does not “wander off”, i.e., a blocking path that ends in an absorbing matching that is either closer or at equal distance to the absorbing set A (Theorem 1).

In Section 4.2 we establish the main result of this article: for roommate markets the set of absorbing matchings equals the set of stochastically stable matchings (Theorem 2). For solvable roommate and marriage markets, this implies the coincidence of stochastically stable and stable matchings (Corollaries 1 and 2). Furthermore, Theorem 2 implies Diamantoudi et al.’s (2004, Theorem 1) “random paths to stability” result (Corollary 3).

4.1 Blocking Paths and the Distance Between Matchings

Given two matchings μ and μ' let $m(\mu, \mu')$ denote the number of agents that have the same mate under μ and μ' , i.e., $m(\mu, \mu') = |\{i \in N : \mu(i) = \mu'(i)\}|$. Interpreting $m(\cdot, \cdot)$ as a measure for the distance between two matchings we say μ' is closer to μ than μ'' (or equivalently, μ is closer to μ' than to μ'') if $m(\mu, \mu') > m(\mu, \mu'')$.⁹

In order to measure the distance between a matching and a set, we define for all matchings μ and $A \subseteq M$, $m(\mu, A) := m(A, \mu) := \max_{\mu' \in A} m(\mu, \mu')$ and say that for all matchings μ and $A, A' \subseteq M$, A is closer to μ than A' (or equivalently, μ is closer to A than to A') if $m(\mu, A) > m(\mu, A')$.

Finally, to measure the distance between two sets, we define for all $A, A' \subseteq M$, $m(A, A') := \max_{\mu \in A} m(\mu, A')$ and we say that for all $A, A', A'' \subseteq M$, A' is closer to A than A'' (or equivalently, A is closer to A' than to A'') if $m(A, A') > m(A, A'')$.

In what follows, we investigate the changes in the distance between matchings and between matchings and absorbing sets due to blocking and mistakes. To this end, we consider an absorbing set $A \in \mathcal{A}(R)$, a matching $\mu \in M \setminus A(R)$, and a matching $\mu' \in A$ with $m(\mu, \mu') = m(\mu, A)$, i.e., there is no matching $\mu'' \in A$ that is closer to μ than μ' .

First, consider the case that the pair $\{i, j\}$ blocks both μ and μ' .

Lemma 4. *Let $\mu, \mu' \in M$ and $\{i, j\} \subseteq N$ be a blocking pair for μ and μ' . Let $\tilde{\mu}$ and $\tilde{\mu}'$ be the matchings that result from μ and μ' by matching $\{i, j\}$. Then, $m(\tilde{\mu}, \tilde{\mu}') \geq m(\mu, \mu')$.*

Proof. We prove $m(\tilde{\mu}, \tilde{\mu}') \geq m(\mu, \mu')$ by showing that for all $k \in N$, $[\mu(k) = \mu'(k)] \Rightarrow [\tilde{\mu}(k) = \tilde{\mu}'(k)]$. Let $k \in N$. Suppose $\mu(k) = \mu'(k)$.

- (i) Let $k \notin \{i, j, \mu(i), \mu(j), \mu'(i), \mu'(j)\}$. Then, $\mu(k) = \tilde{\mu}(k)$ and $\mu'(k) = \tilde{\mu}'(k)$.
- (ii) Let $k \in \{i, j\}$. Note that $\tilde{\mu}(i) = \tilde{\mu}'(i) = j$ and $\tilde{\mu}(j) = \tilde{\mu}'(j) = i$. Thus, $\tilde{\mu}(k) = \tilde{\mu}'(k)$.
- (iii) Let $k \in \{\mu(i), \mu(j), \mu'(i), \mu'(j)\} \setminus \{i, j\}$. Assume, without loss of generality, that $k = \mu(i)$. Thus, $\mu(k) = \mu'(k) = i \neq k$. But then, since $k \notin \{i, j\}$, $\tilde{\mu}(k) = \tilde{\mu}'(k) = k$.

In all cases (i), (ii), and (iii), we can conclude that $\tilde{\mu}(k) = \tilde{\mu}'(k)$. □

Next, we consider the case that there is a blocking pair $\{r, s\}$ with $r \neq s$ for μ that is not a blocking pair for μ' .

Lemma 5. *Let $\mu, \mu' \in M$ be such that $\mu \in I(R) \setminus A(R)$, $\mu' \in A \in \mathcal{A}(R)$, and $m(\mu, \mu') = m(\mu, A)$. Suppose there exists a blocking pair $\{r, s\}$ with $r \neq s$ for μ that is not a blocking pair for μ' . Then, there exists a blocking path $\mu = \mu_1, \dots, \mu_L$ for μ such that $m(\mu_L, A) > m(\mu, A)$.*

Proof. See Appendix B. □

The proof of Lemma 5 parallels the proof of the Claim in Diamantoudi et al. (2004, Theorem 1). The main difference is that we have to assume that there is a blocking pair for μ that is not a blocking pair for μ' , while their analysis is restricted to singleton absorbing sets (i.e., a stable matching μ') which implies the absence of any blocking pair for μ' .

⁹The function $d : M \times M \rightarrow \mathbb{R}$ with $d(\mu, \mu') = |N| - m(\mu, \mu')$ is a distance function on the set of matchings as it satisfies (i) for all $\mu, \mu' \in M$, $d(\mu, \mu') \geq 0$ and $d(\mu, \mu') = 0$ if and only if $\mu = \mu'$; (ii) symmetry, i.e., for all $\mu, \mu' \in M$, $d(\mu, \mu') = d(\mu', \mu)$; and (iii) the triangle inequality, i.e., for all $\mu, \mu', \mu'' \in M$, $d(\mu, \mu'') \leq d(\mu, \mu') + d(\mu', \mu'')$.

For the construction of μ' -trees in the analysis of stochastic stability, we have to find paths from one absorbing set to any other with the least number of mistakes. For this, we will have to prove the existence of a blocking path from a matching $\mu \in M \setminus A(R)$ to a matching in an absorbing set such that this latter matching does not lie strictly further away from a fixed absorbing set $A \in \mathcal{A}(R)$ than μ (Theorem 1). The proof of this result in turn requires an extension of the analysis of single blocking pairs (Lemmas 4 and 5) to paths of blocking pairs. A *path of blocking pairs* for μ is a finite sequence of pairs $\{i_1, j_1\}, \{i_2, j_2\}, \dots, \{i_p, j_p\}$ such that $\{i_1, j_1\}$ is a blocking pair for $\mu_1 := \mu$, $\{i_2, j_2\}$ is a blocking pair for the matching μ_2 that results from μ_1 by matching $\{i_1, j_1\}$, \dots , and $\{i_p, j_p\}$ is a blocking pair for the matching μ_p that results from μ_{p-1} by matching $\{i_{p-1}, j_{p-1}\}$. Let μ_{p+1} be the matching that results from μ_p by matching $\{i_p, j_p\}$. We call μ_{p+1} the *matching induced by* $\{i_1, j_1\}, \{i_2, j_2\}, \dots, \{i_p, j_p\}$. Similarly, we call the blocking path $\mu = \mu_1, \dots, \mu_{p+1}$ the *blocking path for μ induced by* $\{i_1, j_1\}, \{i_2, j_2\}, \dots, \{i_p, j_p\}$.

Lemma 6. *Let $\mu, \mu' \in M$ be such that $\mu \in I(R) \setminus A(R)$, $\mu' \in A \in \mathcal{A}(R)$, $m(\mu, \mu') = m(\mu, A)$. Suppose $\{i_1, j_1\}, \{i_2, j_2\}, \dots, \{i_K, j_K\}$ is a path of blocking pairs for both μ and μ' . Let $\mu = \mu_1, \dots, \mu_K$ be the induced blocking path for μ . Suppose $m(\mu_k, A) \leq m(\mu, A)$ for all $k = 2, \dots, K$. Then, $m(\mu_K, A) = m(\mu, A)$.*

Proof. Let $\mu' = \mu'_1, \dots, \mu'_K$ be the induced blocking path for μ' . By assumption, $m(\mu_1, \mu'_1) = m(\mu_1, A) = m(\mu, A)$. Assume, by induction hypothesis, that $k < K$ is such that $m(\mu_k, \mu'_k) = m(\mu_k, A) = m(\mu, A)$. Then,

$$m(\mu_{k+1}, \mu'_{k+1}) \geq m(\mu_k, \mu'_k) = m(\mu_k, A) = m(\mu, A) \geq m(\mu_{k+1}, A) \geq m(\mu_{k+1}, \mu'_{k+1}),$$

where the first inequality follows from Lemma 4, the equalities from the induction hypothesis, the second inequality from the assumptions, and the last inequality from the definition of m . Hence, all inequalities are in fact equalities and $m(\mu_{k+1}, \mu'_{k+1}) = m(\mu_{k+1}, A) = m(\mu, A)$. Thus, by induction, for $k + 1 = K$, $m(\mu_K, A) = m(\mu, A)$. \square

Lemma 6 facilitates the proof of the following theorem which for any non-absorbing matching μ and for any absorbing set A establishes the existence of a blocking path from μ to an absorbing matching μ' such that μ is not closer to A than μ' (see Figure 1). This result turns out to be the cornerstone for our analysis of stochastic stability.

Theorem 1. Existence of blocking paths that do not “wander off”

Let $A \in \mathcal{A}(R)$ and $\mu \in M \setminus A(R)$. Then, there exists a blocking path $\mu = \mu_1, \dots, \mu_K$ for μ such that $\mu_K \in A(R)$ and $m(\mu_K, A) \geq m(\mu, A)$.

