UTMD Working Paper
The University of Tokyo
Market Design Center

UTMD-074

Harmonious Equilibria in Roommate Problems

P. Jean-Jacques Herings
Tilburg University

Yu Zhou
Nagoya University

February 4, 2025

Harmonious Equilibria in Roommate Problems[*]

P. Jean-Jacques Herings[†] Yu Zhou[‡]

February 4, 2025

Abstract

We study the roommate problem with self-matching agents and preferences with indifferences. Following the approach of Richter and Rubinstein (2024a, 2024b) and Herings (2024), we define equilibria as constraints shaped by social norms, combined with outcomes where agents make optimal choices given those constraints. We propose four new equilibrium notions, conceptually distinct from taboo equilibrium of Richter and Rubinstein (2024a, 2024b) and expectational equilibrium of Herings (2024). Each equilibrium allows an agent a personalized set of contracts, with additional assumptions on minimality of constraints or maximality of permissions. Motivated by outcome-equivalence, we focus on two out of four of these equilibria, the minimal aggregate constraint equilibrium (MACE) and the maximal aggregate permission equilibrium (MAPE), both guaranteed to exist. MACE and MAPE outcomes are not nested, but with strict preferences, an MAPE is an MACE. We further show that taboo equilibrium outcomes, MACE outcomes, and individually rational Pareto-optimal outcomes coincide, while there is an equivalence between the expectational equilibrium outcomes, stable outcomes, and MAPE outcomes, if the former two exist. Finally, we argue that MAPE is logically independent of all concepts that are proposed in the literature to address the non-existence of stable outcomes in the classical roommate problem.

Keywords: Roommate problem, taboo equilibrium, expectational equilibrium, equilibria with minimal constraints and maximal permissions, Pareto optimality, stability.

JEL Classification: C72, C78, D45, D52.

---

[*]We thank Francis Bloch, Daisuke Hirata, Michihiro Kandori, Michael Richter, and Ariel Rubinstein for their helpful comments. 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 (20KK0027, 21H00696, 22K13364, 22H00062). This paper supersedes the previously circulated version, "Taboo Equilibrium and Expectational Equilibrium in Roommate Problems."

[†]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

[‡]Graduate School of Economics, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601, Japan. E-mail: zhouyu_0105@hotmail.com

1 Introduction

In modern economic theory, there are two dominant methodological approaches. The first one is the competitive framework, where prices play the role of coordinating agents' choices and achieve market clearing. The second approach is the game-theoretic one, where agents interact strategically and their equilibrium behavior is described by a strategy profile such that no agent has an incentive to deviate.

In many practical scenarios, prices are either not involved or do not play an explicit role. Examples include marriage markets, resource sharing in common property regimes, and coalition formation. In these cases, agents often behave in ways aligned with social norms. Richter and Rubinstein (2020) describe the example of birthday pie sharing between family members, where agents' choices are driven not by strategic considerations, but by endogenously emerging social norms that foster harmony.

Several notions of equilibrium have been recently proposed to study outcomes in matching models. They neither involve prices nor follow the spirit of standard game-theoretic concepts. One approach results in the notions of para-taboo equilibrium and taboo equilibrium (TE) as proposed by Richter and Rubinstein (2024a) in the setting of roommate problems.[1] In a para-TE, the social norm determines the set of permissible alternatives from which agents can choose. Each agent chooses an optimal alternative from the permissible set. A TE refines the para-TE by requiring the permissible set to contain as many alternatives as possible.

A different approach is the notion of expectational equilibrium (EE) as proposed by Herings (2024) in two-sided many-to-one matching models with or without monetary transfers. At an EE, agents have endogenously determined expectations about whether it is possible or not to write particular bilateral contracts with other agents. Given their expectations about contracts that can potentially be signed, every agent makes an optimal choice. The roots of this concept go back to market-clearing rationing constraints as formulated in Drèze (1975) in his notion of fix-price equilibrium.

Both approaches have in common that agents' choices are restricted by endogenously emerging constraints, but each approach imposes these constraints in its own distinct way. At a para-TE, the set of permissible contracts is the same across agents and a constrained contract is infeasible for both parties involved. A TE emerges as the most lenient norm such that harmony in society is achieved. At an EE, every agent has a personalized set of permissible contracts. If an agent expects a particular bilateral contract to be impossible, the counterparty expects it to be permissible but is unwilling to sign. Unlike TE, the motivation behind EE is not based on social norms that are as lenient as possible, but

[1] Richter and Rubinstein (2020) study para-TE and TE in general environments, including the cake-splitting problem, quorum economies with preferences over clubs with minimal quorum requirements, and Euclidean economies where alternatives form a closed convex subset of some Euclidean space. The implications of para-TE and TE in voting are discussed in Richter and Rubinstein (2021).

rather on expectations about what is feasible, and a constrained contract is infeasible to exactly one party involved.

In this paper, our motivation is to study the emergence of social norms that are as lenient as possible. However, the most lenient social norms can come in different forms. They may apply symmetrically to both sides in a bilateral interaction, while in other cases they may be imposed on a single party. These differences from the two approaches mentioned above are of critical importance, as they influence not only the fairness and enforceability of norms, but also the behavior of the involved parties across diverse socioeconomic contexts. This paper presents conceptual innovations to bridge those two approaches, offering a more nuanced perspective on how social norms shape outcomes.

We study the roommate problem of Gale and Shapley (1962), where agents are matched in a pairwise fashion. The marriage market is obtained as a special case when making particular choices for the agents' preferences. We extend the original setting by allowing self-matching as well as preferences with indifferences. In our context, a match between two agents is represented as signing a contract, which could be with themselves in the case of self-matching. An outcome is a list of contracts such that every agent is involved in exactly one contract. Alternatively, an outcome can be defined as a partition of the set of agents into singletons and doubletons.

An equilibrium consists of a profile of sets of personalized, constrained contracts, and an outcome. Every agent faces a personalized set of permissible contracts and then makes an optimal choice. Unlike TE and EE, a constraint on a contract can apply to one party or both. We propose four new equilibrium notions, depending on whether social norms are expressed as minimality of constraints or maximality of permissions, and whether minimality and maximality are applied at the individual or the aggregate level. This results in the minimal individual constraint equilibrium (MICE), minimal aggregate constraint equilibrium (MACE), maximal individual permission equilibrium (MIPE), and maximal aggregate permission equilibrium (MAPE).

An MICE requires that every agent faces a minimal set of constrained contracts, so it is not possible to achieve an equilibrium outcome by imposing fewer constrained contracts on every agent, and strictly so for at least one. An MACE requires that the aggregate set of constrained contracts, i.e., the union over all individual sets of constrained contracts, be minimal. An MIPE requires that every agent faces a maximal set of permissible contracts, so it is impossible to achieve an equilibrium outcome by allowing more permissible contracts for every agent, and strictly so for at least one. An MAPE requires that the aggregate set of permissible contracts, i.e., the union over all individual sets of permissible contracts, be maximal. It is immediate that MICE and MIPE are equivalent concepts. We show that, in terms of equilibrium outcomes, MACE and MICE are also equivalent concepts. Therefore, we focus our analysis on MACE and MAPE.

We begin by showing the existence of an MACE and an MAPE. While MACE and

MAPE outcomes are not nested, surprisingly, there is always an outcome which is compatible with both concepts. Moreover, an MAPE outcome is also an MACE outcome when agents have strict preferences.

We then show the underlying connection between MACE, MAPE, TE, and EE, despite their conceptual differences. A TE is an MACE, but the reverse may not hold. Still, the outcomes of TE coincide with those of MACE. By definition, an EE is an MAPE, though the reverse may not be true, as an EE may fail to exist even when all agents have strict preferences, while an MAPE does exist. We show that if an EE exists, it is equivalent to an MAPE. Thus, MAPE naturally generalizes EE, extending it to economies where an EE does not exist. Building on the relation between MACE and MAPE, the connections between TE and MAPE, and EE and MACE, are clear.

We further study the normative properties of stability and efficiency. MACE outcomes coincide with individually rational Pareto-optimal outcomes. If at least one stable outcome exists, then MAPE outcomes coincide with stable outcomes. We conclude that TE outcomes coincide with individually rational Pareto-optimal outcomes, and EE outcomes coincide with stable outcomes. In particular, when agents have strict preferences, the set of EE outcomes is a subset of the set of TE outcomes.

A reformulation of MAPE introduces a weaker, novel, notion of stability, called semi-stability. We show the equivalence between semi-stable outcomes and MAPE outcomes. Semi-stability captures the idea that, after removing a minimal set of contracts from the original economy, the reduced economy possesses a stable outcome. When the original economy has a stable outcome, semi-stable outcomes coincide with stable outcomes. When agents have strict preferences, semi-stable outcomes are a selection of individually rational Pareto-optimal outcomes, and this selection is strict for some economies. Thus, semi-stable outcomes address the non-existence issue of stable outcomes while serving as a proper refinement of individually rational Pareto-optimal outcomes. In particular, we argue that MAPE, or equivalently semi-stable outcomes, are independent of all existing concepts addressing the non-existence of stable outcomes in the roommate problem.

Figures 1 and 2 below summarize our main findings. Equivalences in black are always valid, while those in blue are conditional on existence. Our results shed light on the important role that indifferences play in determining the relationship between equilibria.

This work makes a twofold contribution to the existing literature. First and foremost, we make a conceptual contribution to the emerging research on equilibrium notions that incorporate social norms. We propose new equilibrium notions that accommodate different social norms than those embodied in the recently introduced equilibrium notion of TE. The coincidence between MACE and TE outcomes highlights that different social norms can generate the same outcomes. MAPE, on the other hand, is logically independent of TE and serves as a natural extension of EE. This shows that EE, which is based on expectations of what is feasible, can also be founded on social norms. Moreover, unlike EE, MAPE does not suffer from existence issues in the roommate problem. The relation

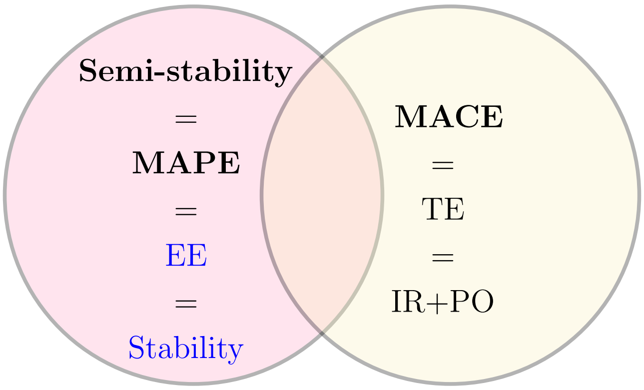

Figure 1: Comparison of equilibrium outcomes under weak preferences.

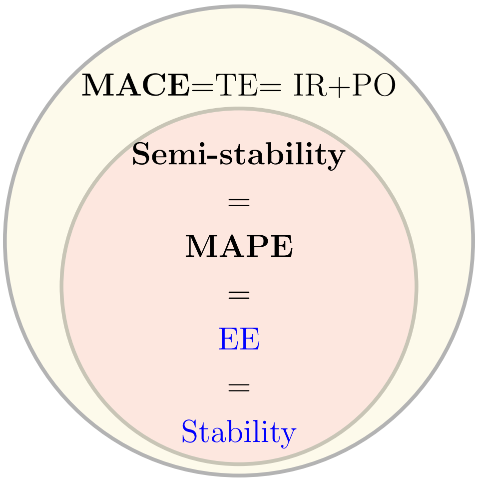

Figure 2: Comparison of equilibrium outcomes under strict preferences.

between MACE and TE, together with the relation between MAPE and EE, helps clarify the connection between TE and EE, which may not be immediately apparent from a direct comparison of their definitions as provided by Richter and Rubinstein (2024a; Chapter 5, 2024b) and Herings (2024). Notably, TE, with outcomes equivalent to those that are individual rational and Pareto-optimal, does not generate new solutions for the roommate problem, while EE does not resolve the non-existence of stable outcomes. On the contrary, MAPE resolves the existence problem and constitutes a proper selection of the set of individually rational and Pareto-optimal outcomes. Our results foster the potential of this research direction by providing new insights into the roommate problem.

