Semi-stability in Roommate Problems*

P. Jean-Jacques Herings† Yu Zhou‡

June 30, 2026

Abstract

We study the roommate problem allowing for preferences with indifferences and agents remaining single. A stable outcome need not exist. We introduce semi-stability, a novel notion that guarantees existence and restores stability through minimal policy or group interventions. We build an equilibrium foundation for semi-stability by showing its equivalence to a newly proposed choice-based equilibrium concept. We show that undominated semi-stable outcomes are Pareto optimal and under strict preferences, every semi-stable outcome is Pareto optimal. Nevertheless, standard structural properties, such as the rural hospital theorem and the lattice properties, fail for semi-stable outcomes. We also discuss possible applications.

Keywords: Roommate problem, semi-stability, choice-based equilibrium, Pareto optimality.

JEL Classification: C72, C78, D47.

*This paper substantially supersedes the previously circulated version, "Harmonious Equilibria in Roommate Problems." We thank two anonymous referees and the co-editor, Rakesh Vohra, for helpful suggestions. We also thank Ata Atay, Francis Bloch, Peter Hammond, Daisuke Hirata, Michihiro Kandori, Manshu Khanna, Andy Mackenzie, Michael Richter, Ariel Rubinstein, Qianfeng Tang, Utku Ünver, and the participants at the CRETA conference, the 14th Conference on Economic Design, PHBS Conference on Economic Theory, the 18th meeting of the Society for Social Choice and Welfare, and various seminars for their helpful comments on the previously circulated version. Yu Zhou gratefully acknowledges financial support from the National Natural Science Foundation of China (72394391) and the Grant-in-aid for Research Activity, Japan Society for the Promotion of Science (22H00062, 26H01960), and JST ERATO (JPMJER2301).

†Department of Econometrics and Operations Research, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, the Netherlands. E-mail: P.J.J.Herings@tilburguniversity.edu

‡College of Business, Southern University of Science and Technology, 1088 Xueyuan Avenue, Nanshan District, Shenzhen, Guangdong, 518055, China; Graduate School of Economics, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601, Japan. E-mail: zhouyu_0105@hotmail.com

1 Introduction

We consider a version of the roommate problem in which agents have preferences over whom they wish to be paired with and are allowed to remain single if they prefer. Moreover, agents' preferences may include indifferences.[1] The goal is to partition the agents into pairs and singletons. Such a partition is called an outcome. Gale and Shapley (1962) show that for an outcome to be desirable, it must not admit a blocking pair, i.e., a pair of agents who would both be strictly better off being matched to each other rather than with their current partners. Unfortunately, a stable outcome need not exist. Even when one exists, it need not be efficient if preferences are allowed to have indifferences.

We propose semi-stability, a novel relaxation of stability that restores stability through minimal policy or group interventions and with favorable efficiency properaties. The literature has addressed non-existence either by restricting agents' preferences or by proposing various relaxations of stability that may fail to be efficient. Section 6 compares semi-stability with these alternative approaches.

Under semi-stability, certain pairs of agents, or matches, are not allowed to form, which prevents them from blocking an outcome. These excluded matches are endogenously determined by a minimality requirement. The exclusion of particular matches yields a stable outcome, but relaxing any of these restrictions would create instability. Semi-stability therefore requires a minimal level of policy or group intervention to restore stability. When a stable outcome exists, the notions of semi-stability and stability coincide. Otherwise, a semi-stable outcome still exists, addressing the non-existence issue.

The idea that a policymaker, like a matching agency, rules out certain pairs has also been proposed in a somewhat different context to address fairness issues. Dur, Gitmez, and Yılmaz (2019) study the school choice model and propose the notion of partial fairness. Partial fairness allows some priority violations, i.e., student-school blocking pairs, to be admissible, but these admissible violations are specified exogenously. In contrast, the excluded matches supporting a semi-stable outcome are determined endogenously by agents' preferences. Moreover, semi-stable outcomes exist in our setting, whereas partially fair outcomes need not exist.

In matching models, establishing the equivalence between competitive equilibrium and cooperative game-theoretic solution concepts is of great interest (see, e.g., Shapley and Shubik (1971), Quinzii (1984), Hatfield, Kominers, Nichifor, Ostrovsky, and Westkamp (2013), Herings (2018), and Candogan, Epitropou, and Vohra (2021)). In this paper, we propose a choice-based equilibrium concept that generalizes the idea underlying competitive equilibrium to settings without prices. Instead of choosing from a budget set, each agent chooses optimally from an option set. Policymakers or groups choose institutional arrangements that determine these option sets through endogenous, agent-specific

[1] Gale and Shapley (1962) introduce the roommate problem with an even number of agents who have strict preferences and prefer being paired to remaining single.

exclusions of matches. A maximal aggregate permission equilibrium (MAPE) consists of an outcome and a profile of such exclusions. MAPE requires the arrangements to induce exclusions that make agents' optimal choices mutually compatible while making the aggregate availability of choices as large as possible. We show that MAPE outcomes coincide with semi-stable outcomes. This result implies that the minimal intervention needed to restore stability is equivalent to maximizing the aggregate availability of choices.

MAPE is related to taboo equilibrium, introduced by Richter and Rubinstein (2024a, 2024b) for roommate problems. Taboo equilibrium differs from MAPE both in the constraints it imposes on agents' choices and in its equivalence to individually rational and Pareto-optimal outcomes.[2] We also compare MAPE with expectational equilibrium, introduced by Herings (2024) for many-to-one matching problems with contracts. Adapting this concept to roommate problems, we show that expectational equilibrium outcomes coincide with stable outcomes. Thus, when a stable outcome exists, the sets of semi-stable outcomes, MAPE outcomes, and expectational equilibrium outcomes all coincide.

We next turn to the efficiency properties of semi-stable outcomes. We show that semi-stable outcomes are weakly Pareto optimal, but need not be Pareto optimal. An undominated semi-stable outcome is defined as a semi-stable outcome that is not dominated by another semi-stable outcome. We show that undominated semi-stable outcomes exist and that they are Pareto optimal. Under strict preferences, all semi-stable outcomes are undominated and therefore Pareto optimal.

We also study structural properties of semi-stable outcomes. First, perturbing indifferences may generate additional semi-stable outcomes in the resulting economies with strict preferences, which are not semi-stable in the original economy. Second, except in special cases, semi-stable outcomes satisfy neither the rural hospital theorem nor the three versions of the lattice properties studied by Gusfield and Irving (1989).

Finally, we discuss some possible applications of our results. The pairwise kidney exchange models studied by Roth, Sönmez, and Ünver (2005) and Sönmez and Ünver (2014) are special cases of our set-up. In such models, undominated semi-stable outcomes coincide with individually rational and Pareto-optimal outcomes. We also discuss the potential applicability of our results to the U.S. Army's battle buddy system.

This paper contributes to the existing literature in the following ways. First, it provides a new relaxation of stability to address the non-existence issue in roommate problems. This notion is grounded in a novel conceptual view that restores stability through minimal intervention while retaining an efficiency rationale. Second, the established equivalence result strengthens the equilibrium interpretation of semi-stability and extends classical equivalence results from matching models with money to settings without money. Finally, the analysis of efficiency and structural properties clarifies the distinctive features of semi-stability and its relevance for practical design problems. An open issue

[2] In an earlier circulated version of this paper, Herings and Zhou (2025) provide a detailed discussion of how MAPE differs from taboo equilibrium and its variants.

is to establish the computational complexity of finding a semi-stable outcome, as well as verifying whether a given outcome is semi-stable.

The remainder of the paper is organized as follows. Section 2 introduces the model and the notion of semi-stability. Section 3 proposes the choice-based equilibrium concepts and establishes their equivalence to semi-stable outcomes. Section 4 studies the efficiency properties of semi-stable outcomes, while Section 5 shows that they do not satisfy several structural properties. In Section 6, we compare our approach with existing approaches that address the non-existence of stable outcomes. Section 7 discusses potential applications. Section 8 presents a short conclusion.

2 Roommate Problems and Semi-stability

Let $N = \{1, \ldots, n\}$ be a finite set of agents. Every agent $i \in N$ has a preference relation $\succsim^i$ defined over N. Preference relations are assumed to be complete and transitive. They are allowed to have indifferences. The associated strict preference and indifference relations are denoted by $\succ^i$ and $\sim^i$, respectively. For $j, k \in N$, $j \succsim^i k$ means that agent i weakly prefers a match with agent j to a match with agent k; $j \succ^i k$ means that i strictly prefers a match with agent j to a match with agent k; and $j \sim^i k$ means that i is indifferent between matching with agent j or agent k. Preferences are said to be strict if the agent is never indifferent between two distinct agents. Let $\succsim = (\succsim^i)_{i \in N}$ denote the profile of preference relations.

A *match* between agents i and j in N is denoted by the set $\{i, j\}$, and is simply written as ij. As the order does not matter, ij and ji indicate the same match. It is possible that $i = j$, which corresponds to the case where agent i remains unmatched. Let $D = \{ij \mid i, j \in N \text{ and } i \neq j\}$ denote the set of all possible matches between two distinct agents, and $I = \{ii \mid i \in N\}$ denote the set of all self-matches. We denote the set of all possible matches by $\overline{Y} = D \cup I$. For every $Y \subseteq \overline{Y}$, the set of matches involving agent i is denoted by $Y^i = \{y \in Y \mid i \in y\}$.

An *outcome* is a mapping $\mu : N \to N$ that assigns to every agent $i \in N$ a partner $\mu(i) \in N$ such that if agent i is matched with agent j, then agent j must be matched with agent i, i.e., for all $i, j \in N$, if $\mu(i) = j$, then $\mu(j) = i$. A mapping $\mu : N \to N$ is an outcome if and only if the collection of matches

$$\mathcal{P}(\mu) = \{ij \in \overline{Y} \mid j = \mu(i)\}$$

is a partition of N. Clearly, any partition $\mathcal{P}$ of N in singletons and pairs uniquely defines the corresponding outcome μ. Let $\mathcal{M}$ denote the set of all possible outcomes.

The primitives of the economy are summarized by $\mathcal{E} = (N, \overline{Y}, \succsim)$.

We next define the notion of stability, a central concept in the matching literature (see, e.g., Roth and Sotomayor (1990)). This notion is known as weak stability in roommate problems allowing for indifferences (see, e.g., Irving and Manlove (2002) and Manlove (2013)).

Definition 2.1 An outcome $\mu \in \mathcal{M}$ of the economy $\mathcal{E} = (N, \overline{Y}, \succsim)$ is *stable* if

(i) for every $i \in N$, $\mu(i) \succsim^i i$,

(ii) there is no pair of distinct agents $ij \in D$ such that $j \succ^i \mu(i)$ and $i \succ^j \mu(j)$.

Condition (i) of Definition 2.1 is known as *individual rationality*. It requires that an agent's partner is weakly preferred to the option of remaining single. Condition (ii) of Definition 2.1 is known as the *no blocking pair* condition. It says that no pair of distinct agents can benefit from leaving their current partner and matching together. We denote the collection of stable outcomes of an economy $\mathcal{E}$ by $\mathcal{S}^*(\mathcal{E})$.

Stable outcomes may not exist in roommate problems (Gale and Shapley, 1962). The following example demonstrates this non-existence issue and illustrates our notation.

Example 2.2 Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy with $N = \{1, 2, 3\}$, so $\overline{Y} = \{11, 12, 13, 22, 23, 33\}$. Also, we have that $D = \{12, 13, 23\}$ and $I = \{11, 22, 33\}$. The set of pairs of distinct agents involving agent 1 is $D^1 = \{12, 13\}$ and the set $I^2 = \{22\}$ corresponds to self-matching by agent 2. Consider a preference profile given by

$$2 \succ^1 3 \succ^1 1,$$
$$3 \succ^2 1 \succ^2 2,$$
$$1 \succ^3 2 \succ^3 3.$$

The set $\mathcal{M}$ consists of four outcomes. Either one of the three possible pairs forms or all agents remain single. For instance, the outcome $(\mu(1), \mu(2), \mu(3)) = (2, 1, 3)$ induces the partition $\mathcal{P}(\mu) = \{12, 3\}$, where agents 1 and 2 form a pair and agent 3 remains single. Clearly, none of the outcomes in $\mathcal{M}$ is stable, i.e., $\mathcal{S}^*(\mathcal{E}) = \emptyset$. $\triangle$

We now introduce a novel, weaker notion of stability, semi-stability, based on interventions in who can match with whom. The strongest intervention would prohibit any agent from matching with any other agent, allowing only self-matching. This trivially guarantees stability in the reduced economy resulting from the interventions, in which self-matching is the only admissible option, but it does so at the cost of eliminating all mutually beneficial matches. Such an outcome is undesirable, as it is not only inefficient but also disregards agents' preferences. Instead, we require that interventions be minimal.

Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy. For every set $C \subseteq D$ of excluded matches, we define the reduced economy $\mathcal{E}_{-C} = (N, \overline{Y} \setminus C, \succsim)$. In the reduced economy $\mathcal{E}_{-C}$, $\overline{Y} \setminus C$ is the set of permissible matches. Obviously, if $C = \emptyset$, then $\mathcal{E}_{-C} = \mathcal{E}$.

An *outcome* in the reduced economy $\mathcal{E}_{-C}$ does not use matches in the set C, so is an element of the set

$$\mathcal{M}_{-C} = \{\mu \in \mathcal{M} \mid \text{ for every } i \in N,\ i\mu(i) \in \overline{Y} \setminus C\}.$$

A mapping $\mu : N \to N$ is an outcome in the reduced economy $\mathcal{E}_{-C}$ if and only if the collection of matches

$$\{ij \in \overline{Y} \setminus C \mid j = \mu(i)\}$$

is a partition of N such that no member of the partition belongs to C.

We can now extend Definition 2.1 to allow for excluded matches.

Definition 2.3 Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy and $C \subseteq D$ be a set of excluded matches. An outcome $\mu \in \mathcal{M}_{-C}$ of the economy $\mathcal{E}_{-C} = (N, \overline{Y} \setminus C, \succsim)$ is *stable* if

(i) for every $i \in N$, $\mu(i) \succsim^i i$,

(ii) there is no pair of distinct agents $ij \in D \setminus C$ such that $j \succ^i \mu(i)$ and $i \succ^j \mu(j)$.

The next proposition shows that an arbitrary individually rational outcome $\mu \in \mathcal{M}$ can be turned into a stable outcome of the reduced economy $\mathcal{E}_{-C}$ by excluding a set of blocking pairs $C \subseteq D$ identified at μ.

Proposition 2.4 *Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy. If μ is an individually rational outcome of $\mathcal{E}$ and $C = \{ij \in D \mid j \succ^i \mu(i)$ and $i \succ^j \mu(j)\}$, then $\mu \in \mathcal{S}^*(\mathcal{E}_{-C})$.*

Proof. Clearly, μ is individually rational, so condition (i) of Definition 2.3 is satisfied. If $ij \in D$ is such that $j \succ^i \mu(i)$ and $i \succ^j \mu(j)$, then, by definition, $ij \in C$, so condition (ii) of Definition 2.3 is satisfied. ∎

We define an outcome to be semi-stable if it is a stable outcome of a reduced economy that results from excluding a minimal set of matches.

Definition 2.5 An outcome $\mu \in \mathcal{M}$ of the economy $\mathcal{E} = (N, \overline{Y}, \succsim)$ is *semi-stable* if

(i) there is $C \subseteq D$ such that $\mu \in \mathcal{S}^*(\mathcal{E}_{-C})$,

(ii) there does not exist $C' \subsetneq C$ such that $\mathcal{S}^*(\mathcal{E}_{-C'}) \neq \emptyset$.

The set C of excluded matches can be interpreted as a policy or group intervention. Condition (i) requires the reduced economy $\mathcal{E}_{-C}$, with permissible matches $\overline{Y} \setminus C$, to admit a stable outcome after the intervention. Condition (ii) makes the intervention minimal: relaxing any match in C leads to instability. We also say that C *supports* the semi-stability of the outcome μ.

We denote the set of semi-stable outcomes of an economy $\mathcal{E}$ by $\mathcal{S}(\mathcal{E})$. It is obvious that $\mathcal{S}^*(\mathcal{E}) \subseteq \mathcal{S}(\mathcal{E})$. The set of stable outcomes can be empty, but the next result shows that the set of semi-stable outcomes cannot. Moreover, if the economy has a stable outcome, then the set of semi-stable outcomes coincides with the set of stable outcomes.

Proposition 2.6 *Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy. The set of semi-stable outcomes of $\mathcal{E}$ is non-empty, i.e., $\mathcal{S}(\mathcal{E}) \neq \emptyset$. Moreover, if the set of stable outcomes of $\mathcal{E}$ is non-empty, i.e., $\mathcal{S}^*(\mathcal{E}) \neq \emptyset$, then $\mathcal{S}^*(\mathcal{E}) = \mathcal{S}(\mathcal{E})$.*

Proof. Observe that $\mathcal{E}_{-D}$ has a unique outcome in which all agents remain single. This outcome is stable in $\mathcal{E}_{-D}$ as it trivially satisfies individual rationality and has no blocking pairs in $D \setminus D = \emptyset$. Since the collection of subsets of D is finite, there is a minimal subset $C \subseteq D$ such that $\mathcal{E}_{-C}$ has a stable outcome. Any such stable outcome belongs to $\mathcal{S}(\mathcal{E})$.

It is clear that $\mathcal{S}^*(\mathcal{E}) \subseteq \mathcal{S}(\mathcal{E})$. Now assume $\mathcal{S}^*(\mathcal{E}) \neq \emptyset$. Suppose there is $C \neq \emptyset$ that supports some $\mu \in \mathcal{S}(\mathcal{E})$, then $\mu \in \mathcal{S}^*(\mathcal{E}_{-C})$. Since $\mathcal{S}^*(\mathcal{E}) \neq \emptyset$, condition (ii) of Definition 2.5 does not hold, contradicting $\mu \in \mathcal{S}(\mathcal{E})$. Consequently, $\mathcal{S}(\mathcal{E}) = \mathcal{S}^*(\mathcal{E})$. ■

The following example illustrates the concept of semi-stability.

Example 2.7 Consider the economy $\mathcal{E}$ of Example 2.2 and the outcome $\mu \in \mathcal{M}$ with $\mathcal{P}(\mu) = \{12, 3\}$. Let $C = \{23\}$, so $\overline{Y} \setminus C = \{11, 12, 13, 22, 33\}$. We show that μ is semi-stable in $\mathcal{E}$. First, we argue that $\mu \in \mathcal{S}^*(\mathcal{E}_{-\mathcal{C}})$: agent 1 is matched with agent 2, agent 1's optimal choice; agent 2 cannot match with agent 3, and is matched with agent 1, the optimal choice among permissible partners; and agent 3 remains single. Example 2.2 shows that $\mathcal{S}^*(\mathcal{E}) = \emptyset$. Thus, C is minimal, so μ is semi-stable.

In the above economy, there are two other semi-stable outcomes, μ' and μ'', where $\mathcal{P}(\mu') = \{13, 2\}$ and $\mathcal{P}(\mu'') = \{23, 1\}$. △

Condition (ii) of Definition 2.1 requires both members of a blocking pair to improve strictly. In roommate problems, there are more demanding notions of stability such as strong stability, where at least one member of the blocking pair improves strictly but the other only needs to improve weakly, and super stability where both members of the blocking pair improve weakly (see, e.g., Irving and Manlove, 2002; Manlove, 2013; Kunysz, 2016). These stronger concepts do not resolve the non-existence problem.[3] As far as existence is concerned, the methodological approach behind semi-stability, an intervention by excluding a minimal set of matches, could also be applied to these concepts. We leave the detailed analysis for future research.

3 Choice-based Equilibrium Concepts and Semi-stability

A classical theme in economics is the equivalence between competitive equilibria and cooperative game-theoretic concepts. In a competitive environment, an agent chooses the best element in the budget set given the current price system. Since our model has no monetary transfers, we replace budget sets with option sets. We study two ways to determine these option sets: through policy or group interventions, and through agents' expectations. The resulting choice-based equilibrium concepts provide equilibrium foundations for semi-stable and stable outcomes, respectively.

[3]That being said, strongly stable outcomes and super stable outcomes satisfy structural properties such as the rural hospital theorem (see, e.g., Subsection 4.5 in Manlove (2013)) and the property studied in Corollary 5.3. In contrast, these properties do not hold for the semi-stable outcomes; see Section 5.

We first define the option set. For every $i \in N$, recall that D^i denotes the set of matches involving agent i and another agent. Let $R^i \subseteq D^i$ denote a set of excluded matches for agent i. Thus, $\overline{Y}^i \setminus R^i$ is agent i's set of permissible matches. Let 2^{D^i} denote the collection of all possible subsets of D^i, so $R^i \in 2^{D^i}$. A profile of excluded matches $R = (R^1, \ldots, R^n)$ belongs to $\Pi_{i \in N} 2^{D^i}$. Note that it is allowed that $ij \in R^i$ but $ij \notin R^j$, i.e., the match ij is excluded for agent i but not for agent j.

The *option set* of agent $i \in N$, given the excluded matches $R^i \subseteq D^i$, is the set of agents with whom agent i can match, defined by

$$\gamma^i(R^i) = \{j \in N \mid ij \in \overline{Y}^i \setminus R^i\}.$$

Since self-matching is never excluded, i.e., $i \in \gamma^i(R^i)$, it follows that $\gamma^i(R^i)$ is non-empty. Accordingly, the *demand set* of agent i given the excluded matches $R^i \subseteq D^i$ is equal to

$$\delta^i(R^i) = \{j \in \gamma^i(R^i) \mid \forall k \in \gamma^i(R^i),\ j \succsim^i k\}.$$

Here, $\delta^i(R^i)$ is the set of $\succsim^i$-maximal agents in agent i's option set.

We begin by defining the notion of para-equilibrium.

Definition 3.1 A *para-equilibrium* of the economy $\mathcal{E} = (N, \overline{Y}, \succsim)$ is a profile of excluded matches and an outcome $(R, \mu) \in \Pi_{i \in N} 2^{D^i} \times \mathcal{M}$ such that, for every $i \in N$, $\mu(i) \in \delta^i(R^i)$.

At a para-equilibrium, each agent is assigned a partner from their demand set.

Next we define a maximal aggregate permission equilibrium.

Definition 3.2 A *maximal aggregate permission equilibrium* (MAPE) of the economy $\mathcal{E} = (N, \overline{Y}, \succsim)$ is a profile of excluded matches and an outcome $(R, \mu) \in \Pi_{i \in N} 2^{D^i} \times \mathcal{M}$ such that

(i) (R, μ) is a para-equilibrium of $\mathcal{E}$,

(ii) there is no para-equilibrium (Q, ν) of $\mathcal{E}$ such that $\cup_{i \in N}(\overline{Y}^i \setminus R^i) \subsetneq \cup_{i \in N}(\overline{Y}^i \setminus Q^i)$.

A MAPE is a para-equilibrium with a maximal aggregate set of permissible matches $\cup_{i \in N}(\overline{Y}^i \setminus R^i)$. The profile R can be interpreted as a policy variable that broadens the available options as much as possible, or as a group matching rule designed to be as lenient as possible. Broader options reduce unnecessary paternalism and allow agents with diverse preferences to pursue outcomes better suited to their needs. At the same time, too few restrictions may make agents' optimal choices mutually incompatible. Thus, the policymaker's interventions or the group's rules play a role analogous to prices in a competitive environment, coordinating individual choices.

The following example illustrates the concept of MAPE.

Example 3.3 Consider the economy of Example 2.2. Let $(R,\mu) \in \prod_{i\in N} 2^{D^i} \times \mathcal{M}$ be such that $R^1 = \emptyset$, $R^2 = \{23\}$, $R^3 = \{13, 23\}$, and $\mathcal{P}(\mu) = \{12, 3\}$. At (R,μ) it holds that $\overline{Y}^1 \setminus R^1 = \{11, 12, 13\}$, $\overline{Y}^2 \setminus R^2 = \{12, 22\}$, and $\overline{Y}^3 \setminus R^3 = \{33\}$. The option sets of the agents are given by $\gamma^1(R^1) = \{1, 2, 3\}$, $\gamma^2(R^2) = \{1, 2\}$, and $\gamma^3(R^3) = \{3\}$.

Below, we show that (R,μ) is a MAPE. First, observe that all agents make their best choice from their option sets. For example, for agent 3, when agents 1 and 2 are not in $\gamma^3(R^3)$, self-matching is the best choice. Thus (R,μ) is a para-equilibrium.

Next, we show the maximality of $\cup_{i\in N}(\overline{Y}^i \backslash R^i) = \overline{Y} \setminus \{23\}$. By contradiction, suppose that there is a para-equilibrium (Q,ν) such that $\cup_{i\in N}(\overline{Y}^i \backslash Q^i) = \overline{Y}$. The outcome ν cannot consist only of self-matches. If it were, then to make ν a para-equilibrium, we would need $Q^1 = D^1$, $Q^2 = D^2$, and $Q^3 = D^3$, contradicting $\cup_{i\in N}(\overline{Y}^i \backslash Q^i) = \overline{Y}$. Thus, under ν one pair of agents is matched, say $ij \in D$, and the third agent, say $k \in N \setminus \{i, j\}$, remains single. To make (Q,ν) a para-equilibrium, we need $k \in \delta^k(Q^k)$, so $Q^k = \{ik, jk\}$. Moreover, since either i or j has k as the top choice, it holds that either $ik \in Q^i$ or $jk \in Q^j$, contradicting $\cup_{i\in N}(\overline{Y}^i \backslash Q^i) = \overline{Y}$. △

Two MAPEs of an economy may have different aggregate sets of permissible matches. We illustrate this in the next example.