We prove Theorem 1 by showing that the following algorithm is well-defined and that it terminates in a finite number of steps.

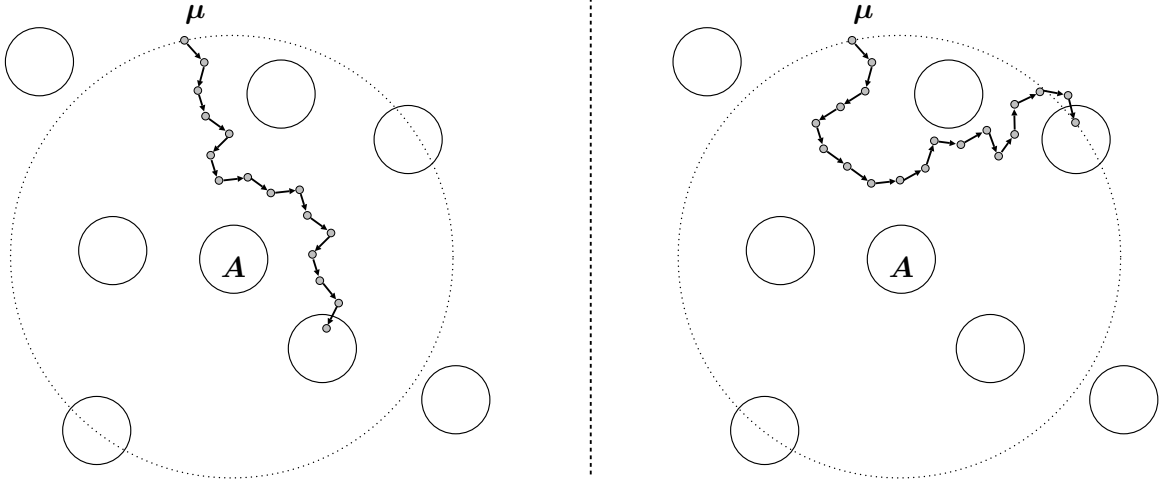


Figure 1: Theorem 1, blocking paths that do not “wander off”

Algorithm. Let $A \in \mathcal{A}(R)$ and $\mu \in M \setminus A(R)$.

Step I. If $\mu \notin I(R)$, then unmatch one by one all agents that have an unacceptable mate under μ . Let $\mu = \mu_1, \dots, \mu_L$ be the corresponding blocking path. Hence, $\mu_L \in I(R)$. Then, $m(\mu_L, A) \geq m(\mu, A)$. (This follows from Lemma 1(iii) $A(R) \subseteq I(R)$.) If $\mu_L \in A(R)$, then stop. Otherwise set $\mu := \mu_L$.

Step II. Let $\mu' \in A$ be such that $m(\mu, A) = m(\mu, \mu')$.

Step III. If each path of blocking pairs for μ

- (i) is also a path of blocking pairs for μ' , and
- (ii) does not induce a matching closer to A ,

then take a path of blocking pairs for μ that induces a blocking path $\mu = \mu_1, \dots, \mu_K$ for μ such that $\mu_K \in A(R)$ and stop. By Lemma 6, $m(\mu_K, A) = m(\mu, A)$.

Step IV. If some path of blocking pairs for μ

- (i) is also a path of blocking pairs for μ' , and
- (ii) induces a matching closer to A ,

then take a path of blocking pairs for μ that induces a blocking path $\mu = \mu_1, \dots, \mu_L$ for μ such that $m(\mu_L, A) > m(\mu, A)$. If $\mu_L \in A(R)$, then set $\mu_K := \mu_L$ and stop. Otherwise set $\mu := \mu_L$ and go to Step II.

Step V. Note that if this step is reached, then there exists a path of blocking pairs for μ that is not a path of blocking pairs for μ' . Take a shortest such path of blocking pairs $\{i_1, j_1\}, \{i_2, j_2\}, \dots, \{i_L, j_L\}$ for μ . Thus, the first $L - 1$ blocking pairs induce a blocking path $\mu = \mu_1, \dots, \mu_L$ for μ and a blocking path $\mu' = \mu'_1, \dots, \mu'_L$ for μ' . Furthermore, the L -th blocking pair for μ , $\{i_L, j_L\}$, is *not* a blocking pair for μ'_L .¹⁰ Then, by Lemma 6, either

- (i) for some k , $1 < k \leq L$, $m(\mu_k, A) > m(\mu, A)$, or
- (ii) $m(\mu_L, A) = m(\mu, A)$.

¹⁰Since μ is individually rational, μ_L is also individually rational. Therefore $i_L \neq j_L$.

Case (i): If $\mu_k \in A(R)$, then set $\mu_K := \mu_k$ and stop. Otherwise set $\mu := \mu_k$ and go to Step II.

Case (ii): If $\mu_L \in A(R)$, then set $\mu_K := \mu_L$ and stop. Otherwise set $\mu := \mu_L$ and $\mu' := \mu'_L$ and go to Step VI.

Step VI. Note that if this step is reached, then μ and μ' are such that $\mu \in I(R) \setminus A(R)$, $\mu' \in A \in \mathcal{A}(R)$, $m(\mu, \mu') = m(\mu, A)$, and there exists a blocking pair $\{r, s\}$ ($r \neq s$) for μ that is not a blocking pair for μ' . By Lemma 5, there exists a blocking path $\mu = \mu_1, \dots, \mu_L$ for μ such that $m(\mu_L, A) > m(\mu, A)$. Take such a path. If $\mu_L \in A(R)$, then set $\mu_K := \mu_L$ and stop. Otherwise set $\mu := \mu_L$ and go to Step II.

Lemma 7. *The algorithm is well-defined and terminates in a finite number of steps.*

Proof. Lemmas 5 and 6 ensure that the algorithm is well-defined. Note that starting from Step II, either the algorithm terminates in Steps III, IV, V, or VI (with a desired matching μ_K), or the algorithm returns to Step II with a matching in $M \setminus A(R)$ that is closer to A than the previous matching μ of Step II. Since the distance to A is finite, the algorithm terminates after finitely many steps. \square

4.2 The Main Result

Theorem 1 is key in the construction of μ -trees to prove the stochastic stability of *all* absorbing matchings.

Theorem 2. Roommate Markets: $SS=A$

For any roommate market R , the set of stochastically stable matchings coincides with the set of absorbing matchings, i.e., $SS(R) = A(R)$.

Proof. Consider a roommate market R . First, we determine the stochastic potential of all absorbing matchings. More precisely, we show that if $|\mathcal{A}(R)| = k$, then for all $\mu \in A(R)$, $\gamma(\mu) = k - 1$. Then, we conclude that $A(R) = SS(R)$.

Let $\mathcal{A}(R) = \{A_1, A_2, \dots, A_k\}$ and $\mu \in A_1$. Note that at least one mistake is needed to connect any two absorbing sets. Hence, for all $\mu \in M$, $\gamma(\mu) \geq k - 1$. We prove $\gamma(\mu) = k - 1$ by constructing a μ -tree T_μ with tree-resistance $k - 1$.

Step 1. Note that we can connect all absorbing matchings in $A_1 \setminus \{\mu\}$ to μ without cycles and without mistakes. This creates a subtree T_1 with tree-resistance $r(T_1) = 0$ of the final μ -tree T_μ .

Next, we show how to connect all absorbing matchings in $A_2 \cup \dots \cup A_k$ to T_1 with one mistake per absorbing set by constructing subtrees T_2, \dots, T_k with $T_1 \subset T_2 \subset \dots \subset T_k \subseteq T_\mu$. For this construction, we order the absorbing sets $A_2 \cup \dots \cup A_k$ according to their distance to A_1 : without loss of generality, $m(A_1, A_1) > m(A_1, A_2) \geq \dots \geq m(A_1, A_k)$ (see Figure 2).

Step 2. Let $\mu_1 \in A_1$ and $\mu_2 \in A_2$ be such that $m(\mu_1, \mu_2) = m(\mu_1, A_2)$, i.e., μ_1 and μ_2 minimize the distance between A_1 and A_2 . Since $\mu_1 \neq \mu_2$, there exists a pair $\{i, j\}$ (possibly $i = j$) such that $\mu_1(i) = j$ and $\mu_2(i) \neq j$. Then, let $\tilde{\mu}_2$ be the matching that results from μ_2 by matching such a pair $\{i, j\}$. Note that $m(\mu_1, \tilde{\mu}_2) > m(\mu_1, \mu_2)$ and therefore, $m(A_1, \tilde{\mu}_2) > m(A_1, A_2)$ and $\tilde{\mu}_2 \notin A_2 \cup \dots \cup A_k$. Furthermore, by construction $r(\mu_2, \tilde{\mu}_2) = 1$ (if $r(\mu_2, \tilde{\mu}_2) = 0$, then $\tilde{\mu}_2 \in A_2$).

Suppose that $\tilde{\mu}_2 \notin A_1$. Then, $\tilde{\mu}_2 \in M \setminus A(R)$ and by Theorem 1, there exists a blocking path $\tilde{\mu}_2, \dots, \hat{\mu}_1$ such that $\hat{\mu}_1 \in A(R)$ and $m(\hat{\mu}_1, A_1) \geq m(\tilde{\mu}_2, A_1)$. Recall that no absorbing matching in $A_2 \cup \dots \cup A_k$ is closer to A_1 than μ_2 (and that $\tilde{\mu}_2$ is closer to A_1 than μ_2). Thus, $\hat{\mu}_1 \in A_1$.