Our findings also reinforce the results established by previous works. Richter and Rubinstein (2024a; Chapter 5, 2024b) already study the notions of para-TE and TE in the roommate problem with strict preferences and without self-matching and show the equivalence between TE outcomes and Pareto-optimal outcomes. Our results imply that, when self-matching is allowed, TE outcomes are equivalent to individually rational Pareto-optimal outcomes and that such an equivalence result also holds when agents' preferences admit indifferences. Herings (2024) shows the equivalence between EE and stable outcomes in two-sided many-to-one matching models. These models are different from the roommate problem studied here. We therefore show that the equivalence between EE and stable outcomes extends to roommate problems.

Second, we contribute new solution concepts to address the issue of non-existence of stable outcomes in the roommate problem. This problem was first identified by Gale and Shapley (1962) and has since become a central issue in the study of the roommate problem. Various solution concepts have been proposed to tackle this challenge, often by introducing weaker notions of stability (see, e.g., Tan (1990), Abraham, Biro, and Manlove (2006), Inarra, Larrea and Molis (2008, 2013), Klaus, Klijn, and Walzl (2010), Biro, Inarra, and Molis (2016), Atay, Mauleon, and Vannetelbosch (2021), and Hirata, Kasuya, and Momoeda (2021, 2023)). MAPE, however, is logically independent of all these solution concepts as is discussed in detail in Section 7. Moreover, MAPE is constructed using a different methodological approach, which could be extended beyond the setting of the roommate problem.

The paper has been organized as follows. Section 2 introduces the model. Section 3 defines TE and EE. Section 4 defines MICE, MACE, MIPE, and MAPE, and analyzes their relations. Section 5 discusses the relation between MACE, MAPE, TE, and EE, and their connection to Pareto optimality and stability is studied in Section 6. Section 7 discusses related solution concepts for roommate problems and Section 8 presents the conclusion.

2 The Roommate Problem

We consider a slight generalization of the matching model as proposed by Gale and Shapley (1962), which is also known as "the roommate problem." In Gale and Shapley (1962), agents have to form matching pairs and matching is one-sided, so agents are not a priori partitioned in two disjoint groups. In addition, agents have strict preferences over matching partners and prefer being matched to remaining single. We formulate this model in terms of bilateral contracts, allow agents to have preferences with indifferences, and also incorporate the case where agents may prefer to be single over being matched with some other agent.

Let N be the finite set of agents. The match between agents i and j in N is described as signing a contract between i and j, which we denote by $\{i, j\}$. Let $\overline{Y} = \{\{i, j\} \mid i, j \in N\}$ be the set of all possible contracts between agents. In case $i = j$, agent i remains unmatched and the self-matching contract is given by $\{i, i\} = \{i\}$. The set of contracts in a subset Y of $\overline{Y}$ which involve agent i is denoted by $Y^i = \{y \in Y \mid i \in y\}$.

The consumption set of agent $i \in N$ is equal to $\overline{Y}^i$. The preferences of agent i are represented by a utility function $u^i : \overline{Y}^i \to \mathbb{R}$. Here we allow agent i to be indifferent between different contracts in the consumption set. Preferences of agent i are said to be *strict* if, for every pair $y, y' \in \overline{Y}^i$ with $y \neq y'$, $u^i(y) \neq u^i(y')$. Let $u = (u^i)_{i \in N}$ be the profile of utility functions.

A set of contracts $A \subseteq \overline{Y}$ is an *outcome* if, for every $i \in N$, $|A^i| = 1$. By the requirement $A \subseteq \overline{Y}$, this definition guarantees that choices by agents in an outcome are mutually compatible, i.e., if $\{i,j\} \in A^i$, then also $\{i,j\} \in A^j$. Let $\mathcal{A}$ be the set of all possible outcomes. For every $i \in N$, the utility function u^i over contracts induces the utility function U^i over outcomes. For every $A \in \mathcal{A}$, if $A^i = \{y\}$, then $U^i(A) = u^i(y)$.

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

3 Taboo Equilibrium and Expectational Equilibrium

In the following, let $S = \{\{i\} \mid i \in N\}$ be the set of contracts that correspond to self-matching and let $D = \{\{i,j\} \mid i,j \in N \text{ and } i \neq j\}$ be the set of contracts that correspond to matching pairs. Clearly, we have that S and D are disjoint and $\overline{Y} = S \cup D$. Let $R^i \subseteq D^i$ be a set of contracts which are not available to agent i, i.e., a set of constrained contracts for agent $i \in N$. The set $\overline{Y}^i \setminus R^i$ then corresponds to the set of permissible contracts for agent i. A self-matching contract is never subject to constraints, so it always belongs to the set of permissible contracts for that agent. Let $R = (R^i)_{i \in N} \in \Pi_{i \in N} 2^{D^i}$ be the profile of sets of constrained contracts.[2]

The *budget set* of agent $i \in N$ given a set of constrained contracts $R^i \subseteq D^i$ is equal to

$$\beta^i(R^i) = \overline{Y}^i \setminus R^i.$$

Note that $\beta^i(R^i)$ consists of all contracts involving agent i that are not in R^i. Since $R^i \subseteq D^i$, it holds that $\{i\} \in \beta^i(R^i)$, so $\beta^i(R^i)$ is non-empty. Accordingly, the *demand set* of agent i given a set of constrained contracts R^i is equal to

$$\delta^i(R^i) = \underset{y \in \beta^i(R^i)}{\arg\max}\ u^i(y).$$

In the following, we first present a unified way to define the equilibrium concepts proposed by Richter and Rubinstein (2024a; Chapter 5, 2024b) and Herings (2024). Subsequently, our uniform treatment suggests four new equilibrium concepts, which are developed in the next section. The primary difference between the equilibria defined in this section and those in the next section lies in the different conditions they impose on the sets of constrained or permissible contracts of the agents.

We begin by defining the notion of para-equilibrium. This notion conveys the common idea behind all our notions of equilibrium, which is that given a set of permissible contracts, every agent chooses an optimal one.

Definition 3.1 A *para-equilibrium* of the economy $\mathcal{E} = (\overline{Y}, u)$ is an element $(R, A) \in \Pi_{i \in N} 2^{D^i} \times \mathcal{A}$ such that, for every $i \in N$, $A^i \subseteq \delta^i(R^i)$.

[2]The power set of a set X is denoted by 2^X.

In a para-equilibrium (R, A), R^i corresponds to the set of constrained contracts for agent $i \in N$. The outcome that results in equilibrium is equal to A. Since A is an outcome, it holds that, for every agent $i \in N$, A^i contains exactly one contract. In a para-equilibrium that contract must be in agent $i's$ demand set.

Richter and Rubinstein (2024a; Chapter 5, 2024b) propose para-taboo equilibrium and taboo equilibrium for the case with strict preferences and no self-matching. Para-taboo equilibrium imposes the constraints resulting from some common social norm. A taboo equilibrium is defined as a para-taboo equilibrium with a minimal profile of sets of constrained contracts.

We begin by defining a para-taboo equilibrium.

Definition 3.2 A *para-taboo equilibrium* of the economy $\mathcal{E} = (\overline{Y}, u)$ is an element $(R, A) \in \Pi_{i \in N} 2^{D^i} \times \mathcal{A}$ such that

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

(ii) for every $\{i, j\} \in D$, $\{i, j\} \in R^i$ if and only if $\{i, j\} \in R^j$.

Para-taboo equilibrium refines para-equilibrium and imposes that constraints apply mutually. Either a contract is available to both contractual parties or to none. As an example, consider $R = (D^i)_{i \in N}$, which implies that self-matching is the only available choice for the agents. It is easy to see that (R, S) is a para-taboo equilibrium.

The profile of sets of constrained contracts R induces the aggregate set of constrained contracts $\cup_{i \in N} R^i$ and the aggregate set of permissible contracts $Y = \overline{Y} \setminus \cup_{i \in N} R^i$. It is therefore also possible to define a para-taboo equilibrium as a pair $(Y, A) \in 2^{\overline{Y}} \times \mathcal{A}$ consisting of an aggregate set of permissible contracts and an outcome. This corresponds to the approach followed in Richter and Rubinstein (2024a).

The definition of a taboo equilibrium is as follows.

Definition 3.3 A *taboo equilibrium* (TE) of the economy $\mathcal{E} = (\overline{Y}, u)$ is an element $(R, A) \in \Pi_{i \in N} 2^{D^i} \times \mathcal{A}$ such that

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

(ii) there is no para-taboo equilibrium (Q, B) such that $\cup_{i \in N} Q^i \subsetneq \cup_{i \in N} R^i$.

As an equivalent alternative to condition (ii), we could also write more explicitly that, for every $i \in N$, $Q^i \subseteq R^i$, and there is $j \in N$ such that $Q^j \subsetneq R^j$. A TE is therefore a para-TE with a minimal profile of sets of constrained contracts. In other words, a TE corresponds to the refinement of para-TE that requires a maximal aggregate set of permissible contracts.

The following example illustrates the concept of TE.

Example 3.4 Take $N = \{1,2,3\}$, so $\overline{Y} = \{\{1\},\{2\},\{3\},\{1,2\},\{1,3\},\{2,3\}\}$. Agents' utility functions are given by:

$$\begin{aligned} u^1(\{1,3\}) &> u^1(\{1,2\}) > & u^1(\{1\}), \\ u^2(\{1,2\}) &> \quad u^2(\{2\}) & > u^2(\{2,3\}), \\ u^3(\{1,3\}) &> \quad u^3(\{3\}) & > u^3(\{2,3\}). \end{aligned}$$

Let (R, A) be such that $R^1 = R^3 = \{\{1,3\}\}$, $R^2 = \emptyset$, and $A = \{\{1,2\},\{3\}\}$. We show that (R, A) is a TE. We begin by showing that (R, A) is a para-TE. For agent 1, it holds that $\beta^1(R^1) = \{\{1\},\{1,2\}\}$. Since $u^1(\{1,2\}) > u^1(\{1\})$, we have $A^1 = \{\{1,2\}\} \subseteq \delta^1(R^1)$. Analogously, we can show that $A^2 = \{\{1,2\}\} \subseteq \delta^2(R^2)$ and $A^3 = \{\{3\}\} \subseteq \delta^3(R^3)$. Thus, condition (i) of Definition 3.2 holds. Moreover, for every $\{i,j\} \in D$, it holds that $\{i,j\} \in R^i$ if and only if $\{i,j\} \in R^j$, so condition (ii) of Definition 3.2 holds as well. We have shown that (R, A) is a para-TE.

We show next that condition (ii) of Definition 3.3 holds for (R, A). We proceed by contradiction. Let (Q, B) be a para-TE such that $\cup_{i\in N} Q^i \subsetneq \cup_{i\in N} R^i$. Since $\cup_{i\in N} R^i = \{\{1,3\}\}$, it holds that $\cup_{i\in N} Q^i = \emptyset$. Then agents 2 and 3 both strictly prefer signing a contract with agent 1 to signing any other contract. Thus $\{1,2\} \in B$ and $\{1,3\} \in B$, contradicting that B is an outcome. Consequently, condition (ii) of Definition 3.3 holds.