Example 3.4 Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy with $N = \{1, 2, 3, 4, 5, 6, 7\}$. Agents have strict preferences with individually rational parts given by

$$\begin{aligned}
&3 \succ^1 7 \succ^1 1\\
&3 \succ^2 2,\\
&1 \succ^3 2 \succ^3 4 \succ^3 3,\\
&3 \succ^4 5 \succ^4 6 \succ^4 4,\\
&6 \succ^5 4 \succ^5 5,\\
&4 \succ^6 5 \succ^6 6,\\
&1 \succ^7 7.
\end{aligned}$$

The MAPE (R,μ) is such that $R^1 = \{13\}$, $R^2 = \{23\}$, $R^3 = \{13, 23\}$, $R^6 = \{46\}$, $R^4 = R^5 = R^7 = \emptyset$, and $\mathcal{P}(\mu) = \{17, 34, 56, 2\}$. The MAPE (Q,ν) is such that $Q^2 = \{23\}$, $Q^4 = \{34\}$, $Q^5 = \{56\}$, $Q^6 = \{46, 56\}$, $Q^7 = \{17\}$, $Q^1 = Q^3 = \emptyset$, and $\mathcal{P}(\nu) = \{13, 45, 2, 6, 7\}$. The MAPE (R,μ) has an aggregate set of permissible matches equal to $\overline{Y} \setminus \{13, 23\}$, whereas the aggregate set of permissible matches at the MAPE (Q,ν) is equal to $\overline{Y} \setminus \{56\}$. △

Our next aim is to establish the equivalence between MAPE outcomes and semi-stable outcomes. We first consider the case with no aggregate restrictions, and show that MAPE outcomes coincide with stable outcomes. To derive these results, it is convenient to generalize a MAPE of $\mathcal{E}$ to reduced economies $\mathcal{E}_{-C} = (N, \overline{Y} \setminus C, \succsim)$ for $C \subseteq D$. The corresponding option and demand sets of agent $i \in N$ are denoted by γ^i_{-C} and δ^i_{-C}, respectively, and are defined by

$$\begin{array}{rcll} \gamma^i_{-C}(R^i) & = & \left\{j \in N \,\middle|\, ij \in (\overline{Y} \setminus C)^i \setminus R^i\right\}, & R^i \subseteq D^i \setminus C^i, \\ \delta^i_{-C}(R^i) & = & \{j \in \gamma^i_{-C}(R^i) \mid \forall k \in \gamma^i_{-C}(R^i),\ j \succsim^i k\}, & R^i \subseteq D^i \setminus C^i. \end{array}$$

Definition 3.1 of a para-equilibrium of the economy $\mathcal{E}$ is extended to economies $\mathcal{E}_{-C}$ for $C \subseteq D$ in the following definition.

Definition 3.5 Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy and $C \subseteq D$ be a set of excluded matches. A *para-equilibrium* of the economy $\mathcal{E}_{-C}$ is $(R, \mu) \in \Pi_{i \in N} 2^{D^i \setminus C^i} \times \mathcal{M}_{-C}$ such that, for every $i \in N$, $\mu(i) \in \delta^i_{-C}(R^i)$.

Since the matches in C cannot be formed in the economy $\mathcal{E}_{-C}$, an outcome is required to belong to $\mathcal{M}_{-C}$ and the endogenously excluded matches R^i for agent $i \in N$ are a subset of $D^i \setminus C^i$.

Next we extend Definition 3.2 of a maximal aggregate permission equilibrium of an economy $\mathcal{E}$ to economies $\mathcal{E}_{-C}$ where $C \subseteq D$.

Definition 3.6 Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy and $C \subseteq D$ be a set of excluded matches. A *maximal aggregate permission equilibrium* (MAPE) of the economy $\mathcal{E}_{-C}$ is $(R, \mu) \in \Pi_{i \in N} 2^{D^i \setminus C^i} \times \mathcal{M}_{-C}$ such that

(i) (R, μ) is a para-equilibrium of $\mathcal{E}_{-C}$,

(ii) there is no para-equilibrium (Q, ν) of $\mathcal{E}_{-C}$ such that

$$\cup_{i \in N} \left((\overline{Y} \setminus C)^i \setminus R^i\right) \subsetneq \cup_{i \in N} \left((\overline{Y} \setminus C)^i \setminus Q^i\right).$$

The following lemma builds connections between equilibria of $\mathcal{E}$ and those of $\mathcal{E}_{-C}$.

Lemma 3.7 *Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy.*

(i) If for some MAPE (R, μ) of $\mathcal{E}$ it holds that $\cup_{i \in N}(\overline{Y}^i \setminus R^i) = \overline{Y}$, then for every MAPE (Q, ν) of $\mathcal{E}$ it holds that $\cup_{i \in N}(\overline{Y}^i \setminus Q^i) = \overline{Y}$.

(ii) Let (R, μ) be a para-equilibrium of $\mathcal{E}$ and let $C = \{ij \in D \mid ij \in R^i \text{ and } ij \in R^j\}$. Then $((R^i \setminus C^i)_{i \in N}, \mu)$ is a MAPE of $\mathcal{E}_{-C}$ such that $\cup_{i \in N} \left((\overline{Y} \setminus C)^i \setminus (R^i \setminus C^i)\right) = \overline{Y} \setminus C$.

(iii) Let $C \subseteq D$. If (R, μ) is a para-equilibrium of $\mathcal{E}_{-C}$, then $((R^i \cup C^i)_{i \in N}, \mu)$ is a para-equilibrium of $\mathcal{E}$.

Proof.

Part (i). Let (R, μ) be a MAPE of $\mathcal{E}$ such that $\cup_{i \in N}(\overline{Y}^i \setminus R^i) = \overline{Y}$. For every MAPE (Q, ν) of $\mathcal{E}$, it cannot be the case that $\cup_{i \in N}(\overline{Y}^i \setminus Q^i) \subsetneq \overline{Y}$ by condition (ii) of Definition 3.2, so $\cup_{i \in N}(\overline{Y}^i \setminus Q^i) = \overline{Y}$.

Part (ii). It clearly holds that $((R^i \setminus C^i)_{i\in N}, \mu) \in \prod_{i\in N} 2^{D^i\setminus C^i} \times \mathcal{M}_{-C}$. For every $i \in N$, we have that $\gamma^i(R^i) = \gamma^i_{-C}(R^i \setminus C^i)$, so

$$\mu(i) \in \delta^i(R^i) = \delta^i_{-C}(R^i \setminus C^i).$$

Thus, $((R^i \setminus C^i)_{i\in N}, \mu)$ is a para-equilibrium of $\mathcal{E}_{-C}$. Moreover, it holds that

$$\cup_{i\in N} \left((\overline{Y} \setminus C)^i \setminus (R^i \setminus C^i)\right) = \cup_{i\in N}(\overline{Y}^i \setminus R^i) = \overline{Y} \setminus C.$$

Thus $((R^i \setminus C^i)_{i\in N}, \mu)$ satisfies condition (ii) of Definition 3.6, so it is a MAPE of $\mathcal{E}_{-C}$.

Part (iii). Since for every $i \in N$, $\gamma^i_{-C}(R^i) = \gamma^i(R^i \cup C^i)$, the conclusion follows. ■

Part (i) of Lemma 3.7 states that if there is a MAPE with an aggregate set of permissible matches equal to $\overline{Y}$, then every MAPE has this property. Part (ii) of Lemma 3.7 says that for any para-equilibrium (R, μ) of $\mathcal{E}$, removing the set of matches C that are *mutually* excluded by both involved parties leads to a reduced economy $\mathcal{E}_{-C}$. In this reduced economy, μ, paired with a profile of excluded matches $(R^i \setminus C^i)_{i\in N}$, is a MAPE with aggregate set of permissible matches equal to $\overline{Y} \setminus C$. Part (iii) of Lemma 3.7 says that if we have a para-equilibrium in a reduced economy $\mathcal{E}_{-C}$, then adding the excluded matches in C^i to those excluded for agent i in this para-equilibrium leads to a para-equilibrium in $\mathcal{E}$. This is because the set of permissible matches for each agent does not change by this operation, so their choices remain the same.

Below, we link stable outcomes of $\mathcal{E}_{-C}$ to MAPE outcomes whose aggregate set of permissible matches is $\overline{Y} \setminus C$.

Proposition 3.8 *Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy and let $C \subseteq D$. Then $\mu \in \mathcal{S}^*(\mathcal{E}_{-C})$ if and only if there is $R \in \Pi_{i\in N} 2^{D^i\setminus C^i}$ such that (R, μ) is a MAPE of $\mathcal{E}_{-C}$ with*

$$\cup_{i\in N} \left((\overline{Y} \setminus C)^i \setminus R^i\right) = \overline{Y} \setminus C.$$

Proof. Let $\mu \in \mathcal{S}^*(\mathcal{E}_{-C})$. For every $i \in N$, let

$$R^i = \{ij \in D^i \setminus C^i \mid j \succ^i \mu(i)\}.$$

By construction of R, it holds that (R, μ) is a para-equilibrium of $\mathcal{E}_{-C}$. We show that $\cup_{i\in N} \left((\overline{Y} \setminus C)^i \setminus R^i\right) = \overline{Y} \setminus C$. By contradiction, suppose $\cup_{i\in N} \left((\overline{Y} \setminus C)^i \setminus R^i\right) \subsetneq \overline{Y} \setminus C$. Then there is $ij \in D \setminus C$ such that $ij \in R^i \cap R^j$. This implies that ij forms a blocking pair in $\mathcal{E}_{-C}$, contradicting that $\mu \in \mathcal{S}^*(\mathcal{E}_{-C})$. Consequently, (R, μ) is a MAPE of $\mathcal{E}_{-C}$ with $\cup_{i\in N} \left((\overline{Y} \setminus C)^i \setminus R^i\right) = \overline{Y} \setminus C$.

Let (R, μ) be a MAPE of $\mathcal{E}_{-C}$ with $\cup_{i\in N} \left((\overline{Y} \setminus C)^i \setminus R^i\right) = \overline{Y} \setminus C$. Since (R, μ) is a para-equilibrium of $\mathcal{E}_{-C}$, it holds that, for every $i \in N$, $\mu(i) \in \delta^i_{-C}(R^i)$. Thus, μ satisfies individual rationality. If there is $ij \in D \setminus C$ such that $j \succ^i \mu(i)$ and $i \succ^j \mu(j)$, then $ij \in R^i \cap R^j$, contradicting $\cup_{i\in N} \left((\overline{Y} \setminus C)^i \setminus R^i\right) = \overline{Y} \setminus C$. Thus, μ has no blocking pair in $\mathcal{E}_{-C}$. It follows that $\mu \in \mathcal{S}^*(\mathcal{E}_{-C})$. ■

We illustrate the intuition underlying Proposition 3.8 for the case $C = \emptyset$. Assume that the economy has a stable outcome μ. If an agent i strictly prefers agent j to $\mu(i)$, stability implies that j does not strictly prefer i to $\mu(j)$. Thus, the match ij only needs to be excluded from i's option set, not from j's. Following this logic, the aggregate set of permissible matches remains equal to $\overline{Y}$. Conversely, if a MAPE has an aggregate set of permissible matches $\overline{Y}$, then no match is excluded by both agents. Hence no blocking pair can exist: if two agents strictly prefer each other to their current partners, para-equilibrium requires their match to be excluded by both agents, a contradiction. Since a MAPE outcome is individually rational, stability follows.

We now establish the equivalence between semi-stable outcomes and MAPE outcomes.

Theorem 3.9 *Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy.*

(i) The set of MAPE outcomes of $\mathcal{E}$ coincides with the set of semi-stable outcomes $\mathcal{S}(\mathcal{E})$.

(ii) Let $\mu \in \mathcal{S}(\mathcal{E})$ be supported by $C \subseteq D$ and let $R \in \Pi_{i \in N} 2^{D^i}$ be such that (R, μ) is a MAPE of $\mathcal{E}$. Then it holds that $C = \overline{Y} \setminus \cup_{i \in N}(\overline{Y}^i \setminus R^i)$.

Proof.

Part (i).

Step 1. A MAPE outcome of $\mathcal{E}$ belongs to $\mathcal{S}(\mathcal{E})$.

Let $(R^*, \mu^*) \in \Pi_{i \in N} 2^{D^i} \times \mathcal{M}$ be a MAPE of $\mathcal{E}$. We define $C = \{ij \in D \mid ij \in R^{*i}$ and $ij \in R^{*j}\}$. Clearly, it holds that $\cup_{i \in N}(\overline{Y}^i \setminus R^{*i}) = \overline{Y} \setminus C$. By Lemma 3.7(ii), $((R^{*i} \setminus C^i)_{i \in N}, \mu^*)$ is a MAPE of $\mathcal{E}_{-C}$ with $\cup_{i \in N}((\overline{Y} \setminus C)^i \setminus (R^{*i} \setminus C^i)) = \overline{Y} \setminus C$. By Proposition 3.8, we have that $\mu^* \in \mathcal{S}^*(\mathcal{E}_{-C})$.