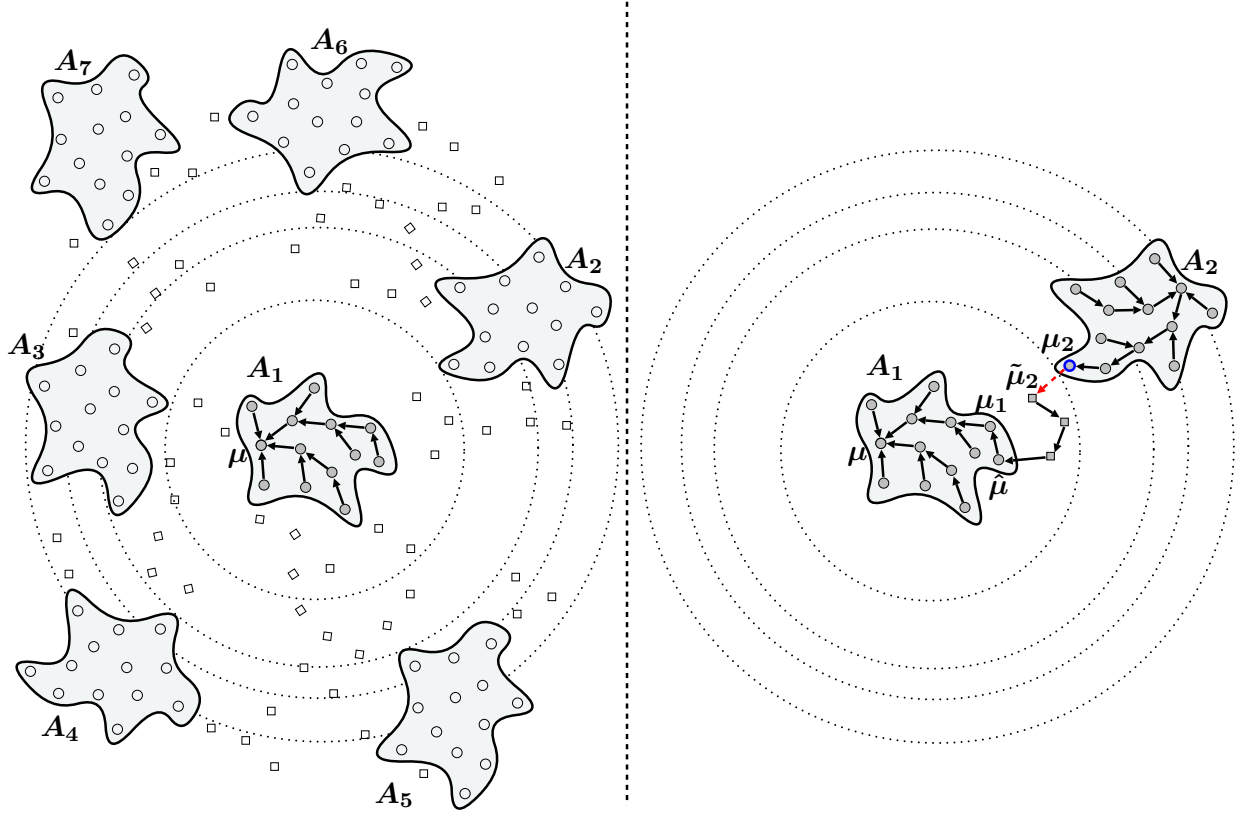


Figure 2: Proof of Theorem 2, Steps 1 and 2

Let $\hat{\mu}$ be the first matching on the path $\mu_2, \tilde{\mu}_2, \dots, \hat{\mu}_1$ that is connected to T_1 . We add the path $\mu_2, \tilde{\mu}_2, \dots, \hat{\mu}$ to T_1 to obtain a subtree with tree-resistance 1 of the final μ -tree T_μ . For $\tilde{\mu}_2 \in A_1$, the path we add to T_1 equals $\mu_2, \tilde{\mu}_2 = \hat{\mu}_1$ and again obtain a subtree with tree-resistance 1 of the final μ -tree T_μ .

Next, we connect all matchings in $A_2 \setminus \{\mu_2\}$ to μ_2 without cycles and without mistakes. Thus, we obtain a subtree T_2 of the final μ -tree T_μ that connects all absorbing matchings in $A_1 \cup A_2$ and some non-absorbing matchings with μ (see Figure 2). Let $S_2 \subset M$ denote the latter set of non-absorbing matchings. The tree-resistance is $r(T_2) = 1$.

Steps 3 up to k . Let $l \in \{3, \dots, k\}$ and assume that we have constructed a subtree T_{l-1} with $r(T_{l-1}) = l - 2$ of the final μ -tree T_μ that connects all absorbing matchings in $A_1 \cup \dots \cup A_{l-1}$ and all non-absorbing matchings in S_{l-1} with μ .

Let $\mu_1 \in A_1$ (not necessarily the same μ_1 as in previous steps) and $\mu_l \in A_l$ be such that $m(\mu_1, \mu_l) = m(\mu_1, A_l)$, i.e., μ_1 and μ_l minimize the distance between A_1 and A_l . Since $\mu_1 \neq \mu_l$, there exists a pair $\{i, j\}$ (possibly $i = j$) such that $\mu_1(i) = j$ and $\mu_l(i) \neq j$. Then, let $\tilde{\mu}_l$ be the matching that results from μ_l by matching such a pair $\{i, j\}$. Note that $m(\mu_1, \tilde{\mu}_l) > m(\mu_1, \mu_l)$ and therefore, $m(A_1, \tilde{\mu}_l) > m(A_1, A_l)$ and $\tilde{\mu}_l \notin A_l \cup \dots \cup A_k$. Furthermore, by construction $r(\mu_l, \tilde{\mu}_l) = 1$ (if $r(\mu_l, \tilde{\mu}_l) = 0$, then $\tilde{\mu}_l \in A_l$).

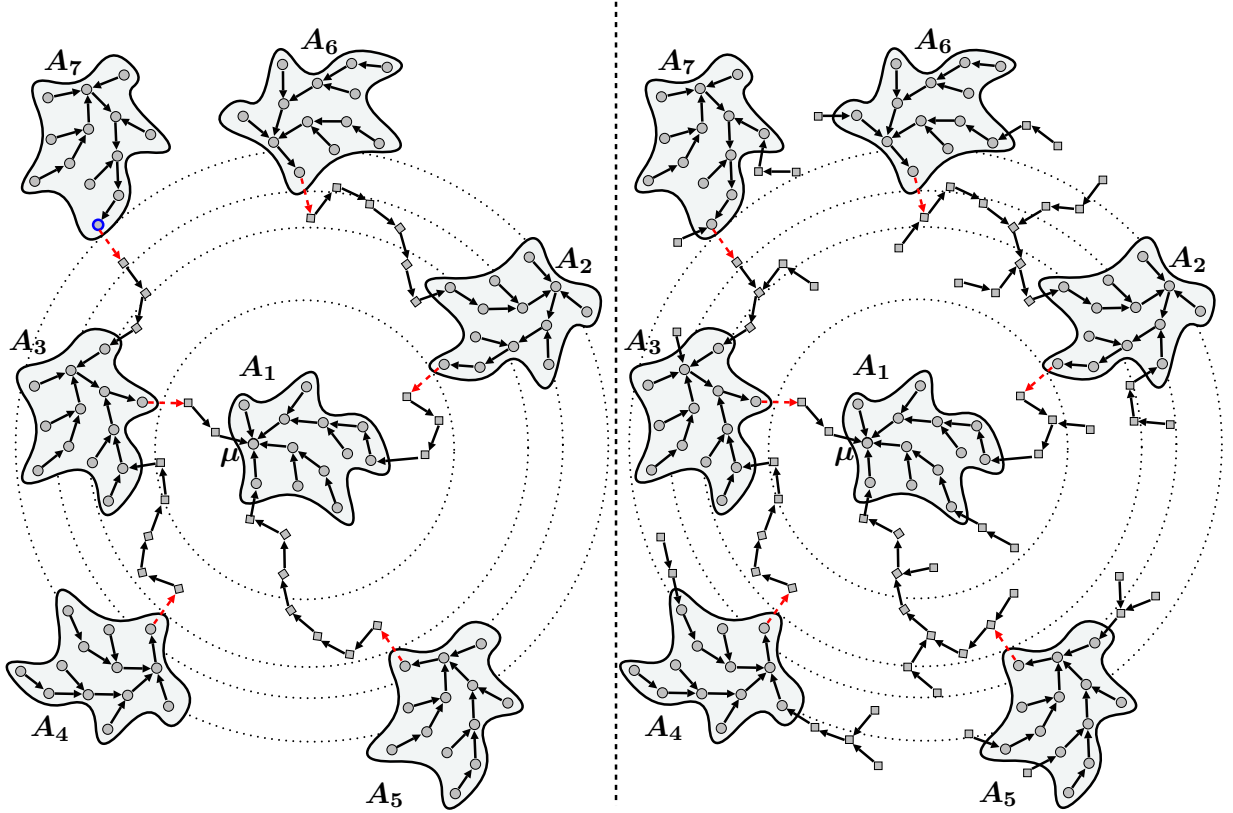


Figure 3: Proof of Theorem 2, Step 7 and Least Resistance μ -Tree

Suppose that $\tilde{\mu}_l \notin A_1 \cup \dots \cup A_{l-1}$. Then, $\tilde{\mu}_l \in M \setminus A(R)$ and by Theorem 1, there exists a blocking path $\tilde{\mu}_l, \dots, \hat{\mu}_1$ (not necessarily the same $\hat{\mu}_1$ as in previous steps) such that $\hat{\mu}_1 \in A(R)$ and $m(\hat{\mu}_1, A_1) \geq m(\tilde{\mu}_l, A_1)$. Recall that no matching in $A_l \cup \dots \cup A_k$ is closer to A_1 than μ_l (and that $\tilde{\mu}_l$ is closer to A_1 than μ_l). Thus, $\hat{\mu}_1 \in A_1 \cup \dots \cup A_{l-1}$ and $\mu_l, \tilde{\mu}_l, \dots, \hat{\mu}_1$ is a path from μ_l into $A_1 \cup \dots \cup A_{l-1}$. Let $\hat{\mu}$ be the first matching on the path $\mu_l, \tilde{\mu}_l, \dots, \hat{\mu}_1$ that is connected to T_{l-1} . We add the path $\mu_l, \tilde{\mu}_l, \dots, \hat{\mu}$ to T_{l-1} to obtain a subtree with tree-resistance $l - 1$ of the final μ -tree T_μ . For $\tilde{\mu}_l \in A_1 \cup \dots \cup A_{l-1}$, the path we add to T_{l-1} equals $\mu_l, \tilde{\mu}_l$ and again obtain a subtree with tree-resistance $l - 1$ of the final μ -tree T_μ .