In conclusion, (R, A) is a TE. △

Herings (2024) introduces a new equilibrium notion for two-sided many-to-one matching models, called expectational equilibrium. It formulates expectations regarding tradable contracts and, in the current context, is defined as follows.

Definition 3.5 An *expectational equilibrium* (EE) of the economy $\mathcal{E} = (\overline{Y}, u)$ is an element $(R, A) \in \Pi_{i\in N} 2^{D^i} \times \mathcal{A}$ such that

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

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

The EE conveys a different idea to derive the constraints on contracts. The second condition of Definition 3.5 requires that in equilibrium constraints are one-sided. It cannot be the case that agents i and j both experience a constraint on contract $\{i,j\}$. Moreover, this condition ensures that the expectations of agents are correct and not the result of coordination failures. Consider a contract $\{i,j\} \in D$. If, at an EE (R, A), agent i expects such a contract not to be available, i.e., $\{i,j\} \in R^i$, then it follows that $\{i,j\} \notin R^j$. Agent j can therefore freely choose to sign the contract $\{i,j\}$, but, since $\{i,j\} \in R^i$ implies $\{i,j\} \notin A$, agent j chooses not to do so. Therefore, agent i is correct in expecting that contract $\{i,j\}$ is not part of agent i's budget set.

A remarkable difference between TE and EE exists in the treatment of constraints. For a TE, if a contract $\{i,j\}$ is not available to agent i, then it is also not available to agent j. This reflects a common social norm that emerges endogenously in a society. As explained above, this is not true for an EE. Here a constraint indicates that a particular contract is out of reach for an agent, since the counterparty of the contract is free to sign it, but chooses not to do so.

The following example illustrates the concept of EE.

Example 3.6 Take $N = \{1, 2, 3\}$, so $\overline{Y} = \{\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}\}$. Agents' utility functions are given by:

$$\begin{aligned} u^1(\{1,3\}) &= u^1(\{1,2\}) > \ u^1(\{1\}), \\ u^2(\{1,2\}) &= \ \ u^2(\{2\}) \ \ > u^2(\{2,3\}), \\ u^3(\{1,3\}) &> \ \ u^3(\{3\}) \ \ > u^3(\{2,3\}). \end{aligned}$$

Let (R, A) be such that $R^1 = \emptyset$, $R^2 = \emptyset$, $R^3 = \{\{1, 3\}\}$, and $A = \{\{1, 2\}, \{3\}\}$. We show that (R, A) is an EE. By the construction of R, it follows that for every $\{i, j\} \in D$, $R^i \cap R^j = \emptyset$, so condition (ii) of Definition 3.5 holds. For agents 1 and 2, it is straightforward that $A^1 \subseteq \delta^1(R^1)$ and $A^2 \subseteq \delta^2(R^2)$. For agent 3, it holds that $\beta^3(R^3) = \{\{3\}, \{2, 3\}\}$. Since $u^3(\{3\}) > u^3(\{2, 3\})$, it holds that $A^3 = \{\{3\}\} = \delta^3(R^3)$. We have shown that (R, A) is an EE. △

4 Equilibria with Minimal Constraints or Maximal Permissions

We now propose four new notions of equilibrium, where each agent has a personalized set of permissible contracts. These equilibria aim to achieve societal harmony by imposing minimal constraints or maximal permissions, either at the individual or the aggregate level. They are conceptually distinct from TE, which targets only a common set of permissible contracts, and from EE, which also rules out common constraints but imposes no minimality assumptions on the set of constrained contracts. We formally examine the relationship between TE, EE, and our proposed equilibria in the next section. We refer to all those equilibria as harmonious equilibria.

We begin by defining a notion of equilibrium that imposes minimal individual sets of constrained contracts.

Definition 4.1 A *minimal individual constraint equilibrium* (MICE) of the economy $\mathcal{E} = (\overline{Y}, u)$ is an element $(R, A) \in \Pi_{i \in N} 2^{D^i} \times \mathcal{A}$ such that

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

(ii) there is no para-equilibrium (Q, B) of $\mathcal{E}$ such that, for every $i \in N$, $Q^i \subseteq R^i$ and there is $j \in N$ with $Q^j \subsetneq R^j$.

The second condition of Definition 4.1 states that an MICE imposes minimal constraints on contracts for each individual agent.

The same idea that requires constraints to be minimal can also be applied at the aggregate level, which results in our next concept.

Definition 4.2 A *minimal aggregate constraint equilibrium* (MACE) of the economy $\mathcal{E} = (\overline{Y}, u)$ is an element $(R, A) \in \Pi_{i\in N} 2^{D^i} \times \mathcal{A}$ such that

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

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

The second condition of Definition 4.2 states that at an MACE, the aggregate set of constrained contracts $\cup_{i\in N} R^i$ is minimal. This is a feature that MACE has in common with TE. Differently from TE, MACE allows for constraints that are one-sided only.

In Example 3.4, (R, A) is both an MICE and an MACE, but this equivalence does not hold in general. In the same example, it is easily verified that (Q, B) such that $Q^1 = Q^3 = \emptyset$, $Q^2 = \{\{1, 2\}\}$, and $B = \{\{1, 3\}, \{2\}\}$ and (Q', B) such that $Q'^1 = Q'^2 = \{\{1, 2\}\}$, and $Q'^3 = \emptyset$ are both MACEs with the same aggregate set of constrained contracts $\{\{1, 2\}\}$. Nevertheless, (Q', B) fails to be an MICE, since (Q, B) is a para-equilibrium with fewer constrained contracts at the individual level. The next result demonstrates that the reverse implication does hold, i.e., an MICE is an MACE.

Proposition 4.3 *Consider an economy* $\mathcal{E} = (\overline{Y}, u)$*. If* $(R^*, A^*) \in \prod_{i\in N} 2^{D^i} \times \mathcal{A}$ *is an MICE of* $\mathcal{E}$*, then* (R^*, A^*) *is an MACE of* $\mathcal{E}$*.*

Proof. Let $(R^*, A^*) \in \prod_{i\in N} 2^{D^i} \times \mathcal{A}$ be an MICE of $\mathcal{E}$. Suppose (R^*, A^*) is not an MACE of $\mathcal{E}$. Then there is a para-equilibrium (Q, B) of $\mathcal{E}$ such that $\cup_{i\in N} Q^i \subsetneq \cup_{i\in N} R^{*i}$. A contract in A^* is not part of R^*, and therefore also not of Q. It follows that, for every $i \in N$, $U^i(B) \geq U^i(A^*)$. For every $i \in N$, let $Q^{*i} = Q^i \cap R^{*i}$. Since for every $i \in N$, the contracts in $Q^i \setminus Q^{*i}$ do not belong to R^{*i}, it holds that the utility of such a contract for agent i is at most $U^i(B)$. Thus, for every $i \in N$, $B^i \subseteq \delta^i(Q^{*i})$, so (B, Q^*) is a para-equilibrium of $\mathcal{E}$. Clearly, for every $i \in N$, $Q^{*i} \subseteq R^{*i}$. By $\cup_{i\in N} Q^i \subsetneq \cup_{i\in N} R^{*i}$, there is $j \in N$ such that $Q^{*j} \subsetneq R^{*j}$. Thus (R^*, A^*) violates condition (ii) of Definition 4.1, contradicting that (R^*, A^*) is an MICE. Consequently, (R^*, A^*) is an MACE of $\mathcal{E}$. ■

The main idea behind the proof of Proposition 4.3 is as follows. Suppose an MICE fails to be an MACE. Then there is a para-equilibrium with a smaller aggregate set of constrained contracts. By taking intersections with the sets of constrained contracts in the MICE, this para-equilibrium can be used to construct a para-equilibrium with weakly smaller sets of constrained contracts for each agent and strictly so for at least one agent than in the MICE, which would lead to a contradiction.

Although we presented an example of an MACE that is not an MICE, the next result shows that the two concepts are equivalent when we focus only on equilibrium outcomes.

Theorem 4.4 *Consider an economy* $\mathcal{E} = (\overline{Y}, u)$*. The set of MACE outcomes of* $\mathcal{E}$ *coincides with the set of MICE outcomes of* $\mathcal{E}$*.*

Proof. In the light of Proposition 4.3, we only have to show that every MACE outcome of $\mathcal{E}$ is also an MICE outcome of $\mathcal{E}$. Let (R, A) be an MACE of $\mathcal{E}$. For every $i \in N$, let

$$R^{*i} = R^i \backslash \{\{i, j\} \in D^i \mid u^i(\{i, j\}) \leq U^i(A^i)\}.$$

Clearly, (R^*, A) is a para-equilibrium such that for every $i \in N$, $R^* \subseteq R^i$. Thus, it follows that $\cup_{i \in N} R^{*i} \subseteq \cup_{i \in N} R^i$. Since (R, A) is an MACE, it holds that (R^*, A) is an MACE.

We show that (R^*, A) is an MICE. By contradiction, suppose not. Then there is a para-equilibrium (Q, B) of $\mathcal{E}$ such that, for every $i \in N$, $Q^i \subseteq R^{*i}$, and there is $j \in N$ such that $Q^j \subsetneq R^{*j}$. Since (R^*, A) is an MACE, it must hold that $\cup_{i \in N} Q^i = \cup_{i \in N} R^{*i}$. By definition of R^{*j}, for every $y \in R^{*j}$, $u^j(y) > U^j(A^j)$. Since $Q^j \subsetneq R^{*j}$, there is $y \in \beta^j(Q^j)$ with $u^j(y) > U^j(A^j)$. It follows that there is $\{j, k\} \in D$ such that $B^j = \{\{j, k\}\}$ and $\{j, k\} \in R^{*j} \backslash Q^j$. By $\cup_{i \in N} Q^i = \cup_{i \in N} R^{*i}$, we have that $\{j, k\} \in Q^k$, so $B^k \neq \{\{j, k\}\}$, leading to a contradiction. Consequently, (R^*, A) is an MICE and A is an MICE outcome. ■

The proof of Theorem 4.4 reveals that if all irrelevant constrained contracts in an MACE are removedthat is, those with utility less than or equal to the utility of the equilibrium outcomean MICE is obtained.

As an alternative to para-equilibria with minimal constraints, one can also consider para-equilibria with maximal permissions. Again, this can be done either at the individual or the aggregate level. We start with the individual level.

Definition 4.5 A *maximal individual permission equilibrium* (MIPE) of the economy $\mathcal{E} = (\overline{Y}, u)$ is an element $(R, A) \in \Pi_{i \in N} 2^{D^i} \times \mathcal{A}$ such that

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

(ii) there is no para-equilibrium (Q, B) of $\mathcal{E}$ such that, for every $i \in N$, $\overline{Y}^i \setminus Q^i \supseteq \overline{Y}^i \setminus R^i$ and there is $j \in N$ with $\overline{Y}^j \setminus Q^j \supsetneq \overline{Y}^j \setminus R^j$.

The second condition of Definition 4.5 is equivalent to the absence of a para-equilibrium (Q, B) of $\mathcal{E}$ such that, for every $i \in N$, $Q^i \subseteq R^i$ and there is $j \in N$ with $Q^j \subsetneq R^j$, which corresponds to the definition of an MICE. We have the following result.