Suppose $\mu^* \notin \mathcal{S}(\mathcal{E})$. Then there is $C' \subsetneq C$ such that $\mathcal{S}^*(\mathcal{E}_{-C'}) \neq \emptyset$. Let $\nu \in \mathcal{S}^*(\mathcal{E}_{-C'})$. By Proposition 3.8, there is $R \in \prod_{i \in N} 2^{D^i \setminus C'^i}$ such that (R, ν) is a MAPE of $\mathcal{E}_{-C'}$ with $\cup_{i \in N} \left((\overline{Y} \setminus C')^i \setminus R^i\right) = \overline{Y} \setminus C'$. By Lemma 3.7(iii), $((R^i \cup C'^i)_{i \in N}, \nu)$ is a para-equilibrium of $\mathcal{E}$. Since $C' \subsetneq C$, it holds that $\overline{Y} \setminus C \subsetneq \overline{Y} \setminus C'$. Thus condition (ii) of Definition 3.2 is violated at (R^*, μ^*), contradicting that (R^*, μ^*) is a MAPE of $\mathcal{E}$. Consequently, $\mu^* \in \mathcal{S}(\mathcal{E})$.

Step 2. An outcome in $\mathcal{S}(\mathcal{E})$ is a MAPE outcome of $\mathcal{E}$.

Let $\mu^* \in \mathcal{S}(\mathcal{E})$. Thus, there is $C \subseteq D$ such that $\mu^* \in \mathcal{S}^*(\mathcal{E}_{-C})$. By Proposition 3.8, there is $R \in \prod_{i \in N} 2^{D^i \setminus C^i}$ such that (R, μ^*) is a MAPE of $\mathcal{E}_{-C}$ with $\cup_{i \in N} \left((\overline{Y} \setminus C)^i \setminus R^i\right) = \overline{Y} \setminus C$. By Lemma 3.7(iii), $((R^i \cup C^i)_{i \in N}, \mu^*)$ is a para-equilibrium of $\mathcal{E}$.

Suppose that $((R^i \cup C^i)_{i \in N}, \mu^*)$ is not a MAPE of $\mathcal{E}$. Then there is a para-equilibrium (Q, ν) of $\mathcal{E}$ such that $\cup_{i \in N}(\overline{Y}^i \setminus Q^i) \supsetneq \cup_{i \in N}(\overline{Y}^i \setminus (R^i \cup C^i))$. Let $C' = \{ij \in D \mid ij \in Q^i$ and $ij \in Q^j\}$. By Lemma 3.7(ii), $((Q^i \setminus C'^i)_{i \in N}, \nu)$ is a MAPE of $\mathcal{E}_{-C'}$ with $\cup_{i \in N}((\overline{Y} \setminus C')^i \setminus (Q^i \setminus C'^i)) = \overline{Y} \setminus C'$. By Proposition 3.8, it holds that $\nu \in \mathcal{S}^*(\mathcal{E}_{-C'})$ with $C' \subsetneq C$. Thus condition (ii) of Definition 2.5 is violated at μ^*, contradicting that $\mu^* \in \mathcal{S}(\mathcal{E})$. Consequently, $((R^i \cup C^i)_{i \in N}, \mu^*)$ is a MAPE of $\mathcal{E}$.

Part (ii). Let $C' = \overline{Y} \setminus \cup_{i \in N}(\overline{Y}^i \setminus R^i)$. By contradiction, suppose that $C \neq C'$. There are two cases. In the first case, there exists $ij \in C \setminus C'$. Since C supports μ to be semi-stable and $ij \in C$, it holds that $j \succ^i \mu(i)$ and $i \succ^j \mu(j)$. On the other hand, since

(R,μ) is a MAPE, by Definition 3.1, $ij \notin C'$ implies $\mu(i) \succsim^i j$ or $\mu(j) \succsim^j i$, leading to a contradiction. In the second case, there exists $ij \in C' \setminus C$. Since (R,μ) is a MAPE and $ij \in C'$, it holds that $j \succ^i \mu(i)$ and $i \succ^j \mu(j)$. On the other hand, since $\mu \in \mathcal{S}(\mathcal{E})$ and $ij \notin C$, it holds that $\mu(i) \succsim^i j$ or $\mu(j) \succsim^j i$, leading to a contradiction. Consequently, $C = C' = \overline{Y} \setminus \cup_{i \in N}(\overline{Y}^i \setminus R^i)$. ∎

To see the intuition behind Theorem 3.9(i), first consider the case where the economy admits a stable outcome. In such a case, semi-stability coincides with stability, so Proposition 3.8 and Lemma 3.7(i) together imply the equivalence result.

Now consider the general case. Let (R^*, μ^*) be a MAPE of $\mathcal{E}$, and let C be the corresponding set of matches excluded by both parties involved. We show that μ^* is semi-stable in $\mathcal{E}$. By Lemma 3.7(ii), μ^* is a MAPE outcome of $\mathcal{E}_{-C}$. By Proposition 3.8, $\mu^* \in \mathcal{S}^*(\mathcal{E}_{-C})$. If μ^* is not semi-stable in $\mathcal{E}$, then removing some smaller set $C' \subsetneq C$ would restore stability, i.e., $\mathcal{S}^*(\mathcal{E}_{-C'}) \neq \emptyset$. By Proposition 3.8, any stable outcome in $\mathcal{S}^*(\mathcal{E}_{-C'})$ is a MAPE outcome of $\mathcal{E}_{-C'}$. Then, by Lemma 3.7(iii), the original economy $\mathcal{E}$ admits a para-equilibrium with aggregate set of permissible matches $\overline{Y} \setminus C'$, which strictly includes $\overline{Y} \setminus C$. This contradicts that μ^* is a MAPE outcome.

Conversely, suppose that $\mu^* \in \mathcal{S}(\mathcal{E})$. We argue that μ^* is a MAPE outcome of $\mathcal{E}$. Let C be the set of excluded matches supporting the semi-stability of μ^*. Then $\mu^* \in \mathcal{S}^*(\mathcal{E}_{-C})$. By Proposition 3.8, μ^* is a MAPE outcome of $\mathcal{E}_{-C}$. Adding the excluded matches C back, Lemma 3.7(iii) yields a para-equilibrium of $\mathcal{E}$ whose outcome is μ^*. This para-equilibrium must be a MAPE of $\mathcal{E}$. Otherwise, Lemma 3.7(ii) and Proposition 3.8 would yield a reduced economy with fewer excluded matches than C that admits a stable outcome, contradicting $\mu^* \in \mathcal{S}(\mathcal{E})$.

Theorem 3.9(ii) shows that the minimal set of blocking pairs whose removal ensures stability coincides with the set of matches excluded by both parties in a MAPE. This result implies that the minimal intervention needed to ensure stability is equivalent to maximizing the availability of options.

Next, we extend the notion of expectational equilibrium, introduced by Herings (2024) in two-sided many-to-one matching models, to roommate problems.

Definition 3.10 An *expectational equilibrium* (EE) of the economy $\mathcal{E} = (N, \overline{Y}, \succsim)$ is $(R,\mu) \in \Pi_{i \in N} 2^{D^i} \times \mathcal{M}$ such that

(i) (R,μ) is a para-equilibrium of $\mathcal{E}$,

(ii) for every $ij \in D$, $R^i \cap R^j = \emptyset$.

The EE concept offers a different perspective on identifying excluded matches. Condition (ii) requires exclusion to be at most one-sided: an agent may be excluded from matching with another agent, while the latter agent remains free to choose the former. In EE, these constraints represent expectations about feasibility. At equilibrium, agents i and j cannot both expect the match $ij \in D$ to be infeasible; this ensures that agents' expectations are correct and not the result of coordination failures. Indeed, consider a match $ij \in D$. If at an EE (R, μ), agent i expects j not to be feasible, i.e., $ij \in R^i$, condition (i) guarantees $\mu(i) \neq j$, and condition (ii) implies $ij \notin R^j$. Agent j is permitted to choose i, but does not do so in equilibrium. Agent i's expectation that matching with j is infeasible is therefore correct.

Despite the differences in motivation, an EE is a MAPE whose aggregate set of permissible matches is $\overline{Y}$, since each match can be excluded by at most one involved party. The following corollary follows from Lemma 3.7, Propositions 2.6 and 3.8, and Theorem 3.9.

Corollary 3.11 *Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy.*

(i) The set of EE outcomes of $\mathcal{E}$ coincides with the set of stable outcomes $\mathcal{S}^*(\mathcal{E})$.

(ii) If $\mathcal{S}^*(\mathcal{E}) \neq \emptyset$, then the sets of EE outcomes of $\mathcal{E}$, MAPE outcomes of $\mathcal{E}$, and semi-stable outcomes $\mathcal{S}(\mathcal{E})$ all coincide.

By Corollary 3.11(i) and Example 2.2, an EE may fail to exist. Thus, MAPE serves as a natural generalization of EE that resolves this non-existence issue. Although the roommate problem is not subsumed by the two-sided matching models studied by Herings (2024), the equivalence between EE outcomes and stable outcomes holds in both settings.

Figure 1 summarizes the equivalence results obtained in this section. Here, SS denotes semi-stability, and S denotes stability. Equivalences in black are always valid, while those in blue are conditional on existence.

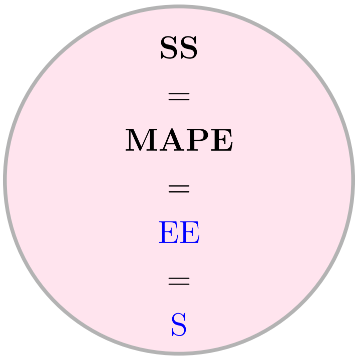

Figure 1: Equivalence between equilibrium outcomes.

4 Efficiency of Semi-stable Outcomes

In this section, we study the efficiency properties of semi-stable outcomes. We begin with the definition of Pareto optimality.

Definition 4.1 An outcome $\mu \in \mathcal{M}$ of the economy $\mathcal{E} = (N, \overline{Y}, \succsim)$ is *Pareto optimal* if there is no outcome $\nu \in \mathcal{M}$ such that

(i) for every $i \in N$, $\nu(i) \succsim^i \mu(i)$,

(ii) for some $j \in N$, $\nu(j) \succ^j \mu(j)$.

The outcome ν in Definition 4.1 is said to *Pareto dominate* the outcome μ.

When agents have preferences with indifferences, stable outcomes may not be Pareto optimal. By Proposition 2.6, a stable outcome is semi-stable. Therefore, a semi-stable outcome may not be Pareto optimal. The economy in the next example has two stable outcomes, where all agents are matched, with one stable outcome Pareto dominating the other. This shows that minimizing the number of excluded pairs or unmatched agents at a semi-stable outcome does not guarantee the Pareto optimality of this outcome.

Example 4.2 Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy with $N = \{1, 2, 3, 4\}$. Agents' preferences are given by

$$\begin{aligned} &2 \sim^1 3 \sim^1 4 \succ^1 1, \\ &1 \sim^2 4 \succ^2 3 \succ^2 2, \\ &1 \succ^3 2 \succ^3 4 \succ^3 3, \\ &1 \sim^4 2 \succ^4 3 \succ^4 4. \end{aligned}$$

Let $\mu, \nu \in \mathcal{M}$ be outcomes with partitions $\mathcal{P}(\mu) = \{12, 34\}$ and $\mathcal{P}(\nu) = \{13, 24\}$, respectively. Both outcomes are stable. Nevertheless, ν Pareto dominates μ. $\triangle$

Example 4.2 raises two questions: whether at least one semi-stable outcome is Pareto optimal, and whether all semi-stable outcomes satisfy some weaker notion of efficiency. We answer both questions affirmatively.

We start with the definition of an undominated semi-stable outcome, which is a semi-stable outcome that is not Pareto dominated by another semi-stable outcome.

Definition 4.3 Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy. A semi-stable outcome $\mu \in \mathcal{S}(\mathcal{E})$ is *undominated* if there does not exist another semi-stable outcome $\nu \in \mathcal{S}(\mathcal{E})$ such that, for every $i \in N$, $\nu(i) \succsim^i \mu(i)$, and there is $j \in N$ such that $\nu(j) \succ^j \mu(j)$.

Since the set of semi-stable outcomes is non-empty and finite, it is straightforward that an undominated semi-stable outcome exists.

Proposition 4.4 *Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy. There is an undominated semi-stable outcome of $\mathcal{E}$.*

Being undominated is less demanding than Pareto optimality, because domination is tested only against other semi-stable outcomes. Nevertheless, we show in Theorem 4.6 that for semi-stable outcomes, being undominated is equivalent to Pareto optimality and individual rationality.