Next, we connect all matchings in $A_l \setminus \{\mu_l\}$ to μ_l without cycles and without mistakes. Thus, we obtain a subtree T_l of the final μ -tree T_μ that connects all absorbing matchings in $A_1 \cup \dots \cup A_l$ and some non-absorbing matchings with μ . Let $S_l \subset M$ denote the latter set of non-absorbing matchings. (Note $S_l \supseteq S_{l-1}$.) The tree-resistance is $r(T_l) = l - 1$.

After k steps we have constructed a subtree T_k of the final μ -tree T_μ that connects all (absorbing) matchings in $A(R)$ and all (non-absorbing) matchings in S_k with μ . The tree-resistance is $r(T_k) = k - 1$ (see Figure 3).

A Least Resistance μ -Tree. Finally, by Theorem 1, we can connect any remaining non-absorbing matching in $M \setminus (A(R) \cup S_k)$ without mistake to T_k through a zero resistance path (as before, paths have to be truncated if they contain already connected matchings). The resulting tree T_μ is a μ -tree with tree-resistance $k - 1$ (see Figure 3). Hence, $\gamma(\mu) = k - 1$.

$A(R)=SS(R)$. As already mentioned in the beginning of the proof, at least one mistake is needed to connect any two absorbing sets. Hence, for all $\mu' \in M$ and $\mu \in A(R)$, $\gamma(\mu') \geq k - 1 = \gamma(\mu)$. Hence, by Lemma 3, $A(R) \subseteq SS(R)$. By Lemma 2, $A(R) \supseteq SS(R)$. Hence, $A(R) = SS(R)$. \square

Theorem 2 and Lemma 1(iv) imply that for solvable roommate and marriage markets stochastic stability and stability coincide.

Corollary 1. Solvable Roommate Markets: $SS=S$

For any solvable roommate market R , the set of stochastically stable matchings coincides with the set of stable matchings, i.e., $SS(R) = S(R)$.

Corollary 2. Marriage Markets: $SS=S$

For any marriage market R , the set of stochastically stable matchings coincides with the set of stable matchings, i.e., $SS(R) = S(R)$.

Remark 1. Robustness of Results for Different “Mistake Dynamics”

Jackson and Watts (2002, Theorem 2) establish a very similar result to Corollary 2. However, they consider a different mistake dynamics where agents only break links by mistake, but do not form new links by mistake. In our model this corresponds to restricting mistakes to single agents, i.e., the only “pairs” $\{i, j\}$ that make mistakes are those with $i = j$.¹¹ We prove in Appendix C that Theorem 2 and Corollaries 1 and 2 also hold for the restricted mistake dynamics (and that therefore we can also obtain Jackson and Watts’s Theorem 2 as a corollary).

Not allowing a pair of (two) agents to match by mistake clearly affects Step 2 (and very similarly, Steps 3 up to k) in the proof of Theorem 2. To be more specific, it can no longer be taken for granted that a pair $\{i, j\}$ that is matched in μ_1 but not in μ_2 can be matched with one mistake (which was sufficient to demonstrate that we only need one mistake to connect any two absorbing sets). However, structural properties of absorbing sets of roommate problems as derived by Ñarra et al. (2008b) and the decomposability of stable matchings (Gusfield and Irving, 1989, Lemma 4.3.9; Klaus and Klijn, 2008, Lemma 2) can be utilized to show that again only one mistake (this time by a single agent) is needed to connect any two absorbing sets (for a detailed discussion see Appendix C).

Finally, note that by definition of our mistake dynamics, if a pair of agents is matched then at most one mistake is made. In other words, they make a mistake if exactly one of the agents is worse off, but they also make only one mistake if both of the agents are worse off. As explained above (see also Appendix C) we in fact only need single agents to make mistakes. Hence, our results are also valid with a mistake dynamics where the matching of two agents that makes both of them worse off counts as two individual mistakes. \triangle

Finally, we obtain Diamantoudi et al.’s (2004) so-called random path to stability result as follows. For any solvable roommate market R , pick a stable matching μ' and construct a μ' -tree using the technique of the proof of Theorem 2. Recall that by Lemma 1(ii) the μ' -tree only contains singleton absorbing sets such that each absorbing set corresponds to exactly one of the stable matchings. Furthermore, the only edges in the μ' -tree with edge-resistance 1 are edges that start from a stable matching. Hence, starting from any *unstable* matching $\mu \in M \setminus S(R)$ the μ' -tree determines a sequence of zero resistance edges – a blocking path for μ – to one of the stable matchings.

¹¹Our notion of the mistake dynamics coincides with that of Jackson and Watts (2002, p. 277) for network formation. They restrict the mistake dynamics only for the analysis of marriage markets.

Corollary 3 (Diamantoudi et al., 2004, Theorem 1). *Random Paths to Stability*

For any solvable roommate market R and any unstable matching $\mu \in M \setminus S(R)$ there exists a blocking path $\mu = \mu_1, \dots, \mu_K$ for μ such that μ_K is stable, i.e., $\mu_K \in S(R)$.

5 Concluding Remarks

We first discuss the role of stochastic stability in “validating” the set of absorbing matchings as the solution concept that properly extends the core to non-solvable roommate markets. Second, with two examples of many-to-many matching and hedonic coalition formation we explain why our main result is a very strong structural result for roommate markets that cannot easily be extended to more general markets.

5.1 Stochastic Stability as Solution Validation Device

We have already discussed in the Introduction that roommate markets are very “simple” markets that lie in the intersection of more general models of matching, network, and coalition formation. Unfortunately, the most prominent solution concept for many basic matching markets – the set of (pairwise/core) stable matchings – is not applicable since for some roommate markets no stable matchings exist. Several solution concepts for non-solvable roommate markets have been discussed. Here we focus on those solution concepts that are “core consistent”, i.e., that coincide with the set of stable matchings whenever it is non-empty. We are aware of four such core consistent extensions:

1. *Absorbing matchings A*: as proven in Lemma 1(iv), for all solvable roommate markets R , $A(R) = S(R)$ (see also Iñarra et al., 2008b, Corollary 2).
2. *Maximum stable matchings MS*: Tan (1990) defines a maximum stable matching as a matching with a “maximum number of disjoint pairs of persons such that these pairs are stable among themselves, i.e., no two persons, who are not paired together but have matched partners, both prefer each other to their partner under the matching.” We denote the set of maximum stable matchings for a roommate market R by $MS(R)$. Since for solvable roommate markets the set of single agents is the same at any stable matching (e.g., Klaus and Klijn, 2008, Theorem 1), it follows that for all solvable roommate markets R , $MS(R) = S(R)$.
3. *Almost stable matchings AS*: Abraham et al. (2006) analyze matchings with a minimal number of blocking pairs and refer to all such matchings as almost stable. We denote the set of almost stable matchings for a roommate market R by $AS(R)$. Since for any stable matching of a solvable roommate market the number of blocking pairs equals zero, it follows that for all solvable roommate markets R , $AS(R) = S(R)$.
4. *P-stable matchings PS*: Iñarra et al. (2008a) use Tan’s (1991) notion of stable partitions for roommate markets to define the set of P -stable matchings (each stable partition P induces a set of P -stable matchings). We denote the set of P -stable matchings for a roommate market R by $PS(R)$. From Tan’s (1991) characterization of solvable roommate markets it follows easily that for all solvable roommate markets R , $PS(R) = S(R)$ (see also Iñarra et al., 2008a, Remark 3).¹²

¹²Note that for all roommate markets R , $A(R) \cap PS(R) \neq \emptyset$ (see Iñarra et al., 2008a,b).

Example 3 in Appendix D describes a (non-solvable) roommate market R such that $[A(R) \subsetneq MS(R)$ and $A(R) \supsetneq MS(R)]$, $[A(R) \subsetneq AS(R)$ and $A(R) \supsetneq AS(R)]$, and $[A(R) \subsetneq PS(R)$ and $A(R) \supsetneq PS(R)]$.

In order to judge which of the above-described solution concepts is the best extension of the core to unsolvable roommate markets, we propose to apply stochastic stability as a second validation criterium (the first being core consistency). A stochastically stable matching is a prediction for the long-run state of the blocking dynamics that is robust with respect to small perturbations. By Corollary 1, for solvable roommate markets the core coincides with the set of stochastically stable matchings, i.e., stochastic stability characterizes the core and no stable matching is singled out as a “better” prediction. By Theorem 2, stochastic stability also characterizes the set of absorbing matchings. Hence, the set of absorbing matchings is the only solution concept for roommate markets that is core consistent *and* shares the stochastic stability characteristic with the core.