Proposition 4.6 *Consider an economy* $\mathcal{E} = (\overline{Y}, u)$. *Then* $(R, A) \in \Pi_{i \in N} 2^{D^i} \times \mathcal{A}$ *is an MIPE of* $\mathcal{E}$ *if and only if* (R, A) *is an MICE of* $\mathcal{E}$.

Proposition 4.3, Theorem 4.4, and Proposition 4.6 motivate us to focus on MACE, since in terms of outcomes this concept is equivalent to MICE and MIPE.

We next consider maximal permissions at the aggregate level. This leads to the following definition.

Definition 4.7 A *maximal aggregate permission equilibrium* (MAPE) of the economy $\mathcal{E} = (\overline{Y}, u)$ is an element $(R, A) \in \Pi_{i \in N} 2^{D^i} \times \mathcal{A}$ such that

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

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

The second condition of Definition 4.7 states that an MAPE is such that the aggregate set of permissible contracts is maximal. At an EE, the aggregate set of permissible contracts is equal to $\overline{Y}$, as all constraints are one-sided. It therefore follows immediately that an EE is an MAPE. On the other hand, at an MAPE, constraints can be two-sided, making MAPE a weaker concept than EE. We elaborate on the comparison between MAPE and EE in Section 5.

Unlike a TE (R, A), where a contract in the aggregate set of permissible contracts $\cup_{i\in N}(\overline{Y}^i \backslash R^i)$ is permissible for both parties, in an MAPE (R, A), a contract in the aggregate set of permissible contracts $\cup_{i\in N}(\overline{Y}^i \backslash R^i)$ can be permissible for only one party.

MAPE and MACE can be different in terms of outcomes. First, consider (R, A) in Example 3.4. It is an MACE, but not an MAPE. Consider a para-equilibrium (Q, B) such that $Q^1 = Q^3 = \emptyset$, $Q^2 = \{\{1, 2\}\}$, and $B = \{\{1, 3\}, \{2\}\}$. Since $\cup_{i\in N}(\overline{Y}^i \setminus Q^i) = \overline{Y}$ and $\cup_{i\in N}(\overline{Y}^i \setminus R^i) = \overline{Y} \setminus \{\{1, 3\}\}$, it follows that condition (ii) of Definition 4.7 is violated at (R, A), so (R, A) is not an MAPE. It also follows easily that A is not an MAPE outcome.

Next, consider (R, A) in Example 3.6. It qualifies as an MAPE because (R, A) is a para-equilibrium with a maximal aggregate set of permissible contracts $\cup_{i\in N}(\overline{Y}^i \backslash R^i) = \overline{Y}$. Nevertheless, (R, A) is not an MACE. Let (Q, B) be such that $Q^1 = Q^2 = Q^3 = \emptyset$ and $B = \{\{1, 3\}, \{2\}\}$. Then (Q, B) is a para-equilibrium with $\cup_{i\in N} Q^i = \emptyset$. However, we have $\cup_{i\in N} R^i = \{\{1, 3\}\}$. Thus condition (ii) of Definition 4.2 is violated at (R, A), so (R, A) is not an MACE. It also follows that B is the only MACE outcome.

Below, we conduct a thorough analysis of the relation between MAPE and MACE.

Theorem 4.8 *Consider an economy $\mathcal{E} = (\overline{Y}, u)$.*

(i) The economy $\mathcal{E}$ has an MACE and an MAPE.

(ii) There is an outcome that is both an MACE outcome and an MAPE outcome of $\mathcal{E}$.

(iii) If agents have strict preferences, then an MAPE outcome of $\mathcal{E}$ is also an MACE outcome of $\mathcal{E}$.

Proof. We prove the three parts in order.
Part (i): It is clear that $((D^i)_{i\in N}, S)$ is a para-equilibrium of $\mathcal{E}$: The only available contract for an agent is the self-matching contract. The finiteness of the number of contracts implies that there are only finitely many para-equilibria, which ensures the existence of MACE and MAPE.

Before proving Part (ii) and Part (iii), we prove the following claim. A para-equilibrium (R, A) is said to be a para-equilibrium supported by a minimal constraint profile R if there is no $Q \in \prod_{i\in N} 2^{D^i}$ such that $\cup_{i\in N} Q^i \subsetneq \cup_{i\in N} R^i$ and (Q, A) is a para-equilibrium.

Claim: Let (R, A) be a para-equilibrium supported by a minimal constraint profile R. Suppose that A is not an MACE outcome. Then there is a para-equilibrium (Q, B) such that for every $i \in N$, $U^i(B^i) \geq U^i(A^i)$ and there is $j \in N$ such that $U^j(B^j) > U^j(A^j)$.

Since A is not an MACE outcome, there is a para-equilibrium (Q, B) such that $\cup_{j\in N} Q^j \subsetneq \cup_{j\in N} R^j$. For every $i \in N$, let $R'^i = (\cup_{j\in N} R^j)^i$, $R' = (R'^i)_{i\in N}$, $Q'^i = (\cup_{j\in N} Q^j)^i$, and $Q' = (Q^{i\prime})_{i\in N}$. Since (R, A) and (Q, B) are para-equilibria, it holds that (Q', B) and (R', A) are also para-equilibria. Since $\cup_{i\in N} Q^i \subsetneq \cup_{i\in N} R^i$, it holds that, for every $i \in N$, $Q'^i \subseteq R'^i$, so $U^i(B^i) \geq U^i(A^i)$.

Now we show that there is $j \in N$ such that $U^j(B^j) > U^j(A^j)$. By contradiction, suppose that for every $i \in N$, $U^i(B^i) = U^i(A^i)$. Since (Q', B) and (R', A) are both para-equilibria and, for every $i \in N$, $Q'^i \subseteq R'^i$, it holds that (Q', A) is also a para-equilibrium with $\cup_{i\in N} Q'^i = \cup_{i\in N} Q^i \subsetneq \cup_{i\in N} R^i$, contradicting that (R, A) is a para-equilibrium supported by a minimal constraint profile R.

Thus the claim holds.

Now we are ready to prove Part (ii) and Part (iii).

Part (ii): An MAPE outcome $A \in \mathcal{A}$ is said to be a Pareto-optimal outcome in the set of MAPE outcomes if there does not exist an MAPE outcome $B \in \mathcal{A}$ such that, for every $i \in N$, $U^i(B^i) \geq U^i(A^i)$ and, for some $j \in N$, $U^j(B^j) > U^j(A^j)$.

By Part (i), an MAPE of $\mathcal{E}$ exists. The finiteness of the number of contracts implies that there is a finite number of MAPE outcomes. Therefore, there is an MAPE (R^*, A^*) such that A^* is Pareto optimal within the set of MAPE outcomes. Moreover, we can choose (R^*, A^*) such that it is supported by a minimal constraint profile R^*.

We show that A^* is an MACE outcome of $\mathcal{E}$. Clearly, A^* is a para-equilibrium outcome. Suppose that A^* is not an MACE outcome. By the above claim, there is a para-equilibrium (Q, B) such that, for every $i \in N$, $U^i(B^i) \geq U^i(A^{*i})$ and there is $j \in N$ such that $U^j(B^j) > U^j(A^{*j})$. For every $i \in N$, let $Q^{*i} = R^{*i} \backslash B^i$. For every $i \in N$, it follows that $B^i \subseteq \delta^i(Q^{*i})$, so (Q^*, B) is a para-equilibrium with $\cup_{i\in N}(\overline{Y}^i \backslash Q^{*i}) \supseteq \cup_{i\in N}(\overline{Y}^i \backslash R^{*i})$. Since (R^*, A^*) is an MAPE, it holds that $\cup_{i\in N}(\overline{Y}^i \backslash Q^{*i}) = \cup_{i\in N}(\overline{Y}^i \backslash R^{*i})$, so (Q^*, B) is also an MAPE. For every $i \in N$, $U^i(B^i) \geq U^i(A^{*i})$ and there is $j \in N$ such that $U^j(B^j) > U^j(A^{*j})$, so we derive a contradiction to A^* being Pareto optimal within the set of MAPE outcomes. Consequently, A^* is an MACE outcome.

Part (iii): Let A^* be an MAPE outcome. Let (R^*, A^*) be an MAPE which is supported by a minimal constraint profile R^*.

Suppose that A^* is not an MACE outcome. By the above claim, there is a para-equilibrium (Q, B) such that, for every $i \in N$, $U^i(B^i) \geq U^i(A^{*i})$ and there is $j \in N$ such that $U^j(B^j) > U^i(A^{*j})$. Clearly, it holds that $B^j \neq A^{*j}$. Let $k \in N$ be such that $B^j = \{\{j, k\}\}$. It holds that $k \neq j$ and $A^{*k} \neq \{\{j, k\}\}$. Since agent k has strict preferences, we have that $U^k(B^k) > U^k(A^{*k})$. As (R^*, A^*) is an MAPE, it holds that $\{j, k\} \in R^{*j}$ and $\{j, k\} \in R^{*k}$. For every $i \in N$, let $Q^{*i} = R^i \backslash B^i$. It follows that, for every $i \in N$, $B^i \subseteq \delta^i(Q^{*i})$, so (Q^*, B) is a para-equilibrium. Note that $\cup_{i\in N}(\overline{Y}^i \backslash Q^{*i}) \supsetneq (\cup_{i\in N}(\overline{Y}^i \backslash R^{*i}) \cup \{\{j,k\}\}$, which contradicts that (R^*, A^*) is an MAPE. Consequently, A^* is an MACE outcome. ■

Part (i) of Theorem 4.8 shows the existence of both an MACE and an MAPE. If we impose constraints such that every agent is only allowed to choose self-matching contracts, combined with an outcome that contains only self-matching contracts, we obtain a para-equilibrium. The finiteness of the number of contracts ensures the existence of both an MACE and an MAPE. As discussed above Theorem 4.8, the outcomes of MACE and MAPE could differ. Therefore, it is surprising that part (ii) of Theorem 4.8 asserts the existence of an outcome that is compatible with both an MACE and an MAPE. If we further restrict our attention to case with strict preferences, part (iii) shows that every MAPE outcome is also an MACE outcome. Still, the reverse may not hold.[3] Reconsider the discussion above Theorem 4.8. Agents in the economy in Example 3.4 all have strict preferences. In such an economy, there is an MACE outcome which is not an MAPE outcome. Thus, part (iii) indicates that MAPE is a stronger concept than MACE (as well as MICE and MIPE). Thus, parts (ii) and (iii) together show that the strictness of preferences plays an important role in determining the relation between MACE and MAPE.

Figures 3 and 4 below summarize the relations between our proposed equilibria.

5 Relation between Harmonious Equilibria

This section analyzes the relation between TE, EE, MACE, and MAPE.

The foundations for defining these equilibria are para-equilibrium and para-TE. We begin by discussing their relation.

First, we present the standard definition of an individually rational outcome.

Definition 5.1 An outcome $A \in \mathcal{A}$ of the economy $\mathcal{E} = (\overline{Y}, u)$ is *individually rational* if, for every $i \in N$, $U^i(A^i) \geq u^i(\{i\})$.

We now present our first equivalence result.