We first prove the following lemma.

Lemma 4.5 *Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy. Let $\mu, \nu \in \mathcal{M}$, $C = \{ij \in D \mid j \succ^i \mu(i)$ and $i \succ^j \mu(j)\}$ and $C' = \{ij \in D \mid j \succ^i \nu(i)$ and $i \succ^j \nu(j)\}$. If ν Pareto dominates μ, then $C' \subseteq C$. Moreover, if there are $j, k \in N$ such that $j = \nu(k)$, $j \succ^k \mu(k)$, and $k \succ^j \mu(j)$, then $C' \subsetneq C$.*

Proof. The Pareto dominance of ν over μ implies that, for every $i \in N$, $\nu(i) \succsim^i \mu(i)$. Thus, for every $ij \in C'$, we have that $j \succ^i \nu(i) \succsim^i \mu(i)$ and $i \succ^j \nu(j) \succsim^j \mu(j)$, so $ij \in C$ and $C' \subseteq C$. Moreover, let $j, k \in N$ be such that $j = \nu(k)$, $j \succ^k \mu(k)$, and $k \succ^j \mu(j)$. Then we have that $jk \in C \setminus C'$, so $C' \subsetneq C$. ■

Lemma 4.5 says that any Pareto improvement of a given outcome yields an outcome with fewer blocking pairs, since agents are weakly better off, so their incentives to block weakly decrease. If such an improvement also strictly benefits two newly paired agents, then at least one blocking pair in the given outcome disappears.

Theorem 4.6 *Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy. A semi-stable outcome of $\mathcal{E}$ is undominated if and only if it is individually rational and Pareto optimal.*

Proof. It is straightforward that a semi-stable outcome that is individually rational and Pareto optimal is undominated. It remains to be shown that an undominated semi-stable outcome is individually rational and Pareto optimal.

Let μ be an undominated semi-stable outcome of $\mathcal{E}$. Individual rationality follows from condition (i) of Definition 2.5.

By contradiction, suppose that μ is Pareto dominated by some outcome $\nu \in \mathcal{M}$ such that, for every $i \in N$, $\nu(i) \succsim^i \mu(i)$, and, for some $j \in N$, $\nu(j) \succ^j \mu(j)$. Let $C = \{ij \in D \mid j \succ^i \mu(i)$ and $i \succ^j \mu(j)\}$ and $C' = \{ij \in D \mid j \succ^i \nu(i)$ and $i \succ^j \nu(j)\}$. By Proposition 2.4, $\mu \in \mathcal{S}^*(\mathcal{E}_{-C})$ and $\nu \in \mathcal{S}^*(\mathcal{E}_{-C'})$. By Lemma 4.5, $C' \subseteq C$. Since $\mu \in \mathcal{S}(\mathcal{E})$, it holds that $C' = C$, so $\nu \in \mathcal{S}(\mathcal{E})$. Since ν Pareto dominates μ, μ is not an undominated semi-stable outcome of $\mathcal{E}$, a contradiction. Thus, no $\nu \in \mathcal{M}$ Pareto dominates μ, so μ is Pareto optimal. ■

The intuition for Theorem 4.6 is as follows. Suppose that an undominated semi-stable outcome μ is Pareto dominated by some outcome ν. Let C and C' denote the sets of blocking pairs at μ and ν, respectively. By Proposition 2.4, μ is stable in the reduced economy obtained by removing C, while ν is stable in the economy obtained by removing C'. Lemma 4.5 gives $C' \subseteq C$. Since μ is semi-stable, it holds that C is minimal, which implies $C = C'$, and hence ν is also semi-stable. Since ν Pareto dominates μ, this contradicts that μ is an undominated semi-stable outcome.

The following result shows that under strict preferences, the standard setting in the roommate problem, every semi-stable outcome is Pareto optimal.

Proposition 4.7 *Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy with strict preferences. Every semi-stable outcome of $\mathcal{E}$ is Pareto optimal.*

Proof. Suppose that $\mu \in \mathcal{S}(\mathcal{E})$ is not Pareto optimal. Then there is $\nu \in \mathcal{M}$ such that, for every $i \in N$, $\nu(i) \succsim^i \mu(i)$, and, for some $j \in N$, $\nu(j) \succ^j \mu(j)$. Let $C = \{ij \in D \mid j \succ^i \mu(i)$ and $i \succ^j \mu(j)\}$ and $C' = \{ij \in D \mid j \succ^i \nu(i)$ and $i \succ^j \nu(j)\}$. By Proposition 2.4, it holds that $\mu \in \mathcal{S}^*(\mathcal{E}_{-C})$ and $\nu \in \mathcal{S}^*(\mathcal{E}_{-C'})$. Let $k = \nu(j)$. Since $k \succ^j \mu(j)$, we have $\mu(j) \neq k$. Moreover, it holds that $j = \nu(k) \succsim^k \mu(k)$ and since preferences are strict, this implies $j \succ^k \mu(k)$. By Lemma 4.5, it holds that $C' \subsetneq C$. This contradicts $\mu \in \mathcal{S}(\mathcal{E})$. Consequently, every $\mu \in \mathcal{S}(\mathcal{E})$ is Pareto optimal. ■

To see why Proposition 4.7 holds, suppose that there is another outcome ν that Pareto dominates μ. By Lemma 4.5, the set of blocking pairs of ν is a proper subset of the blocking pairs of μ. By Proposition 2.4, removing those blocking pairs makes ν stable in the reduced economy. This contradicts the semi-stability of μ.

The following example shows that even under strict preferences, an individually rational and Pareto-optimal outcome may not be semi-stable.

Example 4.8 Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy with $N = \{1, 2, 3, 4\}$. Agents' preferences are given by

$$\begin{aligned} 3 \succ^1 2 \succ^1 4 \succ^1 1, \\ 1 \succ^2 3 \succ^2 4 \succ^2 2, \\ 1 \succ^3 2 \succ^3 4 \succ^3 3, \\ 1 \succ^4 3 \succ^4 2 \succ^4 4. \end{aligned}$$

Let $\mu \in \mathcal{M}$ be such that $\mathcal{P}(\mu) = \{14, 23\}$. Since there is no outcome that makes agents 1, 2, or 3 strictly better off without hurting agent 4, it follows that μ is Pareto optimal. It is clearly individually rational.

Nevertheless, $\mu \notin \mathcal{S}(\mathcal{E})$. Note that the outcome $\nu \in \mathcal{M}$ with $\mathcal{P}(\nu) = \{13, 24\}$ is the unique stable outcome of $\mathcal{E}$, yet $\nu \neq \mu$. △

In the roommate problem with strict preferences, Morrill (2010) proposes algorithms that transform an arbitrary outcome into a Pareto-optimal outcome through Pareto improvements. Nevertheless, Pareto improvements need not transform an arbitrary outcome into a semi-stable outcome. To see this, consider the outcome μ' in Example 4.8 where agents 1 and 4 are matched while the other two agents remain single. Since agents have strict preferences and agent 1 is the best choice for agent 4, there is a unique Pareto improvement over μ' obtained by matching agents 2 and 3. The resulting outcome μ is Pareto optimal, but not semi-stable.

Finally, we define the notion of weak Pareto optimality and show that every semi-stable outcome is weakly Pareto optimal.

Definition 4.9 An outcome $\mu \in \mathcal{M}$ of the economy $\mathcal{E} = (N, \overline{Y}, \succsim)$ is *weakly Pareto optimal* if there is no outcome $\nu \in \mathcal{M}$ such that, for every $i \in N$, $\nu(i) \succ^i \mu(i)$.

The outcome ν in Definition 4.9 is said to *strongly Pareto dominate* the outcome μ. We are now ready to present the following result.

Proposition 4.10 *Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy. Every semi-stable outcome of $\mathcal{E}$ is weakly Pareto optimal.*

Proof. Let $\mu \in \mathcal{S}(\mathcal{E})$. By contradiction, suppose μ is not weakly Pareto optimal. Then there is $\nu \in \mathcal{M}$ such that, for every $i \in N$, $\nu(i) \succ^i \mu(i)$. Let $C = \{ij \in D \mid j \succ^i \mu(i)$ and $i \succ^j \mu(j)\}$ and $C' = \{ij \in D \mid j \succ^i \nu(i)$ and $i \succ^j \nu(j)\}$. By Proposition 2.4, it holds that $\mu \in \mathcal{S}^*(\mathcal{E}_{-C})$ and $\nu \in \mathcal{S}^*(\mathcal{E}_{-C'})$. By Lemma 4.5, we have that $C' \subsetneq C$. Since $\nu \in \mathcal{S}^*(\mathcal{E}_{-C'})$, it cannot be the case that $\mu \in \mathcal{S}(\mathcal{E})$, a contradiction. Consequently, μ is weakly Pareto optimal. ■

The intuition behind Proposition 4.10 is as follows. If a given semi-stable outcome is not weakly Pareto optimal, then there is an outcome that makes every agent strictly better off. If the semi-stable outcome is supported by the set C, then Lemma 4.5 implies that the strongly Pareto dominating outcome is stable in a reduced economy where a proper subset of pairs in C is removed, contradicting the semi-stability of the given outcome.

Figures 2 and 3 summarize the main results presented in this section. PO denotes Pareto optimality and IR denotes individual rationality. USS denotes undominated semi-stability. In Figure 3, it is straightforward that PO+IR is a subset of Weak PO+IR, so this inclusion is omitted.

5 Further Properties of Semi-stable Outcomes

This section studies further properties of semi-stable outcomes. We first relate semi-stable outcomes under preferences with indifferences to those obtained by perturbing the indifferences. We then show that semi-stable outcomes may violate the rural hospital theorem, various versions of lattice properties, and a particular path-connectedness property. All these observations apply to undominated semi-stable outcomes as well.

We now turn to the first issue. Given a preference relation that may exhibit indifferences, a strict preference relation is a *perturbation* of the given preference relation if it preserves all strict preferences and breaks indifferences arbitrarily. Formally, the preference relation $\succsim^{*i}$ is a perturbation of the preference relation $\succsim^i$ of agent $i \in N$ if $\succsim^{*i}$ is strict and, for every $j, k \in N$, $j \succ^i k$ implies $j \succ^{*i} k$. Accordingly, the profile of strict preference relations $\succsim^*$ is a perturbation of the profile of preference relations $\succsim$ if, for every $i \in N$, $\succsim^{*i}$ is a perturbation of $\succsim^i$. Let $\mathcal{R}(\succsim)$ be the collection of profiles of strict preferences corresponding to all possible perturbations of $\succsim$.

Proposition 5.1 *Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy. It holds that*

$$\mathcal{S}(\mathcal{E}) \subseteq \bigcup_{\succsim^* \in \mathcal{R}(\succsim)} \mathcal{S}(N, \overline{Y}, \succsim^*).$$

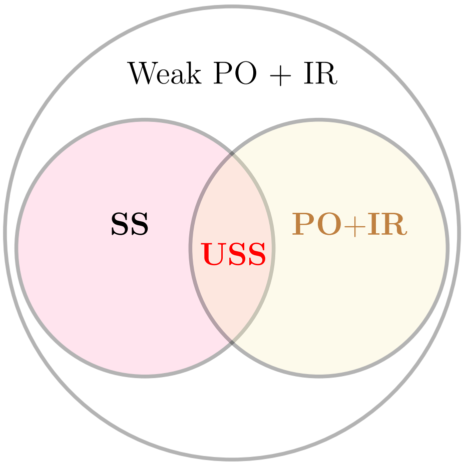

Figure 2: Results when preferences may have indifferences.

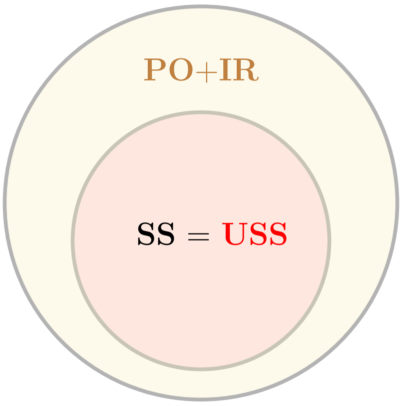

Figure 3: Results with strict preferences.

Proof. Let $\mu \in \mathcal{S}(\mathcal{E})$. We show that there is $\succsim^* \in \mathcal{R}(\succsim)$ such that $\mu \in \mathcal{S}(N, \overline{Y}, \succsim^*)$. Let $C = \{ij \in D \mid j \succ^i \mu(i) \text{ and } i \succ^j \mu(j)\}$. By Proposition 2.4, it holds that $\mu \in \mathcal{S}^*(\mathcal{E}_{-C})$. Let $\succsim^* \in P(\succsim)$ be such that, for every $i \in N$, for every $j \in N \setminus \{\mu(i)\}$ such that $j \sim^i \mu(i)$, $\mu(i) \succ^{*i} j$. It is clear that $\mu \in \mathcal{S}^*(N, \overline{Y} \setminus C, \succsim^*)$.