5.2 Stochastic Stability and Stability for More General Matching Models

In this section we consider extensions of the roommate model for which pairwise stability does not necessarily coincide with core or other stability notions anymore. Without loss of generality, whenever referring to stability, we have pairwise stability (i.e., one of the weakest stability notions in these models) in mind.

In the case of a solvable roommate market, the existence of a blocking path to stability (as proven by Diamantoudi et al., 2004) implies that the set of absorbing matchings coincides with the set of stable matchings (see the proof of Lemma 1(iv)). Therefore, by Lemma 2, $SS \subseteq S$. The stochastic stability of *all* stable matchings (see Theorem 2 and Corollary 1), however, is not implied by the existence of a blocking path to stability but hinges on the existence of blocking paths that do not “wander off” (see Theorem 1). In the following, we provide an example of a many-to-many matching market that demonstrates that the mere existence of a blocking path to stability does not guarantee stochastic stability of all stable matchings.

Kojima and Ünver (2008, Theorem 1) prove the existence of blocking paths to stability in many-to-many matching markets where every agent of one side of the market has substitutable preferences and every agent of the other side of the market has responsive preferences with quota. The many-to-many matching market in Example 1 fulfills these preference conditions, but not all stable matchings are stochastically stable. In particular, blocking paths that do not “wander off” as in Theorem 1 do not always exist.

Example 1. Many-to-Many Matching: $SS \subsetneq S$

We consider a many-to-many matching market with three firms F_1, F_2, F_3 and three workers w_1, w_2, w_3 . Table 1 first lists agents’ preferences in its columns, e.g., firm F_1 ’s preferences are such that $\{w_3, w_2\} P_{F_1} \{w_3, w_1\} P_{F_1} \{w_2, w_1\} P_{F_1} \{w_3\} P_{F_1} \{w_2\} P_{F_1} \{w_1\} P_{F_1} \emptyset \dots$, where \dots means that for F_1 all other sets of workers are worse than having no workers. Note that firms have responsive preferences with quota 2 and workers have substitutable preferences. Second, we list the following matchings in the rows of Table 1: $\mu' = \{\{F_1, w_3\}, \{F_1, w_2\}, \{F_2, w_1\}, \{F_2, w_3\}, \{F_3, w_2\}, \{F_3, w_1\}\}$ and $\mu = \{\{F_1, w_1\}, \{F_2, w_2\}, \{F_3, w_3\}\}$. Both matchings are stable and therefore absorbing.

	firm F_1	firm F_2	firm F_3	worker w_1	worker w_2	worker w_3
μ				$\{F_1\}$	$\{F_2\}$	$\{F_3\}$
μ'	$\{w_3, w_2\}$	$\{w_1, w_3\}$	$\{w_2, w_1\}$	$\{F_2, F_3\}$	$\{F_1, F_3\}$	$\{F_2, F_1\}$
	$\{w_3, w_1\}$	$\{w_1, w_2\}$	$\{w_2, w_3\}$	$\{F_2\}$	$\{F_3\}$	$\{F_1\}$
	$\{w_2, w_1\}$	$\{w_3, w_2\}$	$\{w_1, w_3\}$	$\{F_3\}$	$\{F_1\}$	$\{F_2\}$
	$\{w_3\}$	$\{w_1\}$	$\{w_2\}$			
	$\{w_2\}$	$\{w_3\}$	$\{w_1\}$			
μ	$\{w_1\}$	$\{w_2\}$	$\{w_3\}$			
	\emptyset	\emptyset	\emptyset	\emptyset	\emptyset	\emptyset
	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots

Table 1: Example 1 – preferences and matchings μ and μ' .

To prove that μ is stochastically stable while μ' is not, we proceed in two steps. In Step 1, we construct a path from μ' to μ with one mistake. In Step 2, we show that there is no path from μ to μ' with only one mistake. Observe that μ and μ' are the only stable matchings. Hence, $\gamma(\mu) = 1$ and $\gamma(\mu') > 1$ such that μ is the unique stochastically stable matching (see Lemma 3).

We are therefore left with Steps 1 and 2.

Step 1. To construct a path from μ' to μ with one mistake, suppose the blocking dynamics is in μ' and a single agent, say F_1 , is selected. Then, by one mistake, $\{F_1, w_2\}$ is abandoned. This results in the matching $\mu_1 = \{\{F_1, w_3\}, \{F_2, w_1\}, \{F_2, w_3\}, \{F_3, w_1\}, \{F_3, w_2\}\}$ which can be blocked by $\{F_1, w_1\}$ resulting in $\mu_2 = \{\{F_1, w_3\}, \{F_1, w_1\}, \{F_2, w_3\}, \{F_3, w_2\}\}$. Note that μ_2 is blocked by $\{F_2, w_2\}$ resulting in $\mu_3 = \{\{F_1, w_3\}, \{F_1, w_1\}, \{F_2, w_3\}, \{F_2, w_2\}\}$. Finally, μ_3 can be blocked by $\{F_3, w_3\}$ resulting in $\mu_4 = \{\{F_1, w_1\}, \{F_2, w_2\}, \{F_3, w_3\}\} = \mu$.

Step 2. To show that there is no path with only one mistake from μ to μ' suppose the blocking dynamics is in μ . We distinguish between two cases: (i) a single agent, or (ii) a pair of agents is selected.

For case (i), suppose without loss of generality that F_1 is selected. Then, by one mistake, $\{F_1, w_1\}$ is abandoned and both agents are single. This results in the matching $\mu_1 = \{\{F_2, w_2\}, \{F_3, w_3\}\}$. Observe that this matching cannot be blocked by any single agent or any pair that includes w_2 or w_3 and the only blocking pairs are $\{F_1, w_1\}$, $\{F_2, w_1\}$, and $\{F_3, w_1\}$. If $\{F_1, w_1\}$ is selected, μ is reached. Blocking by $\{F_2, w_1\}$ results in $\mu_2 = \{\{F_2, w_1\}, \{F_2, w_2\}, \{F_3, w_3\}\}$. Matching μ_2 can only be blocked by $\{F_1, w_1\}$ and $\{F_3, w_1\}$. If $\{F_1, w_1\}$ is selected, μ is reached. Blocking by $\{F_3, w_1\}$ results in $\mu_3 = \{\{F_2, w_1\}, \{F_3, w_1\}, \{F_2, w_2\}, \{F_3, w_3\}\}$. But μ_3 can only be blocked by $\{F_1, w_1\}$ resulting in μ . A symmetric argument holds if μ_1 is blocked by $\{F_3, w_1\}$. Hence, all blocking paths from μ_1 lead back to μ .

For case (ii), suppose that the blocking dynamics is in μ and suppose without loss of generality that $\{F_1, w_2\}$ is selected. Then, by one mistake, $\{F_1, w_2\}$ is established. Now observe that the resulting matching, $\mu^* = \{\{F_1, w_1\}, \{F_1, w_2\}, \{F_2, w_2\}, \{F_3, w_3\}\}$, cannot be blocked by any pair that contains w_2 or w_3 , nor by any single agent except for w_2 . But in this case, blocking yields μ .

Hence, our example demonstrates that there is no structure comparable to Theorem 1 for many-to-many matching markets – even if the domain is restricted in such a way that paths to stability can be guaranteed. \diamond

In the absence of a blocking path to stability, singleton absorbing sets (i.e., stable matchings) and non-singleton absorbing sets can co-exist and are candidates for the set of stochastically stable matchings. The following example demonstrates that the topology of an absorbing set is not decisive with respect to stochastic stability. Example 2, taken from Bogomolnaia and Jackson (2002), analyzes a hedonic coalition formation game (see Banerjee et al., 2001; Bogomolnaia and Jackson, 2002) where both absorbing sets (a singleton and a non-singleton set) are stochastically stable.

Example 2. Hedonic Coalition Formation: $SS \supseteq S$

We consider hedonic coalition formation with 3 agents $N = \{1, 2, 3\}$. Table 2 lists agents' preferences in its columns, e.g., agent 1's preferences are such that $\{1, 2\} P_1 N P_1 \{1, 3\} P_1 \{1\}$.

agent 1	agent 2	agent 3
$\{1, 2\}$	$\{2, 3\}$	$\{1, 3\}$
N	N	N
$\{1, 3\}$	$\{1, 2\}$	$\{2, 3\}$
$\{1\}$	$\{2\}$	$\{3\}$

Table 2: Example 2 – preferences.

Observe that N is the unique stable (and core) coalition structure in this example and therefore forms a singleton absorbing set. Moreover, pairwise blocking generates the cycle $\{\{2, 3\}, \{1\}\} \rightarrow \{\{1, 3\}, \{2\}\} \rightarrow \{\{1, 2\}, \{3\}\} \rightarrow \{\{2, 3\}, \{1\}\}$. As each coalition structure in the cycle has only one blocking coalition, the cycle induces a (non-singleton) absorbing set. The only remaining coalition structure is the set of singletons which is obviously not part of an absorbing set.

To see that N and the cycle have a stochastic potential of one, we construct paths with one mistake from N to the cycle and a path with one mistake from the cycle to N .

Suppose the blocking dynamics is in N and the coalition $\{1, 2\}$ is selected. By mistake of agent 2, $\{1, 2\}$ blocks N and the resulting coalition structure is $\{\{1, 2\}, \{3\}\}$ which is a coalition structure in the cycle.