Proposition 5.2 *Consider an economy $\mathcal{E} = (\overline{Y}, u)$. The sets of individually rational outcomes, para-equilibrium outcomes, and para-TE outcomes of $\mathcal{E}$ all coincide.*

Proof. It is straightforward that both para-equilibrium outcomes and para-TE outcomes are individually rational. Let A^* be an individually rational outcome of $\mathcal{E}$. We show that

[3] We remark that even when agents have strict preferences, an MAPE may not be an MACE. Take $N = \{1,2\}$ so $\overline{Y} = \{\{1\},\{2\},\{1,2\}\}$. Assume that $u^1(\{1\}) > u^1(\{1,2\})$ and $u^2(\{2\}) > u^2(\{1,2\})$. Let $R^1 = \{\{1,2\}\}$, $R^2 = \emptyset$, and $A = \{\{1\},\{2\}\}$. Then (R, A) is an MAPE, but not an MACE. To see this, let $Q^1 = \emptyset$ and $Q^2 = \emptyset$. Then (Q, A) is a para-equilibrium with $\cup_{i\in N} Q^i \subsetneq \cup_{i\in N} R^i$.

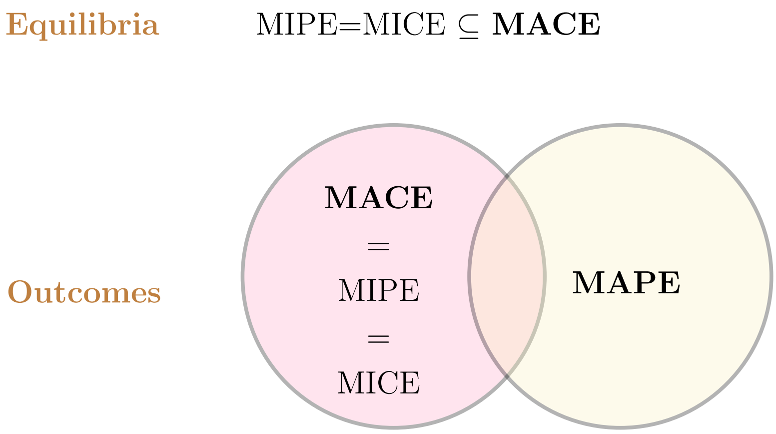

Figure 3: Relation between equilibria with weak preferences.

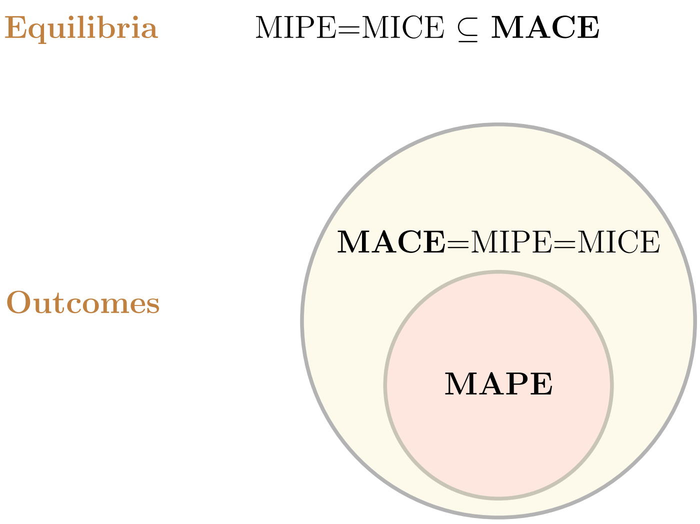

Figure 4: Relation between equilibria with strict preferences.

A^* is both a para-equilibrium outcome and a para-TE outcome of $\mathcal{E}$. For every $i \in N$, let $R^{*i} = D^i \backslash A^{*i}$. Clearly, Definition 3.1 holds for (R^*, A^*), so (R^*, A^*) is a para-equilibrium outcome of $\mathcal{E}$.

It follows that Condition (i) of Definition 3.2 of a TE holds for (R^*, A^*).

Let $y = \{i, j\} \in D$. Clearly, $y \notin A^{*i}$ if and only if $y \notin A^{*j}$. By the construction of R^{*i} and R^{*j}, it holds that $y \in R^{*i}$ if and only if $y \in R^{*j}$. Thus condition (ii) of Definition 3.2 holds as well. We have shown that (R^*, A^*) is a para-TE of $\mathcal{E}$. ■

It is clear that S is an individually rational outcome, so the existence of para-equilibrium and para-TE is ensured. Clearly, a para-TE is a para-equilibrium. Despite the coincidence of their outcomes, a para-equilibrium need not be a para-TE. For instance, consider the economy with strict preferences as described in Example 3.4. Let (Q, B) be such that $Q^1 = Q^3 = \emptyset$, $Q^2 = \{\{1, 2\}\}$, and $B = \{\{1, 3\}, \{2\}\}$. Then (Q, B) is a para-equilibrium, but not a para-TE since only agent 2 is constrained to demand contract $\{1,2\}$.

The following result clarifies the relation between TE and MACE.

Theorem 5.3 *Consider an economy $\mathcal{E} = (\overline{Y}, u)$.*

(i) A TE of $\mathcal{E}$ is an MACE of $\mathcal{E}$.

(ii) The set of TE outcomes of $\mathcal{E}$ coincides with the set of MACE outcomes of $\mathcal{E}$.

Proof.

Part (i): Let (R^*, A^*) be a TE. By contradiction, suppose that (R^*, A^*) is not an MACE. Then there is a para-equilibrium (Q, B) such that $\cup_{i \in N} Q^i \subsetneq \cup_{i \in N} R^{*i}$. For every $i \in N$, let $R^i = (\cup_{j \in N} Q^j)^i$. Clearly, we have that $B \cap (\cup_{j \in N} Q^j) = \emptyset$ and, for every $i \in N$, $R^i \supseteq Q^i$. It follows that, for every $i \in N$, $B^i \subseteq \delta^i(R^i)$. Thus, (R, B) is a para-equilibrium. By the construction of R, $\{i,j\} \in R^i$ if and only if $\{i,j\} \in R^j$, so condition (ii) of Definition 3.2 holds and (R, B) is a para-TE with $\cup_{i \in N} Q^i = \cup_{i \in N} R^i \subsetneq \cup_{i \in N} R^{*i}$. Thus, condition (ii) of Definition 3.3 is violated at (R^*, A^*), contradicting that (R^*, A^*) is a TE. Consequently, (R^*, A^*) is an MACE.

Part (ii): By Part (i), we only need to show that an MACE outcome of $\mathcal{E}$ is a TE outcome of $\mathcal{E}$.

Let (R^*, A^*) be an MACE. For every $i \in N$, let $R^i = (\cup_{j \in N} R^{*j})^i$. For every $\{i,j\} \in D$, it holds that $\{i,j\} \in R^i$ if and only if $\{i,j\} \in R^j$. It is easy to see that (R, A^*) is a para-TE. Moreover, it holds that $\cup_{i \in N} R^i = \cup_{i \in N} R^{*i}$.

Suppose (R, A^*) is not a TE. Then there is a para-TE (Q, B) such that $\cup_{i \in N} Q^i \subsetneq \cup_{i \in N} R^i$. Clearly, (Q, B) is a para-equilibrium. Since $\cup_{i \in N} R^i = \cup_{i \in N} R^{*i}$, it holds that $\cup_{i \in N} Q^i \subsetneq \cup_{i \in N} R^{*i}$. Thus (R^*, A^*) violates condition (ii) of Definition 4.2, contradicting that (R^*, A^*) is an MACE. Consequently, (R, A^*) is a TE. ∎

In terms of outcomes, TE and MACE are equivalent. This shows that it does not make a difference whether constraints are required to be applied on both sides of the contract or not at all, or whether constraints can also be only one-sided. In terms of the equilibria themselves, a TE is an MACE, but the reverse may not be true. For instance, consider the economy with strict preferences described in Example 3.4. Let (Q, B) be such that $Q^1 = Q^3 = \emptyset$, $Q^2 = \{\{1,2\}\}$, and $B = \{\{1,3\}, \{2\}\}$. Clearly (Q, B) is an MACE. However, it is not a TE since only agent 2 is constrained to demand contract $\{1,2\}$.

Replacing MACE with TE in Theorem 4.8, the existence of TE and the relation between MAPE and TE follow.

Corollary 5.4 *Consider an economy $\mathcal{E} = (\overline{Y}, u)$.*

(i) The economy $\mathcal{E}$ has a TE.

(ii) There is an outcome that is both an MAPE outcome and a TE outcome of $\mathcal{E}$.

(iii) If agents have strict preferences, then an MAPE outcome of $\mathcal{E}$ is also a TE outcome of $\mathcal{E}$.

In terms of outcomes, the difference between MACE and MAPE carries over to the difference between TE and MAPE.

Based on the definition, it is clear that an EE is an MAPE. However, the reverse may not hold, since an EE may fail to exist, as illustrated below.[4]

Example 5.5 Take $N = \{1, 2, 3\}$, so $\overline{Y} = \{\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}\}$. Agents' utility functions are given by:

$$u^1(\{1,2\}) > u^1(\{1,3\}) > u^1(\{1\}),$$
$$u^2(\{2,3\}) > u^2(\{1,2\}) > u^2(\{2\}),$$
$$u^3(\{1,3\}) > u^3(\{2,3\}) > u^3(\{3\}).$$

This economy is obtained by removing the fourth agent from Example 3 in Gale and Shapley (1962). In this economy, there is no EE. △

It may therefore be surprising that, when the primitives of the economy are such that an EE does exist, there is a full coincidence between EE and MAPE.

Theorem 5.6 *Consider an economy $\mathcal{E} = (\overline{Y}, u)$.*

(i) An EE of $\mathcal{E}$ is an MAPE of $\mathcal{E}$.

(ii) Assume that an EE of $\mathcal{E}$ exists. Then an MAPE of $\mathcal{E}$ is an EE of $\mathcal{E}$.

Proof. For Part (i), let (R^*, A^*) be an EE of $\mathcal{E}$. Since constraints are one-sided, it holds that $\cup_{i \in N}(\overline{Y}^i \backslash R^{*i}) = \overline{Y}$. Clearly, (R^*, A^*) is an MAPE.

For Part (ii), let (R^*, A^*) be an MAPE of $\mathcal{E}$. By contradiction, suppose that (R^*, A^*) is not an EE. Then there is $\{i, j\} \in D$ such that $\{i, j\} \in R^{*i}$ and $\{i, j\} \in R^{*j}$, so $\cup_{i \in N}(\overline{Y}^i \backslash R^{*i}) \subsetneq \overline{Y}$. By assumption, the economy $\mathcal{E}$ has an EE, say (Q, B). By Part (i), (Q, B) is an MAPE with $\cup_{i \in N}(\overline{Y}^i \backslash Q^i) = \overline{Y}$, contradicting that (R^*, A^*) is an MAPE. ■

Given the existence of an EE, the relationship between EE and MACE is straightforward, as it simply involves replacing MAPE with EE in Theorem 4.8.

We would like to emphasize that, when requiring equilibrium existence and assuming strict preferences, Theorem 4.8, Corollary 5.4, and Theorem 5.6 together suggest that MAPE should be regarded as the strictest concept.

Below, we summarize the relation between TE and EE in a corollary, which follows from Theorem 4.8, Theorem 5.3, and Theorem 5.6. This corollary clarifies the connection between TE and EE as studied in Richter and Rubinstein (2024b) and Herings (2024).

[4]In the roommate problem, if the agents' preferences satisfy the conditions proposed by Tan (1991) or Chung (2000), an EE exists.

Corollary 5.7 *Let $\mathcal{E} = (\overline{Y}, u)$ be an economy.*

(i) Assume that an EE of $\mathcal{E}$ exists. Then there is an outcome that is both an EE and a TE outcome of $\mathcal{E}$.

(ii) If agents have strict preferences, then the set of EE outcomes of $\mathcal{E}$ is a subset of the set of TE outcomes of $\mathcal{E}$.