Suppose $\mu \notin \mathcal{S}(N, \overline{Y}, \succsim^*)$. Then there is $C' \subsetneq C$ and $\nu \in \mathcal{M}$ such that $\nu \in \mathcal{S}^*(N, \overline{Y} \setminus C', \succsim^*)$. We argue that $\nu \in \mathcal{S}^*(N, \overline{Y} \setminus C', \succsim)$. For every $i \in N$, $\nu(i) \succsim^{*i} i$, so either $\nu(i) \succ^{*i} i$ and therefore $\nu(i) \succsim^i i$ or $\nu(i) \sim^{*i} i$ and therefore $\nu(i) = i \sim^i i$, so ν is individually rational in the economy $(N, \overline{Y} \setminus C', \succsim)$.

Suppose there is a blocking pair $ij \in D$ of ν in the economy $(N, \overline{Y} \setminus C', \succsim)$, so $j \succ^i \nu(i)$ and $i \succ^j \nu(j)$. Clearly, we have that $j \succ^{*i} \nu(i)$ and $i \succ^{*j} \nu(j)$, so ij is a blocking pair of ν in the economy $(N, \overline{Y} \setminus C', \succsim^*)$, a contradiction to $\nu \in \mathcal{S}^*(N, \overline{Y} \setminus C', \succsim^*)$. Consequently, there is no blocking pair $ij \in D$ of ν in the economy $(N, \overline{Y} \setminus C', \succsim)$.

We have shown that $\nu \in \mathcal{S}^*(N, \overline{Y} \setminus C', \succsim)$. Since $C' \subsetneq C$, this contradicts $\mu \in \mathcal{S}(\mathcal{E})$. Consequently, it holds that $\mu \in \mathcal{S}(N, \overline{Y}, \succsim^*)$. ■

Proposition 5.1 states that every semi-stable outcome under a preference profile that allows indifferences remains semi-stable for some perturbation. The following example shows that a semi-stable outcome in a perturbed economy may fail to be semi-stable in the original economy.

Example 5.2 Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy with $N = \{1, 2, 3, 4, 5\}$. Agents' preferences are given by

$$\begin{aligned}
&5 \succ^1 2 \succ^1 1 \succ^1 3 \succ^1 4,\\
&1 \succ^2 4 \succ^2 5 \succ^2 2 \succ^2 3,\\
&1 \sim^3 4 \sim^3 5 \sim^3 3 \succ^3 2,\\
&2 \succ^4 3 \succ^4 4 \succ^4 5 \succ^4 1,\\
&3 \succ^5 4 \succ^5 2 \succ^5 1 \succ^5 5.
\end{aligned}$$

Observe that $\mu \in \mathcal{S}^*(\mathcal{E})$ with $\mathcal{P}(\mu) = \{15, 24, 3\}$. Outcome $\nu \in \mathcal{M}$ with $\mathcal{P}(\nu) = \{12, 34, 5\}$ is not stable since 15 forms a blocking pair. It follows that $\nu \notin \mathcal{S}^*(\mathcal{E}) = \mathcal{S}(\mathcal{E})$.

Now consider the perturbation $\succsim^* \in \mathcal{R}(\succsim)$ where $1 \succ^{*3} 4 \succ^{*3} 5 \succ^{*3} 3 \succ^{*3} 2$ and, for every $i \in \{1, 2, 4, 5\}$, $\succsim^{*i} = \succsim^i$. It holds that $\mathcal{S}^*(N, \overline{Y}, \succsim^*) = \emptyset$, whereas $\nu \in \mathcal{S}(N, \overline{Y}, \succsim^*)$, supported by $C = \{15\}$. $\triangle$

If a perturbed economy has a stable outcome, the same outcome is stable in the original economy: restoring indifferences cannot create blocking pairs. Together with Proposition 5.1, we have the following corollary.

Corollary 5.3 *Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy. Then it holds that*

$$\mathcal{S}^*(\mathcal{E}) = \bigcup_{\succsim^* \in \mathcal{R}(\succsim)} \mathcal{S}^*(N, \overline{Y}, \succsim^*).$$

Consequently, an economy has a stable outcome if and only if some perturbed economy has a stable outcome. The same observation for stable outcomes is made by Irving and Manlove (2002, Proposition 1.1) and Chung (2000, Section 5).

In a marriage market with strict preferences, stable outcomes satisfy the rural hospital theorem and the lattice property (see, e.g., Gusfield and Irving (1989) and Roth and Sotomayor (1990)). In roommate problems with strict preferences, stable outcomes, when they exist, still satisfy the rural hospital theorem; see Manlove (2013, Theorem 1.19). However, the lattice property need not hold. Gusfield and Irving (1989, Subsection 4.3.5) instead give two related positive results: one based on median choices across three stable outcomes, and the other based on a semilattice of symmetric differences relative to a fixed stable outcome. We now examine whether these results extend to semi-stable outcomes.

For the rural hospital theorem (see, e.g., Theorem 1.6.3 in Gusfield and Irving (1989) and Section 5.4.4 in Roth and Sotomayor (1990)) to be true in our context, it would require that the set of agents who remain single is the same across all semi-stable outcomes. This theorem applies to stable outcomes in roommate problems with strict preferences, and thus also applies to semi-stable outcomes with strict preferences for which a stable outcome exists. More generally, if multiple semi-stable outcomes are supported by the same set of removed blocking pairs, then they are stable outcomes in the reduced economy and inherit the structural properties of stable outcomes such as the rural hospital theorem.

The rural hospital theorem for semi-stable outcomes does not extend to economies $\mathcal{E}$ for which $\mathcal{S}^*(\mathcal{E}) = \emptyset$. Example 2.7 has no stable outcome, yet it has three semi-stable outcomes, and in each of them a different agent remains single.

A weaker version of the rural hospital theorem would require the number of single agents to be the same across semi-stable outcomes. The next example shows that this weaker requirement can fail even under strict preferences.

Example 5.4 We consider once more the economy $\mathcal{E}$ of Example 3.4. By Theorem 3.9, the outcomes $\mu \in \mathcal{M}$ where $\mathcal{P}(\mu) = \{17, 34, 56, 2\}$ and $\nu \in \mathcal{M}$ such that $\mathcal{P}(\nu) = \{13, 45, 2, 6, 7\}$ are both semi-stable. At μ only agent 2 remains single, but at ν there are three agents, 2, 6, and 7, who remain single. △

Gusfield and Irving (1989, Subsection 4.3.5) show that in roommate problems where agents have strict preferences and stable outcomes exist, the set of stable outcomes may not have a lattice structure. However, they mention two positive results for stable outcomes, when they exist, that are worth examining for semi-stable outcomes. These results are stated for an economy with an even number of agents, strict preferences, and no unacceptable partners.

To state the first result, we need to define the median choice of an agent $i \in N$ given three distinct outcomes $\mu, \mu', \mu'' \in \mathcal{M}$. If i is matched to the same agent in at least two of the three outcomes, that agent is the median choice of i. Otherwise, the median choice of i corresponds to the agent ranked in the middle of these three.

Let $\mathcal{E}$ be an economy with an even number of agents, strict preferences, and no unacceptable partners that has stable outcomes, i.e., $\mathcal{S}^*(\mathcal{E}) \neq \emptyset$. The first result states that, given three distinct outcomes $\mu, \mu', \mu'' \in \mathcal{S}^*(\mathcal{E})$, the function $\nu : N \to N$ where each agent obtains the median choice is an outcome that belongs to $\mathcal{S}^*(\mathcal{E})$. This result applies to semi-stable outcomes as well, since for economies where a stable outcome exists, the set of stable outcomes coincides with the set of semi-stable outcomes. This result generalizes the lattice property of the set of stable outcomes in the marriage problem.

The next example shows that when $\mathcal{S}^*(\mathcal{E}) = \emptyset$, the set of semi-stable outcomes may not contain the agent-wise median of three distinct semi-stable outcomes.

Example 5.5 Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy with $N = \{1, 2, 3, 4\}$. Agents' preferences are given by

$$\begin{aligned}
2 \succ^1 3 \succ^1 4 \succ^1 1,\\
3 \succ^2 1 \succ^2 4 \succ^2 2,\\
1 \succ^3 2 \succ^3 4 \succ^3 3,\\
1 \succ^4 2 \succ^4 3 \succ^4 4.
\end{aligned}$$

In the above economy, $\mathcal{S}^*(\mathcal{E}) = \emptyset$, and $\mathcal{S}(\mathcal{E}) = \{\mu, \mu', \mu''\}$ where $\mathcal{P}(\mu) = \{12, 34\}$, $\mathcal{P}(\mu') = \{13, 24\}$, and $\mathcal{P}(\mu'') = \{14, 23\}$. The outcome μ is supported by $C = \{23\}$, μ' by $C' = \{12\}$, and μ'' by $C'' = \{13\}$.

The function $\nu : N \to N$ where every agent is assigned the median choice among μ, μ', and μ'' is given by $\nu(1) = 3$, $\nu(2) = 1$, $\nu(3) = 2$, and $\nu(4) = 2$. The function ν is not even an outcome. $\triangle$

To state the second result, we need some additional notation. For an outcome $\mu \in \mathcal{M}$, we define the set $U(\mu) = \{(i, j) \in N \times N \mid j \succsim^i \mu(i)\}$, the union of the upper contour sets of the agents at μ. For a fixed outcome $\nu \in \mathcal{M}$, define the symmetric difference between any $\mu \in \mathcal{M}$ and ν as $U_0(\mu) = (U(\mu) \setminus U(\nu)) \cup (U(\nu) \setminus U(\mu))$.

Let $\mathcal{E}$ be an economy with an even number of agents, strict preferences, and no unacceptable partners. The second result states that if $\mathcal{S}^*(\mathcal{E}) \neq \emptyset$, then for every $\nu \in \mathcal{S}^*(\mathcal{E})$, the collection of sets $\{U_0(\mu) \mid \mu \in \mathcal{S}^*(\mathcal{E})\}$ is closed under intersection, i.e., for every $\mu, \mu' \in \mathcal{S}^*(\mathcal{E})$, there is $\mu'' \in \mathcal{S}^*(\mathcal{E})$ such that $U_0(\mu) \cap U_0(\mu') = U_0(\mu'')$. This collection is therefore a meet-semilattice with $U_0(\nu) = \emptyset$ as the minimal element.

As before, this result applies to semi-stable outcomes as well whenever $\mathcal{S}^*(\mathcal{E}) \neq \emptyset$. The next example shows that it may fail when $\mathcal{S}^*(\mathcal{E}) = \emptyset$.

Example 5.6 Consider the economy in Example 5.5. The economy has three semi-stable outcomes, μ, μ', and μ''. It holds that

$$\begin{aligned}
U(\mu) &= \{(1,2),(2,1),(2,3),(3,1),(3,2),(3,4),(4,1),(4,2),(4,3)\},\\
U(\mu') &= \{(1,2),(1,3),(2,1),(2,3),(2,4),(3,1),(4,1),(4,2)\},\\
U(\mu'') &= \{(1,2),(1,3),(1,4),(2,3),(3,1),(3,2),(4,1)\}.
\end{aligned}$$

The symmetric differences with respect to μ are then given by

$$\begin{aligned}
U_0(\mu) &= \emptyset,\\
U_0(\mu') &= \{(1,3),(2,4),(3,2),(3,4),(4,3)\},\\
U_0(\mu'') &= \{(1,3),(1,4),(2,1),(3,4),(4,2),(4,3)\}.
\end{aligned}$$

The intersection of $U_0(\mu')$ and $U_0(\mu'')$ is equal to $\{(1,3),(3,4),(4,3)\}$, which is not equal to any of the symmetric differences with respect to μ. $\triangle$

Finally, we examine the following connectedness property for semi-stable outcomes. Consider an economy $\mathcal{E}$ with strict preferences and $\mu, \mu' \in \mathcal{S}(\mathcal{E})$ supported by C and C', respectively, where $C \neq C'$ and $|C| = |C'|$. Can μ be connected to μ' through a sequence of semi-stable outcomes supported by sets of blocking pairs that replace, one by one, blocking pairs in $C \setminus C'$ with those in $C' \setminus C$? Example 5.7 shows that the answer is no.