Now suppose the blocking dynamics is in $\{\{1, 2\}, \{3\}\}$ and N is proposed. By mistake of agent 1, N blocks $\{\{1, 2\}, \{3\}\}$. ◇

Appendix

A Markov Process Terminology

- A *Markov process* (X, T) is determined by a discrete state space X and a mapping $T : X \times X \rightarrow [0, 1]$ where $T(x, x')$ describes the probability that the state equals $x' \in X$ in period $t+1$ whenever it was in $x \in X$ in period t . Clearly, for all $x \in X$, $\sum_{x' \in X} T(x, x') = 1$. Here we restrict ourselves to *finite, time-homogeneous* Markov processes, i.e., X is a finite set and transition probabilities induced (and captured in T) do not depend on time.
- An *absorbing set* $A \subseteq X$ is a minimal set of states that once entered throughout the dynamic process is never abandoned.
- An absorbing set A is *aperiodic* whenever it does not contain any deterministic and non-trivial cycle, i.e., there is no sequence of at least two states $x_1, x_2, \dots, x_n \in A$ such that for all $i = 1, \dots, n-1$, $x_i \in A$ and $T(x_i, x_{i+1}) = T(x_n, x_1) = 1$. Note that a sufficient condition for the aperiodicity of an absorbing set A is that for some $x \in A$, $0 < T(x, x) < 1$, i.e., the Markov process exhibits sufficient (but not complete) inertia.
- By the *weak fundamental theorem of Markov processes* every aperiodic absorbing set $A \subseteq X$ corresponds to exactly one *invariant distribution* $p : X \rightarrow [0, 1]$ with $p \cdot T = p$ and support A , i.e., $\sum_{x \in A} p(x) = 1$. If all absorbing sets of a Markov process are aperiodic, then its set of invariant distributions is defined as the convex hull of the invariant distributions of its absorbing sets. The support of an invariant distribution p of such a Markov process is therefore a (non-empty) collection of absorbing sets.
- By the *fundamental theorem of Markov processes* the unique invariant distribution p that is induced by an aperiodic absorbing set $A \subseteq X$ describes the probability $p(x)$ that the process will be at state $x \in A$ if it reached a state in A and propagated forever, i.e., for all $x \in A$ and all probability distributions $\tilde{p} : A \rightarrow [0, 1]$, $p(x) = (\lim_{k \rightarrow \infty} \tilde{p} \cdot T^k)(x)$.
- A Markov process is *ergodic* if it has a unique absorbing set.
- A Markov process is *irreducible* if it is ergodic and the unique absorbing set coincides with the state space X .
- A *perturbed Markov process* (X, T^ϵ) is a Markov process such that for all $x, x' \in X$, $\lim_{\epsilon \rightarrow 0} T^\epsilon(x, x') = T(x, x')$, and for all $\epsilon > 0$, $T^\epsilon(x, x') > 0$ implies that there is an $r \geq 0$ with $0 < \lim_{\epsilon \rightarrow 0} \epsilon^{-r} T^\epsilon(x, x') < \infty$. Here we restrict ourselves to irreducible and aperiodic perturbed Markov processes. Therefore, the invariant distribution p^ϵ of (X, T^ϵ) is unique.
- The *limit invariant distribution* p^* of a Markov process (X, T) is the invariant distribution p^ϵ of a perturbed process (X, T^ϵ) in the limit of $\epsilon \rightarrow 0$. Young (1993, Theorem 4(i)) implies that $p^* := \lim_{\epsilon \rightarrow 0} p^\epsilon$ exists and is an invariant distribution of (X, T) .
- A state in the support of p^* is *stochastically stable*.

B Proof of Lemma 5

The proof parallels the proof of the Claim in Diamantoudi et al. (2004, Theorem 1). To extend their claim from solvable roommate markets to all roommate markets we assume that there is a blocking pair for μ that is not a blocking pair for μ' . (For solvable roommate markets there are no blocking pairs for μ' at all.) Step 9 of the proof crucially depends on this assumption. Moreover, we employ a slightly different measure for the distance between two matchings: Diamantoudi et al. (2004) count the numbers of pairs (including singletons) that are common to both matchings, whereas we count the number of agents who have the same mate under both matchings.¹³

Step 1. By Lemma 1(iii), $\mu' \in A$ implies $\mu' \in I(R)$.

Step 2. If μ is blocked by a pair that is matched in μ' , then this blocking pair induces a matching μ_2 with $m(\mu_2, \mu') \geq m(\mu, \mu') + 1$ and Lemma 5 is satisfied with $\mu_L = \mu_2$. Hence, we can make the following assumption.

D1. $\mu \in I(R)$ and there exists no blocking pair $\{i, j\}$, $i \neq j$, for μ such that $\mu'(i) = j$.

If μ' is blocked by a pair that is matched in μ , then this blocking pair induces a matching $\mu'_2 \in A$ with $m(\mu, \mu'_2) \geq m(\mu, \mu') + 1$; contradicting $m(\mu, \mu') = m(\mu, A)$. Hence, we can make the following assumption.

D2. $\mu' \in I(R)$ and there exists no blocking pair $\{i, j\}$, $i \neq j$, for μ' such that $\mu(i) = j$.

Step 3. Define $f : N \rightarrow N$ by

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

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

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

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

“ \Leftarrow ” follows from the definition of f .

“ \Rightarrow ” follows from $\mu, \mu' \in I(R)$ and the strictness of preferences.

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

$$[f(i) = \mu(i) \Rightarrow f(j) = \mu'(j)] \text{ and } [f(i) = \mu'(i) \Rightarrow f(j) = \mu(j)].$$

Obviously, both statements are true if $\mu(j) = \mu'(j) (= i)$. Hence, assume $\mu(j) \neq \mu'(j)$. To prove the first part, assume by contradiction that $f(i) = \mu(i)$ and $f(j) = \mu(j) \neq \mu'(j)$. This implies $i = \mu(j) P_j \mu'(j)$. But as $\mu'(j) \neq \mu(j) = i$, it follows that $\mu(i) \neq \mu'(i)$ and $j = \mu(i) P_i \mu'(i)$. This violates D2 because then $\{i, j\}$ is a blocking pair for μ' even though they are matched at μ . A symmetric argument proves the second part by using D1. This implies that for all $i \in N$,

$$[\mu(i) P_i \mu'(i) \Rightarrow \mu'(f(i)) P_{f(i)} \mu(f(i))] \text{ and } [\mu'(i) P_i \mu(i) \Rightarrow \mu(f(i)) P_{f(i)} \mu'(f(i))].$$

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

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

¹³In contrast to $m(\cdot, \cdot)$ as introduced in Section 4, the measure in Diamantoudi et al. (2004) cannot be normalized to a distance function on the set of matchings, because two identical and two distinct matchings may well have the same number of common pairs and singletons.

“ \Leftarrow ” follows directly from the definition of f .

“ \Rightarrow ” follows for $f(i) = i$ by Step 4. So, suppose that $f(i) = j \neq i$. Without loss of generality assume that $f(i) = j = \mu(i)$. Then, by Step 5, $f(j) = \mu'(j)$. The assumptions $f(f(i)) = i$ and $f(i) = j$ imply $f(j) = i$. Hence, $\mu'(j) = i$, and this together with $\mu(i) = j$ implies that i and j are mates at μ and μ' . Thus, $\mu(i) = \mu'(i)$.

Step 7. Denote by $f^t(i)$ the t -th application of f , i.e., $f^1(i) = f(i)$ and $f^t(i) = f(f^{t-1}(i))$. Then for all $i \in N$, there exists a $t \in \mathbb{N}$ such that $f^t(i) = i$.

In the sequence $\sigma_i = (i, f^1(i), f^2(i), \dots)$ the finiteness of N implies that at least some $j \in N$ has to appear more than once. Hence, there exists a \tilde{t} such that $f^{\tilde{t}}(i) = f^p(i)$ for some $p < \tilde{t}$. Let t be the minimal such \tilde{t} . Then, $f^t(i) = i$. To see this, suppose by way of contradiction that $f^t(i) \neq i$. Then the sequence σ_i equals

$$\sigma_i = (i = i_0, i_1, i_2, \dots, i_{p-1}, i_p, i_{p+1}, \dots, i_{t-1}, i_t = i_p, i_{p+1}, \dots) \text{ where } i_p \neq i.$$

By minimality of t , all agents in $\{i, i_1, \dots, i_{t-1}\}$ are distinct. Since $i_p = f(i_{p-1}) = f(i_{t-1})$, i_p has to be part of two distinct pairs in μ and μ' , respectively. Strict preferences imply that $f(i_p)$ is either i_{p-1} or i_{t-1} . In either case, however, $f(f(i_p)) = i_p$ and Step 6 implies $\mu(i_p) = \mu'(i_p)$. A contradiction.

As a result, each agent belongs to a unique cycle and the set of cycles induces a partition of N . We denote the cycle that i belongs to by c_i and the set of agents in c_i by $S_i \subseteq N$. The partition then equals $\{S_i\}_{i \in N}$.

Step 8. For all $i \in N$, if $\mu(i) \neq \mu'(i)$, then $|S_i| \geq 4$ and $|S_i|$ is even.

Steps 4 and 6 exclude cycles of length 1 or 2, respectively. Together with $|S_i|$ even, it follows that $|S_i| \geq 4$. The second part follows from the alternation property in Step 5. Suppose without loss of generality that agent $i_0 = i$ (strictly) prefers matching μ . Then, by Step 5, $i_1 = f(i)$ prefers μ' , $i_2 = f(f(i))$ prefers μ , etc. So, all agents i_h with h being an even number prefer μ and all agents i_h with h being an odd number prefer μ' . As $\mu'(i) \neq \mu(i)$, only an agent i_{h^*} with h^* being an odd number can close the cycle. Hence, $|S_i| = |\{i_0, i_1, \dots, i_{h^*}\}|$ is even.