Finally, we note that the set of outcomes corresponding to any of the equilibrium notions defined in Section 3 and Section 4 may not satisfy the standard "lattice property" or the "rural hospital theorem." For a detailed discussion, see Appendix A.

6 Efficiency and Stability

We now study the normative properties of efficiency and stability for MACE and MAPE.

We first give the definition of a Pareto-optimal outcome.

Definition 6.1 An outcome $A \in \mathcal{A}$ of the economy $\mathcal{E} = (\overline{Y}, u)$ is *Pareto optimal* if there does not exist an outcome $B \in \mathcal{A}$ such that

(i) for every $i \in N$, $U^i(A^i) \geq U^i(B^i)$,

(ii) for some $j \in N$, $U^j(A^j) > U^j(B^j)$.

A stable outcome is defined as follows.

Definition 6.2 An outcome $A \in \mathcal{A}$ of the economy $\mathcal{E} = (\overline{Y}, u)$ is *stable* if

(i) A is individually rational,

(ii) there does not exist $\{i, j\} \in D$ such that $u^i(\{i, j\}) > U^i(A^i)$ and $u^j(\{i, j\}) > U^j(A^j)$.

Condition (ii) of Definition 6.2 says that no two agents can mutually benefit from re-matching. It coincides with the standard "no blocking pair" condition when defining a stable outcome (Gale and Shapley, 1962).

In our setting, there is a finite number of contracts, which implies a finite number of outcomes. Therefore, there exists a Pareto-optimal outcome. However, there may be no stable outcome, e.g., in the economy described in Example 5.5.

The following result shows the relation between MACE outcomes and MAPE outcomes on the one hand and stable outcomes and Pareto-optimal outcomes on the other.

Theorem 6.3 *Let $\mathcal{E} = (\overline{Y}, u)$ be an economy.*

(i) The set of MACE outcomes of $\mathcal{E}$ coincides with the set of individually rational Pareto-optimal outcomes of $\mathcal{E}$.

(ii) Assume that a stable outcome of $\mathcal{E}$ exists. Then the set of MAPE outcomes of $\mathcal{E}$ coincides with the set of stable outcomes of $\mathcal{E}$.

Proof.

Part (i): We first show that an MACE outcome of $\mathcal{E}$ is individually rational and Pareto optimal. Let (R^*, A^*) be an MACE of $\mathcal{E}$. Condition (i) of Definition 4.2 and Proposition 5.2 imply that A^* satisfies individual rationality.

Suppose that A^* is not Pareto optimal. Let $B \in \mathcal{A}$ be such that, for every $i \in N$, $U^i(B^i) \geq U^i(A^{*i})$ and, for some $j \in N$, $U^j(B^j) > U^j(A^{*j})$. For every $i \in N$, let $Q^i = R^{*i} \setminus B^i$. Since $U^j(B^j) > U^j(A^{*j})$, it holds that $\cup_{i\in N} Q^i \subsetneq \cup_{i\in N} R^{*i}$. It is easy to see that, for every $i \in N$, $B^i \subseteq \delta^i(Q^i)$ and (Q, B) is a para-equilibrium. Then $\cup_{i\in N} Q^i \subsetneq \cup_{i\in N} R^{*i}$ implies that (R^*, A^*) violates condition (ii) of Definition 4.2, contradicting that (R^*, A^*) is an MACE. Consequently, A^* is Pareto optimal.

We show next that an individually rational Pareto-optimal outcome of $\mathcal{E}$ is an MACE outcome of $\mathcal{E}$. Let $A^* \in \mathcal{A}$ be an individually rational Pareto-optimal outcome of $\mathcal{E}$. For every $i \in N$, let

$$R^{*i} = \{\{i, j\} \in D^i \mid u^i(\{i, j\}) > U^i(A^{*i}) \text{ or } u^j(\{i, j\}) > U^j(A^{*j})\}.$$

Clearly, (R^*, A^*) is a para-equilibrium of $\mathcal{E}$.

Suppose that (R^*, A^*) is not an MACE. Then there is a para-equilibrium (Q, B) such that $\cup_{i\in N} Q^i \subsetneq \cup_{i\in N} R^{*i}$. By the construction of R^*, for every $\{i, j\} \in D$, $\{i, j\} \in R^{*i}$ if and only if $\{i, j\} \in R^{*j}$. Thus, for every $i \in N$, $Q^i \subseteq R^{*i}$ and there is $j \in N$ such that $Q^j \subsetneq R^{*j}$. Clearly, for every $i \in N$, it holds that $U^i(B^i) \geq U^i(A^{*i})$. Let $k \in N$ be such that $\{j, k\} \in R^{*j} \setminus Q^j$. By the construction of R^{*j} and R^{*k} it holds that $u^j(\{j, k\}) > U^j(A^{*j})$ or $u^k(\{j, k\}) > U^k(A^{*k})$. Moreover, we have that $U^j(B^j) \geq u^j(\{j, k\})$ and $U^k(B^k) \geq u^k(\{j, k\})$, so $U^j(B^j) > U^j(A^{*j})$ or $U^k(B^k) > U^k(A^{*k})$. Recall that, for every $i \in N$, $U^i(B^i) \geq U^i(A^{*i})$. This contradicts that A^* is a Pareto-optimal outcome. Consequently, (R^*, A^*) is an MACE.

Part (ii): Assume that a stable outcome of $\mathcal{E}$ exists.

We show first that a stable outcome of $\mathcal{E}$ is an MAPE outcome of $\mathcal{E}$. Let A^* be a stable outcome. For every $i \in N$, let

$$R^{*i} = \{\{i, j\} \in D^i \mid u^i(\{i, j\}) > U^i(A^{*i})\}.$$

We show that (R^*, A^*) is an MAPE.

By the construction of R^*, it is easy to verify that condition (i) of Definition 4.7 holds. To show that condition (ii) of Definition 4.7 holds, we proceed by contradiction. Suppose that (Q, B) is a para-equilibrium with $\cup_{i\in N}(\overline{Y}^i \backslash Q^i) \supsetneq \cup_{i\in N}(\overline{Y}^i \backslash R^{*i})$. Then there is $\{i, j\} \in D$ such that $\{i, j\} \in (\cup_{i\in N}(\overline{Y}^i \backslash Q^i)) \setminus (\cup_{i\in N}(\overline{Y}^i \backslash R^{*i})$. By the construction of R^{*i} and R^{*j}, it follows that $u^i(\{i, j\}) > U^i(A^{*i})$ and $u^j(\{i, j\}) > U^j(A^{*j})$, contradicting that A^* is a stable outcome. Consequently, (R^*, A^*) is an MAPE.

By the construction of R^* and the fact that A^* is a stable outcome, it also follows that $\{i, j\} \in R^{*i}$ implies $\{i, j\} \notin R^{*j}$. Thus (R^*, A^*) is an MAPE with $\cup_{i\in N}(\overline{Y}^i \setminus R^{*i}) = \overline{Y}$.

We show next that an MAPE outcome of $\mathcal{E}$ is stable. Let (R^*, A^*) be an MAPE of $\mathcal{E}$.

Since a stable outcome of $\mathcal{E}$ exists, the above analysis implies that such a stable outcome, paired with some aggregate set of permissible contracts equal to $\overline{Y}$, becomes an MAPE. Thus, by condition (ii) of Definition 4.7, it follows that $\cup_{i\in N}(\overline{Y}^i \setminus R^{*i}) = \overline{Y}$.

Suppose that A^* is not stable. By condition (i) of Definition 4.7 and Proposition 5.2, for every $i \in N$, $U^i(A^i) \geq u^i(\{i\})$. Thus, there is $\{i,j\} \in D$ such that $u^i(\{i,j\}) > U^i(A^{*i})$ and $u^j(\{i,j\}) > U^j(A^{*j})$. It follows that $\{i,j\} \in R^{*i} \cap R^{*j}$, which contradicts $\cup_{i\in N}(\overline{Y}^i \setminus R^{*i}) = \overline{Y}$. Consequently, A^* is stable. ■

Theorem 6.3 compares our equilibrium concepts in terms of efficiency and stability. An MACE outcome is Pareto optimal but not always stable, while an MAPE outcome is stable when a stable outcome exists, though not necessarily Pareto optimal.

By Theorems 4.8, 5.3, and 6.3, we have the following result.

Corollary 6.4 *Let $\mathcal{E} = (\overline{Y}, u)$ be an economy.*

(i) The set of TE outcomes of $\mathcal{E}$ coincides with the set of individually rational Pareto-optimal outcomes of $\mathcal{E}$.

(ii) If agents have strict preferences, then the set of stable outcomes of $\mathcal{E}$ is a subset of the set of individually rational Pareto-optimal outcomes of $\mathcal{E}$.

Richter and Rubinstein (2024a; Chapter 5, 2024b) show the equivalence between TE outcomes and Pareto-optimal outcomes. They assume that every agent strictly prefers to match with someone else, rather than being self-matched. Under this assumption, every Pareto-optimal outcome is individually rational. Without this assumption, a Pareto-optimal outcome may not be individually rational. Thus, part (i) of Corollary 6.4 is a natural generalization of earlier results in the current context, i.e., both Pareto optimality and individual rationality are necessary to characterize the TE outcomes.[5]

As shown above, EE and stable outcomes may not exist. Theorems 5.6 and 6.3 imply the equivalence between EE outcomes and stable outcomes, conditional on their existence. However, this equivalence holds independently of the existence result.

Proposition 6.5 *Let $\mathcal{E}$ be an economy. The set of EE outcomes of $\mathcal{E}$ coincides with the set of stable outcomes of $\mathcal{E}$.*

Proof. We first show that an EE outcome of $\mathcal{E}$ is stable. Let (R^*, A^*) be an EE. Clearly, condition (i) of Definition 3.5 and Proposition 5.2 imply that condition (i) of Definition 6.2 holds at A^*. For every $\{i,j\} \in D$ such that $u^i(\{i,j\}) > U^i(A^*)$, condition (i) of Definition 3.5 implies $\{i,j\} \in R^{*i}$. By condition (ii) of Definition 3.5, it follows that

[5] If, as an alternative, we define para-TE and TE by taking $R \in \prod_{i\in N} 2^{\overline{Y}^i}$ rather than $R \in \prod_{i \in N} 2^{D^i}$, then it is possible that some agents cannot choose self-matching contracts. Then a para-TE and a TE may not be individually rational, but it is still possible to obtain an equivalence between TE outcomes and Pareto-optimal outcomes.

$\{i,j\} \notin R^{*j}$, so by condition (i) of Definition 3.5, $u^j(\{i,j\}) \leq U^j(A^*)$. Thus condition (ii) of Definition 6.2 holds at A^*. Consequently, A^* is stable.

We show next that a stable outcome of $\mathcal{E}$ is an EE outcome of $\mathcal{E}$. Let A^* be a stable outcome. For every $i \in N$, let

$$R^{*i} = \{\{i,j\} \in D^i \mid u^i(\{i,j\}) > U^i(A^{*i})\}.$$

We show that (R^*, A^*) is an EE. By the construction of R^*, condition (i) of Definition 3.5 clearly holds. If there is $\{i,j\} \in D$ such that $R^{*i} \cap R^{*j} \neq \emptyset$, then $\{i,j\}$ forms a blocking pair, which would violate condition (ii) of Definition 6.2. Thus condition (ii) of Definition 3.5 holds. Consequently, (R^*, A^*) is an EE. ■

Herings (2024) shows the equivalence between EE outcomes and stable outcomes in two-sided matching models with and without transfers. The current setting is not subsumed by two-sided matching models, so Proposition 6.5 demonstrates that such an equivalence relation extends to the roommate problem.