Example 5.7 Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy with $N = \{1, 2, 3, 4, 5\}$. Agents have strict preferences given by

$$\begin{aligned} &2 \succ^1 4 \succ^1 5 \succ^1 3 \succ^1 1, \\ &3 \succ^2 5 \succ^2 1 \succ^2 4 \succ^2 2, \\ &4 \succ^3 1 \succ^3 2 \succ^3 5 \succ^3 3, \\ &5 \succ^4 2 \succ^4 3 \succ^4 1 \succ^4 4, \\ &1 \succ^5 3 \succ^5 4 \succ^5 2 \succ^5 5. \end{aligned}$$

We begin with the following observations. First, no outcome in $\mathcal{E}$, $\mathcal{E}_{-\{13\}}$, $\mathcal{E}_{-\{35\}}$ or $\mathcal{E}_{-\{13,35\}}$ is stable. This is because, in any outcome of these economies, at least one agent $i \in N$ remains single. That agent i forms a blocking pair with agent $i-1$, with agent 0 interpreted as agent 5. Second, no outcome in $\mathcal{E}_{-\{23\}}$ or $\mathcal{E}_{-\{13,23\}}$ is stable. The same argument as in the first point applies, except for outcomes of those economies where agent 3 is single. For such outcomes to be stable, all other agents must be paired. However, the outcomes where 12 or 14 forms are blocked by 35, and the outcome where 15 forms is blocked by 12. Thus, none of them is stable. Third, no outcome in $\mathcal{E}_{-\{15\}}$ or $\mathcal{E}_{-\{15,35\}}$ is stable. The same argument as in the first point applies, except for outcomes of those economies where agent 1 is single. For such outcomes to be stable, all other agents must be paired. However, the outcomes where 23 or 24 forms are blocked by 13, and the outcome where 25 forms is blocked by 45. Thus, none of them is stable. Fourth, by the same arguments as above, no outcome in $\mathcal{E}_{-\{15,23\}}$ is stable.

Now consider outcomes $\mu, \mu' \in \mathcal{M}$ with $\mathcal{P}(\mu) = \{23, 45, 1\}$ and $\mathcal{P}(\mu') = \{12, 45, 3\}$. We show $\mu, \mu' \in \mathcal{S}(\mathcal{E})$ where μ is supported by $C = \{13, 15\}$ and μ' by $C' = \{23, 35\}$. Direct verification gives $\mu \in \mathcal{S}^*(\mathcal{E}_{-C})$ and $\mu' \in \mathcal{S}^*(\mathcal{E}_{-C'})$. The above observations show that proper subsets of C and C' give rise to reduced economies without stable outcomes.

To connect μ to μ' through a sequence of semi-stable outcomes supported by sets of blocking pairs that replace, one by one, blocking pairs in $C \setminus C'$ with those in $C' \setminus C$, one needs one of the reduced economies $\mathcal{E}_{-\{13,23\}}$, $\mathcal{E}_{-\{13,35\}}$, $\mathcal{E}_{-\{15,23\}}$, and $\mathcal{E}_{-\{15,35\}}$ to have a stable outcome. The above observations again show that this is not the case. Thus, μ cannot be connected to μ' in the manner described above.

By Proposition 4.7, the conclusion extends to undominated semi-stable outcomes. $\triangle$

6 Alternative Approaches to Address Non-existence

The literature addresses the non-existence of stable outcomes in roommate problems through two approaches. One approach imposes restrictions on preferences to guarantee existence. The other approach adopts alternative solution concepts.

We begin with the first approach. In the roommate problem with strict preferences, Tan (1991) presents a necessary and sufficient condition on preferences that guarantees the existence of a stable outcome. When preferences need not be strict, Chung (2000) proposes the no odd ring condition, which is sufficient but not necessary for existence. This condition states that there is no sequence with an odd number of agents $(i_1, i_2, \ldots, i_k)$, where $k \geq 3$, such that (i) for every odd index j with $1 \leq j \leq k$, $i_{j+1} \succ^{i_j} i_{j-1} \succsim^{i_j} i_j$, and (ii) for every even index j with $1 < j \leq k$, $i_{j+1} \succsim^{i_j} i_{j-1} \succsim^{i_j} i_j$.

In our approach, removing a minimal set of blocking pairs C from $\overline{Y}$ is equivalent to a transformation of agents' preferences such that every agent $i \in N$ ranks all partners in C^i below self-matching. A stable outcome then exists in the transformed economy. Note that a given preference profile can induce different transformed profiles, depending on which minimal set of blocking pairs is removed. When the original preferences are strict, the transformed profile must satisfy Tan's condition for the existence of a stable outcome in the transformed economy. Nevertheless, when the original profile has indifferences, the transformed profile may contain an odd ring, yet the transformed economy still has a stable outcome.[4] Thus, the preference restriction implied by our approach is different from the no odd ring condition given by Chung (2000).

The second approach generally focuses on roommate problems with strict preferences. Tan (1990) proposes a notion of stability called *maximum stability.* An outcome is maximally stable if it matches as many pairs as possible while satisfying internal stability in the sense that the matched agents cannot form blocking pairs. Closely related is the concept of *maximum internal stability* of Biró, Iñarra, and Molis (2016), which looks for outcomes with the largest set of internally stable pairs. Note that maximum stable outcomes are maximum internally stable. There are two further refinements of maximum internally stable outcomes: $P-$stable outcomes, proposed by Iñarra, Larrea, and Molis (2008), and locally stable outcomes, proposed by Vandomme, Crama, and Baratto (2025). Biró et al. (2016) propose *maximum irreversibility.* A set of pairs is irreversible if the agents involved in the pairs cannot form a blocking pair and also cannot find a partner outside the set with whom they would prefer to match. Outcomes with the largest set of irreversible pairs are called maximum irreversible. Next, they propose *Q-stability* as the set of outcomes in the intersection of the maximum internally stable and maximum irreversible outcomes, and show that this set is non-empty.

Example 3 of Biró et al. (2016) shows that the intersection of the set of individually rational and Pareto-optimal outcomes and the set of maximum internally stable outcomes can be empty. Example 4 of Biró et al. (2016) presents an economy where a semi-stable

[4]Consider the economy in Example 3.4, allowing agent 1 to be indifferent between self-matching and matching with agent 7. In this economy, μ specified in Example 3.4 is still a semi-stable outcome supported by $C = \{13, 23\}$. Construct the transformed profile as suggested above, which has $(4, 5, 6)$ as an odd ring. Nevertheless, μ is a stable outcome in the transformed economy.

outcome, namely their almost stable outcome, is not maximum irreversible. Thus, semi-stability is independent of maximum stability, maximum internal stability, $P-$stability, $Q-$stability, local stability, and maximum irreversibility.

Another way to address non-existence takes issue with the myopic nature of deviations and allows for the possibility that one deviation is followed by another. Hirata, Kasuya, and Tomoeda (2021) allow for deviations by coalitions that are not necessarily singletons and pairs. A deviation by a coalition $S \subseteq N$ from $\mu \in \mathcal{M}$ to $\nu \in \mathcal{M}$ requires that, for every $i \in S$, $\nu(i) \succ^i \mu(i)$, $\nu(S) = S$, for every $i \in N \setminus S$ such that $\mu(i) \in S$, $\nu(i) = i$, and, for every $i \in N \setminus S$ such that $\mu(i) \in N \setminus S$, $\nu(i) = \mu(i)$. A deviation by coalition S from $\mu \in \mathcal{M}$ to $\nu \in \mathcal{M}$ is called robust up to depth $k \in \mathbb{N}$ if no agent in S is worse off in any sequence of up to k additional deviations following the deviation to ν. An outcome $\mu \in \mathcal{M}$ is said to be stable against robust deviations up to depth $k \in \mathbb{N}$ if there is no coalition S with a deviation from μ that is robust up to depth k. Hirata et al. (2021) show that there exists an outcome which is stable against robust deviations up to depth 3. We use the following example from their introduction to show that an outcome which is stable against robust deviations up to depth 3 may not be semi-stable while a semi-stable outcome may not be stable against robust deviations up to depth 3.

Example 6.1 Let $\mathcal{E} = (N, \overline{Y}, \succsim)$ be an economy with $N = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Agents have strict preferences with individually rational parts given by

$$\begin{aligned}
&2 \succ^1 9 \succ^1 1,\\
&3 \succ^2 1 \succ^2 2,\\
&4 \succ^3 2 \succ^3 3,\\
&5 \succ^4 3 \succ^4 4,\\
&6 \succ^5 4 \succ^5 5,\\
&7 \succ^6 5 \succ^6 6,\\
&8 \succ^7 6 \succ^7 7,\\
&9 \succ^8 7 \succ^8 8,\\
&1 \succ^9 8 \succ^9 9.
\end{aligned}$$

Hirata et al. (2021) argue that the outcome $\mu \in \mathcal{M}$ with $\mathcal{P}(\mu) = \{12, 34, 56, 78, 9\}$ is not stable against robust deviations up to depth 3. Let $R^1 = R^3 = R^5 = R^7 = \emptyset$, $R^2 = \{23\}$, $R^4 = \{45\}$, $R^6 = \{67\}$, $R^8 = \{89\}$, and $R^9 = \{19, 89\}$. However, μ is a semi-stable outcome supported by $C = \{89\}$. Next consider the outcome $\nu \in \mathcal{M}$ with $\mathcal{P}(\nu) = \{12, 45, 78, 3, 6, 9\}$. Hirata et al. (2021) argue that ν is stable against robust deviations up to depth 2 and therefore also up to depth 3. Nevertheless, ν is not semi-stable. Indeed, any C' supporting ν to be semi-stable must contain $\{23, 56, 89\}$. Since $C \subsetneq C'$, ν cannot be semi-stable. $\triangle$

Atay, Mauleon, and Vannetelbosch (2021) extend the notion of weak stability of Klijn and Massó (2003) as defined for marriage problems to roommate problems. A blocking pair is said to be weak if one of its agents belongs to another blocking pair with a more preferred partner. An outcome is weakly stable if every blocking pair is weak. Atay et al. (2021) show that every roommate problem has an outcome that is both weakly stable and weakly Pareto optimal and that the collection of such outcomes coincides with the bargaining set. They study the economy $\mathcal{E}$ of Example 5.5 in their Example 3 and conclude that the only outcome that is both weakly stable and weakly Pareto optimal is $\nu \in \mathcal{M}$ with $\mathcal{P}(\nu) = \{14, 2, 3\}$. From our analysis of Example 5.5, we find that $\nu \notin \mathcal{S}(\mathcal{E})$. Moreover, ν is not Pareto optimal as it is dominated by the semi-stable outcome $\mu'' \in \mathcal{S}(\mathcal{E})$ with $\mathcal{P}(\mu'') = \{14, 23\}$.

Hirata, Kasuya, and Tomoeda (2023) call a deviation by coalition S from $\mu \in \mathcal{M}$ to $\nu \in \mathcal{M}$ strongly robust if all agents in S are better off in any additional deviation. An outcome $\mu \in \mathcal{M}$ is said to be weakly stable against robust deviations if there is no coalition S with a strongly robust deviation from μ. Hirata et al. (2023) show that there is an outcome in the bargaining set which is weakly stable against robust deviations. Thus, it follows from Example 5.5 that a weakly stable outcome against robust deviations may not be semi-stable. Conversely, the semi-stable outcome μ in Example 5.4 is not weakly stable against robust deviations. This is because the deviation from μ where agents 1 and 3 match is strongly robust, since they are each other's top choice and thus remain matched under any subsequent deviations.

To address non-existence, the literature has also studied set-valued solutions, such as the absorbing set proposed by Iñarra, Larrea, and Molis (2013) and the myopic stable set of Demuynck, Herings, Saulle, and Seel (2019). An absorbing set of outcomes has the property that each outcome can be reached from any other outcome by a sequence of pairwise deviations, whereas pairwise deviations can never lead to an outcome outside the set. The myopic stable set adds the requirement of iterated external stability, i.e., from any outcome not in the set there is a sequence of deviations that ends with an outcome in the set. For roommate problems, the myopic stable set coincides with the union of the absorbing sets. An outcome in the absorbing set may not be semi-stable while a semi-stable outcome may not be in the absorbing set. The former is illustrated by Example 5.5, where the outcomes μ, μ', and μ'' all belong to the unique absorbing set, and therefore to the myopic stable set, but the pair 23 has a deviation from μ to an outcome where 1 and 4 become single, which also belongs to the absorbing set, but is not semi-stable. On the other hand, the outcome μ in Example 5.4 is semi-stable, but cannot be part of an absorbing set, since after a deviation by the pair 13, who are each other's top choices, no sequence of subsequent deviations can return to μ. Another set-valued concept, stochastic stability, is studied by Klaus, Klijn, and Walzl (2010) for roommate problems. They show that the set of stochastically stable outcomes coincides with the set of outcomes in the absorbing sets, so stochastic stability is distinct from semi-stability.

Finally, in roommate problems where indifferences are allowed, Abraham, Biró, and Manlove (2006) suggest almost stability to address the non-existence problem. This concept removes the minimum number of pairs needed for a stable outcome to exist in the reduced roommate problem. The study by Abraham et al. (2006) also applies to roommate problems with strict preferences. It is clear that an almost stable outcome is semi-stable, but the converse need not hold. In Example 3.4, both μ and ν are semi-stable outcomes, where μ is supported by $C = \{13, 23\}$ and ν by $C' = \{56\}$. Only ν, which has fewer blocking pairs, is an almost stable outcome.