Step 9. Recall that there is a blocking pair $\{r, s\} \subseteq N$ ($r \neq s$) for μ that is not a blocking pair for μ' . Hence, there is an agent $h \in \{r, s\}$ with $\mu' P_h \mu$ (if $\mu R_h \mu'$ for all $h \in \{r, s\}$ then $\{r, s\}$ also blocks μ' , which would be a contradiction). Without loss of generality let $r = 1$ prefer μ' over μ and denote $c_1 = (1, 2, \dots, p)$ (according to Step 8, $p \geq 4$ and even). For each $i \in \{1, 2, \dots, p-1\}$, Step 5 and $\mu' P_1 \mu$ imply that $\mu'(i) = i + 1$ if i is odd, $\mu(i) = i + 1$ if i is even, and $\mu(p) = 1$. Hence,

$$\begin{aligned} \mu' &= \{\{1, 2\}, \{3, 4\}, \dots, \{p-3, p-2\}, \{p-1, p\}, \dots\}, \\ \mu &= \{\{2, 3\}, \{4, 5\}, \dots, \{p-2, p-1\}, \{p, 1\}, \dots\}. \end{aligned}$$

Let μ_2 be the matching that results from $\mu_1 = \mu$ by matching blocking pair $\{r = 1, s\}$. As $f(2) = \mu(2) = 3$ by Step 5, and therefore 2 prefers $3 = \mu(2)$ over $1 = \mu'(2)$, it follows that $s \neq 2$ such that $\{1, s\}$ is not a pair in μ' . Hence,

$$m(\mu_2, \mu') = \begin{cases} m(\mu, \mu') & \text{if } \mu(s) \neq \mu'(s); \\ m(\mu, \mu') - 1 & \text{if } \mu(s) = \mu'(s) = s; \\ m(\mu, \mu') - 2 & \text{if } \mu(s) = \mu'(s) \neq s. \end{cases}$$

Note that at matching μ_2 , agent p is unmatched. Then, individual rationality of μ' implies $\mu' P_p \mu_2$ with $\mu'(p) = p - 1$. Note that

$$\mu_2(p-1) = \begin{cases} 1 & \text{if } s = p-1; \\ p-1 & \text{if } s = p-2; \\ p-2 & \text{if } s \neq p-1, p-2. \end{cases}$$

By Step 5, $f(p-1) = \mu'(p-1) = p$ and therefore $\mu' P_{p-1} \mu_2$ if $\mu_2(p-1) \in \{p-1, p-2\}$. Hence, $\{p-1, p\}$ blocks μ_2 if $s \neq p-1$.

Case 1. $s \neq p-1$. Then, $\{p-1, p\}$ blocks μ_2 . Let μ_3 be the matching that results from μ_2 by matching this blocking pair. If $\mu(s) \neq \mu'(s)$, then $m(\mu_3, \mu') = m(\mu, \mu') + 2$ as desired for Lemma 5.

Now suppose $\mu(s) = \mu'(s)$ (which implies $s \notin \{1, \dots, p\}$). Then $\mu_3 = \{\{2, 3\}, \{4, 5\}, \dots, \{p-2\}, \{p-1, p\}, \{1, s\}, \dots, \{\mu'(s)\}, \dots\}$ in case $\mu'(s) \neq s$ and $\mu_3 = \{\{2, 3\}, \{4, 5\}, \dots, \{p-2\}, \{p-1, p\}, \{1, s\}, \dots\}$ in case $\mu'(s) = s$. In either case, $m(\mu_3, \mu') \geq m(\mu, \mu')$. Observe that $p-2$ is unmatched in μ_3 .

Suppose that $p \geq 6$. Then, individual rationality of μ' and $f(p-3) = \mu'(p-3) = p-2$ implies that $\{p-3, p-2\}$ blocks μ_3 . Let μ_4 be the matching that results from μ_3 by matching this blocking pair. Then, $m(\mu_4, \mu') \geq m(\mu, \mu') + 2$ as desired for Lemma 5.

Now suppose $p < 6$, i.e., $p = 4$ (Step 8). Recall that $\{1, s\}$ blocks μ (and, hence, s prefers 1 over $\mu(s) = \mu'(s)$). Since $\{r = 1, s\}$ is not a blocking pair for μ' , agent 1 prefers $\mu'(1) = 2$ over s . But for $p = 4$ and $\mu(s) = \mu'(s)$, $\mu_3 = \{\{2\}, \{3, 4\}, \{1, s\}, \dots, \{\mu'(s)\}\}$ in case $\mu'(s) \neq s$ and $\mu_3 = \{\{2\}, \{3, 4\}, \{1, s\}, \dots\}$ in case $\mu'(s) = s$. In either case, μ_3 is blocked by $\{1, 2\}$. The matching that results from μ_3 by matching this blocking pair equals $\mu_4 = \{\{1, 2\}, \{3, 4\}, \dots, \{s\}, \dots, \{\mu'(s)\}\}$ in case $\mu'(s) \neq s$ and $\mu_4 = \{\{1, 2\}, \{3, 4\}, \dots, \{s\}, \dots\}$ in case $\mu'(s) = s$. Note that $m(\mu_4, \mu') \geq m(\mu, \mu') + 2$ as desired for Lemma 5.

Case 2. $s = p-1$ (which implies $\mu(s) \neq \mu'(s)$).

Suppose $p \geq 6$. Then,

$$\begin{aligned} \mu &= \{\{2, 3\}, \{4, 5\}, \dots, \{p-2, s\}, \{p, 1\}, \dots\}, \\ \mu_2 &= \{\{2, 3\}, \{4, 5\}, \dots, \{p-2\}, \{p\}, \{1, s\}, \dots\}, \end{aligned}$$

and the individual rationality of μ' together with $f(p-3) = \mu'(p-3) = p-2$ implies that $\{p-3, p-2\}$ blocks μ_2 . The matching that results from μ_2 by matching this blocking pair equals $\mu_3 = \{\{2, 3\}, \{4, 5\}, \dots, \{p-4\}, \{p-3, p-2\}, \{p\}, \{1, s\}, \dots\}$ and exhibits $m(\mu_3, \mu') = m(\mu, \mu') + 2$ as desired for Lemma 5.

Finally, suppose $p = 4$, i.e., $\{r, s\} = \{1, 3\}$. Then,

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

Since $\{r, s\} = \{1, 3\}$ does not block μ' , $2 P_1 3$ or $4 P_3 1$. Suppose without loss of generality that $2 P_1 3$. Then, $\{1, 2\}$ blocks μ_2 and the matching that results from μ_2 by matching this blocking pair equals $\mu_3 = \{\{1, 2\}, \{3\}, \{4\}, \dots\}$ and exhibits $m(\mu_3, \mu') = m(\mu, \mu') + 2$ as desired for Lemma 5.

C The Restricted Mistake Dynamics

In Section 3.2 we define perturbations of the dynamic process, or the *mistake dynamics*, as follows. We assume that in every period a pair $\{i, j\}$ (possibly $i \neq j$) is selected. If the pair is a blocking pair and not yet matched, then they can match without making a mistake. However, if the pair is not a blocking pair, we assume that they end up being matched (by mistake) with probability $\epsilon > 0$. This definition of the mistake dynamics coincides with that of Jackson and Watts (2002, p. 277) for network formation. However, in their analysis of stochastically stable matchings for marriage problems, Jackson and Watts (2002, Section 5) employ a more restrictive mistake dynamics:¹⁴ a pair $\{i, j\}$ with $i \neq j$ that is matched becomes by mistake unmatched in the next period with probability $\epsilon > 0$. In our model this corresponds to restricting mistakes to single agents, i.e., the only “pairs” $\{i, j\}$ that make mistakes are those with $i = j$ (who then by mistake matches with himself, i.e., he unmatched from his current partner). We refer to this dynamics as the *restricted mistake dynamics*. Hence, our original mistake dynamics allows for a larger set of mistakes than the restricted mistake dynamics; in particular, we allow for the matching of a non-blocking pair $\{i, j\}$ with $i \neq j$.

In the remainder of this appendix we consider the restricted mistake dynamics. This restriction only affects one argument in Step 2 (and the corresponding arguments in Steps 3 up to k) in the proof of Theorem 2. In the following, we show how these steps can be modified such that the proof of Theorem 2 also works for the restricted mistake dynamics.

Modified proof of Theorem 2. Consider Step 2 in the proof of Theorem 2 and let μ_1 and μ_2 be as defined there, i.e., $\mu_1 \in A_1$ and $\mu_2 \in A_2$ with $A_1, A_2 \in \mathcal{A}(R)$ such that μ_1 and μ_2 minimize the distance between A_1 and A_2 .