7 Stability in Roommate Problems

The roommate problem has been long studied in the literature. The central issue is resolving the non-existence of stable outcomes. One approach is to impose restrictions on agents preferences to ensure a stable outcome (Tan, 1991; Chung, 2000).

The other approach is to consider alternative solution concepts. The literature has focused on proposing new solutions that weaken the notion of stability (see, e.g., Tan (1990), Abraham et al. (2006), Inarra et al. (2008, 2013), Klaus et al. (2010), Biro et al. (2016), Atay et al. (2021), and Hirata et al. (2021, 2023)). In the following, we study the implications of MAPE as a new stability concept and compare MAPE with the solution concepts proposed in the literature.

If the set of stable outcomes of an economy is non-empty, it coincides with the set of MAPE outcomes. Even if the economy does not have a stable outcome, an MAPE still exists.[6] This suggests using MAPE outcomes as a solution for roommate problems.

In the following, we first propose a weaker notion of stability called semi-stability, and show that semi-stable outcomes are equivalent to MAPE outcomes. Next, we argue that the set of semi-stable outcomes refines the set of individually rational Pareto-optimal outcomes. Finally, we compare semi-stable outcomes with existing solution concepts and conclude that semi-stability is a novel concept distinct from all the others.

Let $\mathcal{E} = (\overline{Y}, u)$ and $C \subseteq D$ be given. We denote the economy $\mathcal{E}_C = (\overline{Y} \backslash C, u')$, where $\overline{Y} \backslash C$ is the set of all possible contracts, and u' is the restriction of u to $\overline{Y} \backslash C$, i.e., for every $i \in N$, for every $y \in \overline{Y}^i \backslash C^i$, $u'^i(y) = u^i(y)$. If $C = \emptyset$, then obviously $\mathcal{E}_\emptyset = \mathcal{E}$.

[6] If an economy has a stable outcome, then every MAPE of such an economy shares the same aggregate set of permissible contracts $\overline{Y}$. Otherwise, different MAPEs may have different aggregate sets of permissible contracts. Appendix B provides an illustrative example.

Definition 7.1 An outcome $A \in \mathcal{A}$ of the economy $\mathcal{E} = (\overline{Y}, u)$ is *semi-stable* if

(i) there is $C \subseteq D$ such that A is a stable outcome of $\mathcal{E}_C$,

(ii) there does not exist $C' \subseteq D$ such that $C' \subsetneq C$ and the economy $\mathcal{E}_{C'}$ has a stable outcome.

The first condition indicates that semi-stability is weaker than stability, as removing some contracts from the original economy results in fewer blocking pairs. Clearly, if A is a stable outcome of $\mathcal{E}$, then $C = \emptyset$. The second condition requires the set of removed contracts to be minimal. Notably, if the original economy has a stable outcome, then an outcome is semi-stable if and only if it is stable.

Now we are ready to present the following equivalence result.

Theorem 7.2 *Let $\mathcal{E} = (\overline{Y}, u)$ be an economy. The set of MAPE outcomes of $\mathcal{E}$ coincides with the set of semi-stable outcomes of $\mathcal{E}$.*

Proof.

Step 1: An MAPE outcome of $\mathcal{E}$ is a semi-stable outcome of $\mathcal{E}$.

Let (R^*, A^*) be an MAPE of $\mathcal{E}$. Let $C = D \backslash \cup_{i \in N} (D^i \backslash R^{*i})$. It is easy to verify that A^* is an EE outcome of $\mathcal{E}_C$. By Proposition 6.5, A^* is a stable outcome of $\mathcal{E}_C$.

Suppose that A^* is not semi-stable of $\mathcal{E}$. Then there is $C' \subseteq D$ such that $C' \subsetneq C$ and $\mathcal{E}_{C'}$ has a stable outcome. Let B be a stable outcome of $\mathcal{E}_{C'}$. By Proposition 6.5, B is an EE outcome of $\mathcal{E}_{C'}$. Thus, there is $R \in \prod_{i \in N} 2^{D^i}$ such that, for every $i \in N$, $R^i \subseteq D^i \backslash C'^i$ and (R, B) is an EE of $\mathcal{E}_{C'}$.

For every $i \in N$, let $R'^i = R^i \cup C'^i$. Clearly, (R', B) is a para-equilibrium of $\mathcal{E}$. It also holds that, for every $i \in N$, $\overline{Y}^i \backslash R'^i = (\overline{Y}^i \backslash C'^i) \backslash R^i$. Since (R, B) is an EE of $\mathcal{E}_{C'}$, it holds that $\cup_{i \in N}((\overline{Y}^i \backslash C'^i) \backslash R^i) = \overline{Y} \backslash C'$. Together with $C = \overline{Y} \backslash \cup_{i \in N} (\overline{Y}^i \backslash R^{*i})$ and $C \supsetneq C'$, we have that $\cup_{i \in N}(\overline{Y}^i \backslash R'^i) = \overline{Y} \backslash C' \supsetneq \overline{Y} \backslash C = \cup_{i \in N}(\overline{Y}^i \backslash R^{*i})$. Thus condition (ii) of Definition 4.7 is violated at (R^*, A^*), contradicting that (R^*, A^*) is an MAPE of $\mathcal{E}$.

Consequently, A^* is semi-stable.

Step 2: A semi-stable outcome of $\mathcal{E}$ is an MAPE outcome of $\mathcal{E}$.

Let A^* be a semi-stable outcome of $\mathcal{E}$. Thus, there is $C \subseteq D$ such that A^* is a stable outcome of $\mathcal{E}_C$ and there does not exist $C' \subseteq D$ such that $C' \subsetneq C$ and $\mathcal{E}_{C'}$ has a stable outcome. By part (ii) of Theorem 6.3, there is $R \in \prod_{i \in N} 2^{D^i}$ such that, for every $i \in N$, $R^i \subseteq D^i \backslash C^i$ and (R, A^*) is an MAPE of $\mathcal{E}_C$.

For every $i \in N$, let $R'^i = R^i \cup C^i$. Clearly, (R', A^*) is a para-equilibrium of $\mathcal{E}$.

Suppose that (R', A^*) is not an MAPE of $\mathcal{E}$. Then there is a para-equilibrium (Q, B) of $\mathcal{E}$ such that $\cup_{i \in N}(\overline{Y}^i \backslash Q^i) \supsetneq \cup_{i \in N}(\overline{Y}^i \backslash R'^i)$. For every $i \in N$, it holds that $\overline{Y}^i \backslash R'^i = \overline{Y}^i \backslash (R^i \cup C^i) = (\overline{Y}^i \backslash C^i) \backslash R^i$. Since (R, A^*) is an MAPE of $\mathcal{E}_C$, it holds that $\cup_{i \in N}((\overline{Y}^i \backslash C^i) \backslash R^i) = \overline{Y} \backslash C$, so $\cup_{i \in N}(\overline{Y}^i \backslash R'^i) = \overline{Y} \backslash C$. Now let $C'' \subseteq D$ be such that $\overline{Y} \backslash C'' = \cup_{i \in N}(\overline{Y}^i \backslash Q^i)$. By $\cup_{i \in N}(\overline{Y}^i \backslash Q^i) \supsetneq \cup_{i \in N}(\overline{Y}^i \backslash R'^i)$, we have that $C'' \subsetneq C$. It is easy to verify that (Q, B) is an EE of $\mathcal{E}_{C''}$. By Proposition 6.5, B is a stable outcome of $\mathcal{E}_{C''}$

with $C'' \subsetneq C$. Thus condition (ii) of Defintion 7.1 is violated at A^*, contradicting that A^* is semi-stable. Consequently, (R', A^*) is an MAPE of $\mathcal{E}$. ■

By parts (i) and (iii) of Theorem 4.8, part (i) of Theorem 6.3, and Thereom 7.2, we obtain the following result.

Corollary 7.3 *Let $\mathcal{E} = (\overline{Y}, u)$ be an economy. If all agents have strict preferences, the set of semi-stable outcomes of $\mathcal{E}$ is non-empty and, moreover, it is a subset of the set of individually rational Pareto-optimal outcomes of $\mathcal{E}$.*

In the canonical roommate problem where all agents have strict preferences, Corollary 7.3 shows that semi-stable outcomes resolve the issue regarding the existence of stable outcomes. In particular, the set of semi-stable outcomes *refines* the set of Pareto-optimal outcomes. In some economies, both with and without stable outcomes, this refinement is proper as illustrated by the following two examples.

First, consider the economy described in Example 3.4. This economy has a stable outcome, $\{\{1,3\},\{2\}\}$, which is both Pareto optimal and semi-stable. In contrast, the outcome $\{\{1,2\},\{3\}\}$ is Pareto optimal, but it is not a semi-stable outcome. Second, consider an economy that merges the one of Example 3.4 with the one in Example 5.5, where the agents in the latter economy are relabeled as 4, 5, and 6. Agents retain the same preferences over contracts with agents from their own economy, but regard contracts with agents from the other economy as strictly inferior to self-matching contracts. It is easy to verify that, in the combined economy, there is no stable outcome, and an individually rational Pareto-optimal outcome may not be semi-stable.

We end this section by providing a thorough comparison which shows that semi-stable outcomes are distinct from all existing concepts addressing the non-existence of stable outcomes in the roommate problem where agents have strict preferences.

Abraham et al. (2006) propose a notion of stability called "almost stability." The key idea is to remove the minimum number of contracts, each representing a blocking pair, so that there is a stable outcome in the reduced economy. Almost stability is stronger than semi-stability and the set of almost stable outcomes is a subset of the set of semi-stable outcomes. In Appendix B, we give an example of an economy with two different semi-stable outcomes. Only the one where a smaller number of contracts is removed from the economy is almost stable.

Tan (1990) proposes a notion of stability called "maximum stability," which is also known as maximum internal stability when agents are allowed to remain unmatched (Biro et al., 2016). Inarra et al. (2008) propose a notion of stability called "$P-$stability." Biro et al. (2016) propose "maximum irreversibility" and "$Q-$stability." Biro et al. (2016) provide an example (Example 3, Page 78) to show that the intersection of the set of individually rational Pareto-optimal outcomes and the set of "maximum internally stable" outcomes can be empty. When agents have strict preferences, by Corollary 7.3, the set of semi-stable outcomes is a subset of the set of individually rational Pareto-optimal outcomes. Moreover, Biro et al. (2016) show that the sets of $Q-$stable outcomes and $P-$stable outcomes are subsets of the set of maximum internally stable outcomes. Thus semi-stability is independent of $Q-$stability, $P-$stability, and maximum internally stability. Notice that a $Q-$stable outcome is also a maximum irreversible outcome and the matching μ mentioned in Example 4 of Biro et al. (Page 78, 2016) yields a semi-stable outcome which is not a maximum irreversible outcome. Thus, semi-stability is independent of "maximum irreversibility."

Inarra et al. (2013) propose the "absorbing set" as the solution concept. An outcome in the absorbing set may not be semi-stable while a semi-stable outcome may not be in the absorbing set.[7] Thus the absorbing set and our solution concept are logically independent. Klaus et al. (2010) propose a notion of stability, called "stochastic stability." They show that the set of stochastically stable outcomes coincide with the set of outcomes in the absorbing sets, so stochastic stability is distinct from semi-stability.