Almost stability is independent of undominated semi-stability: the former need not be Pareto optimal, while the latter may not minimize the number of blocking pairs to be removed to restore stability in the reduced economy. Consider the economy obtained by merging those in Examples 3.4 and 4.2. Agents in both examples are treated as distinct and retain their original preferences for agents in their original economy, but rank all agents from the other economy strictly below their self-matches. The merged outcome μ^* consisting of μ in Example 3.4 and ν in Example 4.2 is undominated semi-stable, but not almost stable. This is because the merged outcome μ^{**} consisting of ν in Example 3.4 and μ in Example 4.2 involves a strictly smaller number of blocking pairs. Indeed, μ^{**} is almost stable, but not undominated semi-stable. Recall that in Example 4.2, ν Pareto dominates μ. Consequently, the merged outcome consisting of ν in Example 3.4 and ν in Example 4.2 Pareto dominates μ^{**}. Manlove (2013) considers a variation of almost stability where the number of agents involved in blocking pairs should be as small as possible, a concept that is called min-block-agent stability in Chen, Hermelin, Sorge, and Yedidsion (2018). The same examples as above illustrate the independence between undominated semi-stability and min-block-agent stability.

7 Discussion

Pairwise kidney exchange is a leading application of the roommate problem (see, e.g., Roth et al. (2005) and Sönmez and Ünver (2014)). Each agent, called a patient, comes with a donor. When two patients are paired, they exchange donors. Preferences are often assumed to be *dichotomous*: for every agent i, $N \setminus \{i\}$ is partitioned into at most two sets, N_1^i and N_0^i, such that for every $j, k \in N_1^i$, $j \sim^i k \succ^i i$ and, for every $j, k \in N_0^i$, $i \succ^i j \sim^i k$. Thus, for agent i, any other patient's donor is either compatible or incompatible.

The key solution concept in pairwise kidney exchange is Pareto optimality (see, e.g., Roth et al. (2005) and Sönmez and Ünver (2014)). This model is a special case of ours, but the dichotomous structure of preferences yields stronger results.

Proposition 7.1 *Let $\mathcal{E}$ be an economy with dichotomous preferences.*

(i) The set of stable outcomes of $\mathcal{E}$ is non-empty, i.e., $\mathcal{S}^*(\mathcal{E}) = \mathcal{S}(\mathcal{E}) \neq \emptyset$.

(ii) Any two individually rational and Pareto-optimal outcomes of $\mathcal{E}$ have the same number of agents who remain single and this number is the minimum over all individually rational outcomes in $\mathcal{M}$.

(iii) The set of individually rational and Pareto-optimal outcomes of $\mathcal{E}$ coincides with the set of undominated semi-stable outcomes of $\mathcal{E}$.

Proof. Part (i) follows from Chung (2000) and part (ii) follows from Roth et al. (2005).

For part (iii), by part (i) and Theorem 4.6, we only need to show that an individually rational and Pareto-optimal outcome is stable. Let μ be such an outcome. By contradiction, suppose that μ is not stable. The dichotomous preferences imply that for every $i \in N$ with $\mu(i) \neq i$, agent i has no incentive to deviate. Thus, if μ is not stable, the only possible blocking is formed by two single agents i and j under μ, i.e., $\mu(i) = i$ and $\mu(j) = j$. Pairing i and j while leaving all other agents' assignments unchanged generates an outcome that Pareto dominates μ, contradicting the Pareto optimality of μ. ■

Under dichotomous preferences, stable outcomes exist, and undominated semi-stable outcomes coincide with individually rational and Pareto-optimal outcomes. In particular, Proposition 7.1 implies that any semi-stable outcome that minimizes the number of unmatched agents is undominated, so Pareto optimal. However, under dichotomous preferences, a semi-stable outcome need not be Pareto optimal. For example, consider $N = \{1, 2, 3, 4\}$ with $N_1^1 = \{3, 4\}$, $N_1^2 = \{4\}$, $N_1^3 = \{1\}$, and $N_1^4 = \{1, 2\}$. Matching only agents 1 and 4 is stable, hence semi-stable, but it is Pareto dominated by the outcome that matches 1 with 3 and 2 with 4.

Pairwise kidney exchange incorporates the designer's objectives through priority structures that determine how agents are matched and how outcomes are compared. Priority-based mechanisms are designed to select among individually rational and Pareto-optimal outcomes while achieving desirable properties such as fairness and strategy-proofness (see, e.g., Roth et al. (2005) and Sönmez and Ünver (2014)). By Proposition 7.1(iii), these mechanisms can be viewed as selections from undominated semi-stable outcomes. Our analysis applies to more general preferences, leaving room to accommodate additional institutional constraints and to serve policy goals in a broader range of environments.

Another potential application of the roommate problem is the U.S. Army's "battle buddy system," which pairs soldiers who live, train, and deploy together for mutual assistance and support.[5] Stability is relevant because the system relies on mutual trust and cooperation. Instability may lead to isolation and undermine psychological safety and reliability. Yet instability may be unavoidable, arising either from preference misalignment or from administrative restrictions.[6]

In the battle buddy system, semi-stability can serve as a baseline for addressing the non-existence of stable outcomes, with minimal intervention as the guiding principle. It identifies feasible outcomes from which a final assignment can be selected. Refinements

[5]See the official document *Enlisted Initial Entry Training Policies and Administration* for details: https://adminpubs.tradoc.army.mil/regulations/TR350-6.pdf

[6]For the former, see anecdotal evidence from the soldier forum: https://www.reddit.com/r/army/. For the latter, see the description of the Navy Buddy Enlistment Program: https://www.liveabout.com/navy-buddy-enlistment-program-3332890.

of semi-stable outcomes such as undominated semi-stable outcomes and almost stable outcomes then reflect different objectives, see Section 6. If the designer prioritizes efficiency, e.g., by relaxing some administrative restrictions, undominated semi-stability is appropriate. If the goal is instead to minimize the number of blocking pairs, almost stability is more suitable. Thus, the relevant refinement depends on whether priority is given to efficiency or to stability to the maximum extent.

8 Conclusion

This paper proposes semi-stability, a novel stability notion that addresses the non-existence of stable outcomes in roommate problems. It is based on minimal policy or group interventions to restore stability with favorable efficiency properties. These features set it apart from existing approaches to address non-existence.

We provide an equilibrium foundation for semi-stability by establishing its equivalence to a newly proposed notion of maximal aggregate permission equilibrium. We also establish several efficiency properties: every semi-stable outcome is weakly Pareto optimal; every undominated semi-stable outcome is Pareto optimal; and, under strict preferences, every semi-stable outcome is Pareto optimal. In contrast, semi-stable outcomes generally fail to satisfy the rural hospital theorem and various versions of lattice properties.

Our results suggest potential applications to practical design problems, including pairwise kidney exchange and the battle buddy system. Future research can incorporate richer institutional details and evaluate design improvements based on our theoretical insights.

References

[1] Abraham, David J., Biró, Péter, Manlove, David F., 2006. "Almost stable" matchings in the roommates problem. In Thomas Erlebach and Giuseppe Persiano (Eds.), Proceedings of WAOA 2005: the 3rd International Workshop on Approximation and Online Algorithms, 6-7 October, 2005, Lecture Notes in Computer Science Vol. 3879. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1-14.

[2] Atay, Ata, Mauleon, Ana, Vannetelbosch, Vincent, 2021. A bargaining set for roommate problems. Journal of Mathematical Economics 94, 102465, 1-10.

[3] Biró, Péter, Iñarra, Elena, Molis, Elena, 2016. A new solution concept for the roommate problem: $\mathcal{Q}$-stable matchings. Mathematical Social Sciences 79, 74-82.

[4] Candogan, Ozan, Epitropou, Markos, Vohra, Rakesh V., 2021. Competitive equilibrium and trading networks: A network flow approach. Operations Research 69, 114-147.

[5] Chen, Jiehua, Hermelin, Danny, Sorge, Manuel, Yedidsion, Harel, 2018. How hard is it to satisfy (almost) all roommates? arXiv preprint, arXiv:1707.04316, 1-27.

[6] Chung, Kim-Sau, 2000. On the existence of stable roommate matchings. Games and Economic Behavior 33, 206-230.

[7] Demuynck, Thomas, Herings, P. Jean-Jacques, Saulle, Riccardo D., Seel, Christian, 2019. The myopic stable set for social environments. Econometrica 87, 111-138.

[8] Dur, Umut, Gitmez, A. Arda, Yılmaz, Özgür, 2019. School choice under partial fairness. Theoretical Economics 14, 1309-1346.

[9] Gale, David, Shapley, Lloyd S., 1962. College admissions and the stability of marriage. The American Mathematical Monthly 69, 9-15.

[10] Gusfield, Dan, Irving, Robert W., 1989. The stable marriage problem: structure and algorithms. MIT Press.

[11] Hatfield, John W., Kominers, Scott D., Nichifor, Alexandru, Ostrovsky, Michael, Westkamp, Alexander, 2013. Stability and competitive equilibrium in trading networks. Journal of Political Economy 121, 966-1005.

[12] Herings, P. Jean-Jacques, 2018. Equilibrium and matching under price controls. Journal of Economic Theory 177, 222-244.

[13] Herings, P. Jean-Jacques, 2024. Expectational equilibria in many-to-one matching models with contracts. Journal of Economic Theory 216, 105799, 1-14.

[14] Herings, P. Jean-Jacques, Zhou, Yu, 2025. Harmonious equilibria in roommate problems. CentER Discussion Paper 2025-001, Tilburg University, 1-30.

[15] Hirata, Daisuke, Kasuya, Yusuke, Tomoeda, Kentaro, 2021. Stability against robust deviations in the roommate problem. Games and Economic Behavior 130, 474-498.

[16] Hirata, Daisuke, Kasuya, Yusuke, Tomoeda, Kentaro, 2023. Weak stability against robust deviations and the bargaining set in the roommate problem. Journal of Mathematical Economics 105, 102818, 1-7.

[17] Iñarra, Elena, Larrea, Concepcion, Molis, Elena, 2008. Random paths to P-stability in the roommate problem. International Journal of Game Theory 36, 461-471.

[18] Iñarra, Elena, Larrea, Concepcion, Molis, Elena, 2013. Absorbing sets in roommate problems. Games and Economic Behavior 81, 165-178.

[19] Irving, Robert W., Manlove, David F., 2002. The stable roommates problem with ties. Journal of Algorithms 43, 85-105.

[20] Klaus, Bettina, Klijn, Flip, Walzl, Markus, 2010. Stochastic stability for roommate markets. Journal of Economic Theory 145, 2218-2240.

[21] Klijn, Flip, Massó, Jordi, 2003. Weak stability and a bargaining set for the marriage model. Games and Economic Behavior 42, 91-100.

[22] Kunysz, Adam, 2016. The strongly stable roommates problem. Proceedings of ESA '16: the 24th Annual European Symposium on Algorithms. Vol. 57. Leibniz International Proceedings in Informatics, pp. 60:1-60:15.

[23] Manlove, David, 2013. Algorithmics of matching under preferences. World Scientific.

[24] Morrill, Thayer, 2010. The roommates problem revisited. Journal of Economic Theory 145, 1739-1756.

[25] Quinzii, Martine, 1984. Core and competitive equilibria with indivisibilities. International Journal of Game Theory 13, 41-60.

[26] Richter, Michael, Rubinstein, Ariel, 2024a. Unilateral stability in matching problems. Journal of Economic Theory 216, 105780, 1-9.

[27] Richter, Michael, Rubinstein, Ariel, 2024b. No Prices No Games! Four Economic Models. Open Book Publishers, Cambridge, UK.

[28] Roth, Alvin E., Sotomayor, Marilda, 1990. Two-Sided Matching: A study in game-theoretic modeling and analysis. Cambridge University Press.

[29] Roth, Alvin E., Sönmez, Tayfun, Ünver, M. Utku, 2005. Pairwise kidney exchange. Journal of Economic Theory 125, 151-188.

[30] Shapley, Lloyd S., Shubik, Martin, 1971. The assignment game I: The core. International Journal of Game Theory 1, 111-130.

[31] Sönmez, Tayfun, Ünver, M. Utku, 2014. Altruistically unbalanced kidney exchange. Journal of Economic Theory 152, 105-129.

[32] Tan, Jimmy J.M., 1990. A maximum stable matching for the roommates problem. BIT Numerical Mathematics 30, 631-640.

[33] Tan, Jimmy J.M., 1991. A necessary and sufficient condition for the existence of a complete stable matching. Journal of Algorithms 12, 154-178.

[34] Vandomme, Elise, Crama, Yves, Baratto, Marie, 2025. Locally stable matchings for the roommates problem. Working Paper, 1-26.