According to Theorems 6 and 7 in Iñarra et al. (2008a), the set of agents in a roommate problem R can be uniquely divided into *satisfied* and *dissatisfied* agents. In an absorbing set $A \in \mathcal{A}(R)$, any satisfied agent is matched with exactly one agent (possibly with himself), while any dissatisfied agent has at least two different matches in A . Moreover, the restriction of an absorbing set to the set of satisfied agents is a stable matching and the restrictions of any two absorbing sets to the set of dissatisfied agents are identical.¹⁵ Since μ_1 and μ_2 minimize the distance between A_1 and A_2 , dissatisfied agents are matched with the same partners in μ_1 and μ_2 . Hence, $\mu_1 \neq \mu_2$ implies that there exists a pair of satisfied agents $\{i, j\}$ such that $\mu_1(i) = j$, $\mu_2(i) = k$ and $k \neq j$.¹⁶ As the restrictions of μ_1 and μ_2 to the set of satisfied agents are stable, they are also decomposable, i.e., $j = \mu_1(i) P_i \mu_2(i) = k$ implies $l := \mu_2(j) P_j \mu_1(j) = i$ and $k = \mu_2(i) P_i \mu_1(i) = j$ implies $i = \mu_1(j) P_j \mu_2(j) = l$.¹⁷ Suppose, without loss of generality, that $j = \mu_1(i) P_i \mu_2(i) = k$ (and therefore $l = \mu_2(j) P_j \mu_1(j) = i$). Now consider the matching that results from μ_2 by unmatching $\{j, l\}$. In this matching, agent j is single and $\{i, j\}$ is a blocking pair (because i prefers j to his current mate k). Denote the matching that results from

¹⁴See the corresponding discussion in (Jackson and Watts, 2002, p. 285).

¹⁵This is for instance reflected by the figures that accompany the proof of Theorem 2: all absorbing sets have the same number of matchings.

¹⁶Similarly as for stable matchings we can conclude that any satisfied agent who is alone in some absorbing matching is also alone in all other absorbing matchings. For solvable roommate markets, this statement boils down to the well-known result that the set of single agents is the same for all stable matchings (see Gusfield and Irving, 1989, Theorem 4.5.2 or Klaus and Klijn, 2008, Theorem 1 for an elementary proof). As i and j are satisfied agents, we then can conclude that $i \neq j$.

¹⁷See Gusfield and Irving, 1989, Lemma 4.3.9 or Klaus and Klijn, 2008, Lemma 2 for an elementary proof.

matching $\{i, j\}$ by $\tilde{\mu}_2$. Note that $m(\mu_1, \tilde{\mu}_2) > m(\mu_1, \mu_2)$ and therefore, $m(A_1, \tilde{\mu}_2) > m(A_1, A_2)$. By construction, it takes only one mistake to connect μ_2 and $\tilde{\mu}_2$. Having constructed $\tilde{\mu}_2$, Step 2 of the proof proceeds in exactly the same way as in Section 4.2. Theorem 1 guarantees that there is a blocking path from $\tilde{\mu}_2$ to a matching $\hat{\mu}_1 \in A_1$ such that A_2 can be connected with $\mu \in A_1$ with one mistake.

Similarly, for Step $l \in \{3, \dots, k\}$, Iñarra et al. (2008b, Theorems 6 and 7) and decomposability imply the existence of a path with only one mistake from $\mu_l \in A_l$ with $m(A_1, \mu_l) = m(A_1, A_l)$ to $\tilde{\mu}_l$ with $m(A_1, \tilde{\mu}_l) > m(A_1, A_l)$. Again, Theorem 1 guarantees that there is a blocking path from $\tilde{\mu}_l$ to a matching $\hat{\mu} \in A_1 \cup \dots \cup A_{l-1}$ such that A_l can be connected with the μ -tree constructed in Steps 2 to $l-1$ with one mistake. The rest of the proof of Theorem 2 is not altered by the restriction of the mistake dynamics. Hence, for every $A \in \mathcal{A}(R)$ and $\mu \in A$, there exists a μ -tree with resistance $|\mathcal{A}(R)| - 1$ which implies that $A(R) = SS(R)$ and reestablishes Theorem 2 for the restricted mistake dynamics. \square

Note that the above proof utilizes structural properties of the roommate problem (decomposability and the structural properties of absorbing matchings concerning satisfied and dissatisfied agents as established by Iñarra et al., 2008b) that are not necessary if we use our original mistake dynamics. Having established Theorem 2 for the restricted mistake dynamics, we also obtain Corollaries 1 and 2 for the restricted mistake dynamics. In particular, we obtain Jackson and Watts's (2002) Theorem 2, as a corollary to Theorem 2 for the restricted mistake dynamics.

D A Roommate Market

The following example was kindly provided by Elena Molis.

Example 3. We consider a non-solvable roommate market with 8 agents $N = \{1, \dots, 8\}$. Table 3 lists agents' preferences in its columns, e.g., agent 1's preferences are such that $2 P_1 3 P_1 4 P_1 6 P_1 5 P_1 7 P_1 8 P_1 1$.

agent 1	agent 2	agent 3	agent 4	agent 5	agent 6	agent 7	agent 8
2	3	1	6	4	7	5	3
3	1	2	3	7	4	6	8
4	4	4	5	1	2	1	1
6	5	5	1	2	3	2	2
5	6	6	2	3	1	3	7
7	8	7	7	6	5	4	4
8	7	8	8	8	8	8	5
1	2	3	4	5	6	7	6

Table 3: Example 3 – preferences.

The set of absorbing matchings equals $A(R) = \{\mu_4, \mu_5, \mu_6, \mu_7\}$ and the set of P -stable matchings equals $PS(R) = \{\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6\}$ where

$$\begin{aligned}
 \mu_1 &= [\{1\}, \{2, 3\}, \{4, 5\}, \{6, 7\}, \{8\}], \\
 \mu_2 &= [\{1, 2\}, \{3\}, \{4, 5\}, \{6, 7\}, \{8\}], \\
 \mu_3 &= [\{1, 3\}, \{2\}, \{4, 5\}, \{6, 7\}, \{8\}], \\
 \mu_4 &= [\{1\}, \{2, 3\}, \{4, 6\}, \{5, 7\}, \{8\}], \\
 \mu_5 &= [\{1, 2\}, \{3\}, \{4, 6\}, \{5, 7\}, \{8\}], \\
 \mu_6 &= [\{1, 3\}, \{2\}, \{4, 6\}, \{5, 7\}, \{8\}], \\
 \mu_7 &= [\{1, 2\}, \{3, 8\}, \{4, 6\}, \{5, 7\}].
 \end{aligned}$$

Hence, $A(R) \subsetneq PS(R)$ and $A(R) \supsetneq PS(R)$. Furthermore, observe that $\mu_7 \notin MS(R)$ and $\mu_1 \in MS(R)$, which implies $A(R) \subsetneq MS(R)$ and $A(R) \supsetneq MS(R)$. Finally, since $\mu_5 \notin AS(R)$ and $\mu_1 \in AS(R)$ we also have $A(R) \subsetneq AS(R)$ and $A(R) \supsetneq AS(R)$. \diamond

References

- Abraham, D. J., Biró, P., and Manlove, D. F. (2006): ““Almost Stable” Matchings in the Roommates Problem.” In T. Erlebach and G. Persiano, editors, *Proceedings of WAOA 2005, Lecture Notes in Computer Science Vol. 3879*, pages 1–14. Springer, Berlin/Heidelberg.
- Banerjee, S., Konishi, H., and Sönmez, T. (2001): “Core in a Simple Coalition Formation Game.” *Social Choice and Welfare*, 18: 135–153.
- Bochet, O., Klaus, B., and Walzl, M. (2007): “Dynamic Recontracting Processes with Multiple Indivisible Goods.” Meteor Research Memorandum 07/018 (<http://ideas.repec.org/p/dgr/umamet/2007018.html>).
- Bogomolnaia, A. and Jackson, M. (2002): “The Stability of Hedonic Coalition Structures.” *Games and Economic Behavior*, 38: 201–230.
- Diamantoudi, E., Miyagawa, E., and Xue, L. (2004): “Random Paths to Stability in the Roommate Problem.” *Games and Economic Behavior*, 48: 18–28.
- Gale, D. and Shapley, L. S. (1962): “College Admissions and the Stability of Marriage.” *American Mathematical Monthly*, 69: 9–15.
- Gusfield, D. and Irving, R. W. (1989): *The Stable Marriage Problem: Structure and Algorithms*. The MIT Press, Cambridge.
- Iñarra, E., Larrea, C., and Molis, E. (2008a): “Random Paths to P -stability in the Roommate Problem.” *International Journal of Game Theory*, 36: 461–471.
- Iñarra, E., Larrea, C., and Molis, E. (2008b): “The Stability of the Roommate Problem Revisited.” Mimeo (February).
- Jackson, M. O. and Watts, A. (2002): “The Evolution of Social and Economic Networks.” *Journal of Economic Theory*, 106: 265–295.
- Kandori, M., Mailath, G. J., and Rob, R. (1993): “Learning, Mutation, and Long-Run Equilibria in Games.” *Econometrica*, 61: 29–56.
- Klaus, B. and Klijn, F. (2008): “Smith and Rawls Share a Room: Stability and Medians.” Meteor Research Memorandum 08/009 (<http://ideas.repec.org/p/dgr/umamet/2008009.html>).
- Kojima, F. and Ünver, U. (2008): “Random Paths to Pairwise Stability in Many-to-Many Matching Problems: A Study on Market Equilibration.” *International Journal of Game Theory*, 36: 473–488.
- Roth, A. E. and Vande Vate, J. (1990): “Random Paths to Stability in Two-Sided Matching.” *Econometrica*, 58: 1475–1480.
- Serrano, R. and Volij, O. (2007): “Mistakes in Cooperation: the Stochastic Stability of Edgeworth’s Recontracting.” *Economic Journal*, forthcoming.
- Tan, J. J. M. (1990): “A Maximum Stable Matching for the Roommates Problem.” *BIT*, 29: 631–640.
- Tan, J. J. M. (1991): “A Necessary and Sufficient Condition for the Existence of a Complete Stable Matching.” *Journal of Algorithms*, 12: 154–178.
- Young, H. P. (1993): “The Evolution of Conventions.” *Econometrica*, 61: 57–84.