Hirata et al. (2021) propose a notion of stability against robust deviations up to depth k. A deviation is robust up to depth k if no deviators are worse-off after at most k subsequent deviations compared to the status quo. An outcome is stable against robust deviations up to depth k if there is no deviation that is robust up to depth k. They show that there is a stable outcome against robust deviations up to depth 3. However, we use an example from their introduction (replicated in Appendix C) to show that a stable outcome 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.

Atay et al. (2021) show the non-emptiness and coincidence of the "bargaining set" and the set of "weakly stable" and "weakly efficient" outcomes. Taking Example 3 in their paper for instance, there is a unique weakly stable and weakly efficient outcome, i.e., the bargaining set is a singleton. This outcome is not semi-stable. Nevertheless, there is a semi-stable outcome in the same example, which does not belong to the bargaining set. Hirata et al. (2023) propose a notion called "weak stability against robust deviations," which has a non-empty intersection with the bargaining set. Thus, using Example 3 in Atay et al. (2021), the fact that the bargaining set is a singleton implies that a weakly stable outcome against robust deviations may not be semi-stable. One may verify that the semi-stable outcome A in Appendix B is not weakly stable against robust deviations.[8]

[7] In Example 3 given by Inarra et al. (2013), they show that the almost stable outcome μ_8, which is also a semi-stable outcome, is not in the absorbing set. In the same example, one can easily verify that the outcome μ_5 in the absorbing set is not a semi-stable outcome.

[8] Consider the MAPE (R, A) as derived in Appendix B. The deviation formed by agents 1 and 3 is strongly robust because agents 1 and 3 are each other's top choice, and once these two agents are matched, no beneficial rematching with other agents occurs. Note that a weakly stable outcome should not permit any strongly robust deviations.

8 Conclusion

This paper studies the classical roommate problem while allowing for self-matching agents and preferences with indifferences. Building on the equilibrium approach of Richter and Rubinstein (2024a; 2024b) and Herings (2024), we propose four new equilibrium concepts. One of these concepts, MAPE, is used to address the non-existence of stable outcomes in roommate problems.

In the current context, agents can freely sign contracts with one another. Our results naturally extend to scenarios where agents are limited in their contracting options or where contracts take more abstract forms - such as involving the same pair of agents but differing in terms. Although existence of an equilibrium is ensured, an issue for further research is the study of adjustment processes that end up in an equilibrium.

Standard competitive equilibrium concepts, where prices coordinate agent choices, and game-theoretic solutions, such as Nash equilibrium, are well-studied in the matching literature. However, little is known about equilibria based on social norms or expectations, which differ methodologically and conceptually from the standard concepts. This paper, along with Richter and Rubinstein (2024a; 2024b) and Herings (2024), highlights the potential of these equilibrium concepts to provide new insights into existing frameworks.

Appendix

A The Lattice Property and the Rural Hospital Theorem

Take $N = \{1, 2, 3\}$, so $\overline{Y} = \{\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}\}$. Agents' utility functions are given by:

$$\begin{aligned} u^1(\{1,3\}) >& \ \ u^1(\{1\}) &> u^1(\{1,2\}), \\ u^2(\{2,3\}) >& \ \ u^2(\{2\}) &> u^2(\{1,2\}), \\ u^3(\{1,3\}) =& \ u^3(\{2,3\}) &> \ \ u^3(\{3\}). \end{aligned}$$

In such an economy, there are only two stable outcomes: $A = \{\{1, 3\}, \{2\}\}$ and $B = \{\{2, 3\}, \{1\}\}$. Agent 1 prefers A to B, while agent 2 prefers B to A. Consequently, there is no stable outcome that is unanimously preferred by all agents over the other stable outcome. Furthermore, agent 1 is assigned a self-matching contract in B, while agent 2 is assigned a self-matching contract in A. This implies that different stable outcomes assign self-matching contracts to different agents. As a result, the lattice property and the rural hospital theorem, as defined in the standard sense (Roth and Sotomayor, 1990), fail to hold for stable outcomes.

In the above economy, the sets of MACE outcomes, MAPE outcomes, stable outcomes, and individually rational Pareto-optimal outcomes all coincide. Therefore, the observations made above apply to other equilibrium outcomes as well.

B Multiplicity of the Aggregate Set of Permissible Contracts

Take $N = \{1,2,3,4,5,6\}$, so $\overline{Y} = \{\{1\},\{2\},\{3\},\{4\},\{5\},\{6\}\} \cup \{\{i,j\} \mid i,j \in N$ with $i \neq j\}$. Agents have strict preferences over contracts. For simplicity, we omit contracts strictly less preferred than self-matching ones. Agents' utility functions are given by:

$$
\begin{aligned}
&u^1(\{1,3\}) > \ u^1(\{1\}),\\
&u^2(\{2,3\}) > \ u^2(\{2\}),\\
&u^3(\{1,3\}) > u^3(\{2,3\}) \ > \ u^3(\{3,4\}) \ > \ u^3(\{3\}),\\
&u^4(\{3,4\}) > u^4(\{4,5\}) \ > \ u^4(\{4,6\}) \ > \ u^4(\{4\}),\\
&u^5(\{5,6\}) > u^5(\{4,5\}) \ > \ \ u^5(\{5\}),\\
&u^6(\{4,6\}) > u^6(\{5,6\}) \ > \ \ u^6(\{6\}).
\end{aligned}
$$

Let $A = \{\{1\},\{2\},\{3,4\},\{5,6\}\}$, $R^1 = \{\{1,3\}\}$, $R^2 = \{\{2,3\}\}$, $R^3 = \{\{1,3\},\{2,3\}\}$, $R^4 = R^5 = \emptyset$, and $R^6 = \{\{4,6\}\}$. Then (R, A) is an MAPE with $\cup_{i\in N}(\overline{Y}^i \backslash R^i) = \overline{Y}\backslash\{\{1,3\},\{2,3\}\}$. Let $B = \{\{1,3\},\{2\},\{4,5\},\{6\}\}$, $Q^1 = \emptyset$, $Q^2 = \{\{2,3\}\}$, $Q^3 = \emptyset$, $Q^4 = \{\{3,4\}\}$, $Q^5 = \{\{5,6\}\}$, and $Q^6 = \{\{4,6\},\{5,6\}\}$. Then (Q, B) is an MAPE with $\cup_{i\in N}(\overline{Y}^i \backslash Q^i) = \overline{Y}\backslash\{\{5,6\}\}$. Both (R, A) and (Q, B) are MAPEs, but $\cup_{i\in N}(\overline{Y}^i \backslash Q^i)$ and $\cup_{i\in N}(\overline{Y}^i \backslash R^i)$ are different, even in terms of cardinality.

C Stability against Robust Deviations and MAPE

Take $N = \{1,2,3,4,5,6,7,8,9\}$, so $\overline{Y} = \{\{i\} \mid i \in N\} \cup \{\{i,j\} \mid i,j \in N$ with $i \neq j\}$. Agents have strict preferences over contracts. For simplicity, we omit contracts strictly less preferred than self-matching ones. Agents' utility functions are given by:

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

This economy corresponds to the one described in the introduction of Hirata et al. (2021) for the case with nine agents.

Let $A = \{\{1,2\},\{3,4\},\{5,6\},\{7,8\},\{9\}\}$, which corresponds to $\mu_{(9)}$ in Hirata et al. (2021). Let $R^1 = R^3 = R^5 = R^7 = \emptyset$, $R^2 = \{\{2,3\}\}$, $R^4 = \{\{4,5\}\}$, $R^6 = \{\{6,7\}\}$, $R^8 = \{\{8,9\}\}$, and $R^9 = \{\{1,9\},\{8,9\}\}$. One can verify that (R, A) is an MAPE with $\cup_{i\in N}(\overline{Y}^i \backslash R^i) = \overline{Y}\backslash\{\{8,9\}\}$. However, Hirata et al. (2021) argue that $\mu_{(9)}$ is not stable against robust deviations up to depth 3. Let $B = \{\{1,2\},\{3\},\{4,5\},\{6\},\{7,8\},\{9\}\}$, which corresponds to $\mu'_{(9)}$ in Hirata et al. (2021). They argue that $\mu'_{(9)}$ is stable against robust deviations up to depth 2, which implies that $\mu'_{(9)}$ is stable against robust deviations up to depth 3. Nevertheless, B is not an MAPE outcome. To see this, for any Q such that (Q, B) is a para-equilibrium, it holds that $\cup_{i\in N}(\overline{Y}^i \backslash Q^i) \subseteq \overline{Y}\backslash\{\{2,3\},\{5,6\},\{8,9\}\}$. Since (R, A) is a para-equilibrium with $\cup_{i\in N}(\overline{Y}^i \backslash R^i) = \overline{Y}\backslash\{\{8,9\}\}$, we conclude that B is not an MAPE outcome.

References

[1] Abraham, D. J., Biro P., Manlove, D.F., 2006. "Almost stable" matchings in the roommates problem. In R. Fleischer and G. Trippen (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, (pp. 1-14). Berlin, Heidelberg: Springer Berlin Heidelberg.

[2] Atay, A., Mauleon, A., Vannetelbosch, V., 2021. A bargaining set for roommate problems. Journal of Mathematical Economics 94, 102465.

[3] Biro, P., Inarra, E., Molis, E., 2016. A new solution concept for the roommate problem: Q-stable matchings. Mathematical Social Sciences, 79, 74-82.

[4] Chung, K.S., 2000. On the existence of stable roommate matchings. Games and Economic Behavior 33(2), 206-230.

[5] Drèze, J.H., 1975. Existence of an exchange equilibrium under price rigidities. International Economic Review 16, 301-320.

[6] Gale, D., Shapley, L.S., 1962. College admissions and the stability of marriage. The American Mathematical Monthly 69(1), 9-15.

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

[8] Hirata, D., Kasuya, Y., Tomoeda, K., 2021. Stability against robust deviations in the roommate problem. Games and Economic Behavior 130, 474-498.

[9] Hirata, D., Kasuya, Y., Tomoeda, K., 2023. Weak stability against robust deviations and the bargaining set in the roommate problem. Journal of Mathematical Economics 105, 102818, 1-7.

[10] Inarra, E., Larrea, C., Molis, E., 2008. Random paths to P-stability in the roommate problem. International Journal of Game Theory 36, 461-471.

[11] Inarra, E., Larrea, C., Molis, E., 2013. Absorbing sets in roommate problems. Games and Economic Behavior 81, 165-178.

[12] Klaus, B., Klijn, F., Walzl, M., 2010. Stochastic stability for roommate markets. Journal of Economic Theory 145(6), 2218-2240.

[13] Richter, M., Rubinstein, A., 2020. The permissible and the forbidden. Journal of Economic Theory 188, 105042, 1-21.

[14] Richter, M., Rubinstein, A., 2021. Holding a group together: Non-game theory versus game theory. The Economic Journal 131(638), 2629-2641.

[15] Richter, M., Rubinstein, A., 2024a. Unilateral stability in matching problems. Journal of Economic Theory 216, 105780, 1-9.

[16] Richter, M., Rubinstein, A., 2024b. No Prices No Games! Four Economic Models. Open Book Publishers, Cambridge, UK.

[17] Roth, A. E., Sotomayor, M., 1990. Two-Sided Matching: A study in game-theoretic modeling and analysis two-sided matching. Cambridge University Press.

[18] Tan, J.J., 1990. A maximum stable matching for the roommates problem. BIT Numerical Mathematics 30(4), 631-640.

[19] Tan, J.J., 1991. A necessary and sufficient condition for the existence of a complete stable matching. Journal of Algorithms 12(1), 154-178.