UTS UNIVERSITY OF TECHNOLOGY SYDNEY

UTS Business School – Department of Economics
Dr Chau Chak Wing Building
14 – 28 Ultimo Road, Ultimo NSW 2007

T: +61 2 9514 7713
economics@uts.edu.au

Department of Economics
Working Paper Series

'Weak Stability against Robust Deviations and the Bargaining Set in the Roommate Problem'

Daisuke Hirata [1]
Yusuke Kasuya [2]
Kentaro Tomoeda[3]

[1] Hitotsubashi University
[2] Kobe University
[3] The University of Technology Sydney

No. 1 / 2022 **July 2022**

Weak Stability against Robust Deviations and the Bargaining Set in the Roommate Problem*

Daisuke Hirata† Yusuke Kasuya‡ Kentaro Tomoeda§

This Version: July 9, 2022

Abstract

We propose a new solution concept in the roommate problem, weak stability against robust deviations (wSaRD), by weakening a similar concept of Hirata et al. (2021). We identify a common sufficient condition for wSaRD and weak stability of Klijn and Massó (2003). We can always construct a weakly efficient matching satisfying this condition. Consequently, we can always find a wSaRD matching within the bargaining set. This is in contrast with the original, stronger concept of Hirata et al. (2021), which does not always intersect with the bargaining set.

1 Introduction

The roommate problem is a problem to partition finite agents into pairs (roommates) and singletons. Such a partition is called a matching, and it is called stable if no group of agents can profitably deviate from it by rematching among themselves. Unlike the marriage problem, an instance of the roommate problem may not possess a stable matching (Gale and Shapley, 1962). At the same time, the roommate problem is a special case of coalition and network formation, since a pair can be seen as a coalition or as a network link. For those reasons, the roommate problem is regarded as "an important benchmark for the development of solution concepts for matching, network, and coalition formation models" (Klaus et al., 2010). Indeed, a variety of solution concepts for the roommate problem have been proposed and studied, often by weakening

*Some results in this study were first reported in Kasuya and Tomoeda (2012), then incorporated in the working paper version of Hirata et al. (2020), but omitted from the published version (Hirata et al., 2021). We thank Isa Hafalir, Michihiro Kandori, Mamoru Kaneko, Bettina Klaus, Fuhito Kojima, Akihiko Matsui, Manabu Toda, Alvin E. Roth, Zaifu Yang, and seminar participants at various places for their comments.

†Hitotsubashi University. Email: d.hirata@r.hit-u.ac.jp

‡Kobe University. Email: kasuya@econ.kobe-u.ac.jp

§University of Technology Sydney. Email: Kentaro.Tomoeda@uts.edu.au

stability (e.g, Abraham et al., 2006; Atay et al., 2021; Biró et al., 2016; Inarra et al., 2008; Iñarra et al., 2013; Klaus et al., 2010; and Tan, 1990).

In this line of literature, Hirata et al. (2021) recently propose a solution concept called *stability against robust deviations* (for short, SaRD).[1] A deviation from a matching is defined to be robust if any deviator is weakly better off than at the original matching, after any possible subsequent deviation following her own. A matching is called SaRD if no deviation from it is robust. A robust deviation would be particularly attractive for the deviators in that they will never be strictly worse off even if their own deviation triggers another deviation. The idea of SaRD is to rule out such attractive deviations, if not all possible ones.

As one might have noticed, the concept of SaRD shares a spirit with the *bargaining set*.[2] In cooperative game theory, an imputation belongs to the bargaining set (e.g., Aumann and Maschler, 1964; Zhou, 1994), if any objection against it has a counterobjection. In the matching context (Klijn and Massó, 2003; Atay et al., 2021), imputations and objections become a synonym for matchings and deviations, respectively. Therefore, SaRD and the bargaining set are similar in that both require every possible deviation/objection should be precluded by another deviation/objection.

At the same time, the two concepts also have a subtle difference: The definition of SaRD, on the one hand, presumes that subsequent deviations take place *after* an initial deviation is actually implemented, and it is the expectation of the former that would *indirectly* prevent the latter. In the case of the bargaining set, on the other hand, a counterobjection is an alternative to and would *directly* prevent an original objection, assuming that the former can be proposed *before* the latter is implemented. To judge which concept is more sensible, thus, one would need to scrutinize how deviations/objections are formed and implemented in each application of the model.

In this paper, we propose a weakening of SaRD, which we call *weak stability against robust deviations* (for short, wSaRD), by considering a slightly stronger notion of robust deviations. According to the above definition, a deviation is robust even if there is a chance that after a subsequent deviation, some deviators become indifferent to the original matching. If there is a slight but positive cost to form a deviation (which is outside of the formal model), however, a potential deviator would be reluctant to incur that cost when she would end up being indifferent in terms of matching. Building on this idea, we define a deviation from a matching to be *strongly robust* if any deviator is *strictly better off* than at the original matching, after any possible subsequent deviation. We say that a matching is wSaRD if no deviation is strongly robust.

[1] Precisely speaking, the following is what they call "SaRD up to depth 1."

[2] For other solution concepts related to SaRD, see also Barberà and Gerber (2003), Kurino (2020), and Troyan et al. (2020).

The main result of this paper (the Theorem) is twofold. First, a wSaRD matching always exists in the roommate problem. Second, the intersection of the set of wSaRD matchings and the bargaining set is always non-empty. It should be noted that neither of these holds true for the original SaRD; i.e., no SaRD matching exists in some problem instances, and even when one does, the set of all SaRD matchings may be disjoint from the bargaining set.[3] When the two sets are disjoint, in order to judge between the two solution concepts, one would need to look into the subtle difference between them that originates from institutional details behind the formal model. In contrast, the intersection of wSaRD matchings and the bargaining set, which we show non-empty in this paper, could be considered as a solution concept free from such details, potentially reconciling the divergence of the two related ideas.

To begin, we provide a simple sufficient condition for a matching to be wSaRD (Lemma 1). As it is straightforward to construct a matching meeting it, the existence of a wSaRD matching alone immediately follows (the Corollary). The key is that the sufficient condition allows a mutually-acceptable pair of agents to remain both single (i.e., unmatched). Once a pair of singles deviate and match with each other, neither will ever be strictly worse off after any subsequent deviation than at the original state of being unmatched. Therefore, such a deviation by a pair of singles is always robust. However, it is not necessarily strongly robust, because one of them can become single again after some subsequent deviation. Our sufficient condition indicates an easy way to construct a wSaRD matching by leaving some particular mutually-acceptable pairs unmatched, and as such, it is not sufficient for SaRD.

Actually, the same condition is sufficient not only for wSaRD but also for *weak stability* (Klijn and Massó, 2003), which is another related concept (Lemma 2). A pairwise deviation from a given matching (i.e., a blocking pair) is said to be weak if one of them can find a better partner who also forms a pairwise deviation with her. And a matching is called weakly stable if any pairwise deviation from it is weak. Weak stability essentially imposes the same restriction as of the bargaining set but applies it only to pairwise deviations. Indeed, Klijn and Massó (2003) show in the marriage problem that a matching is in the bargaining set if and only if it is weakly stable and weakly efficient. Atay et al. (2021) recently extend this characterization to the roommate problem.

In establishing our main result, thus, we only need to construct a matching that satisfies weak efficiency as well as the aforementioned condition. It should be noted

[3] These points are already pointed out by Hirata et al. (2021). To overcome the nonexistence, they weaken SaRD by parameterizing the robustness of deviations. Unlike wSaRD here, however, the resulting set of matchings may still be disjoint with the bargaining set, no matter how far they weaken SaRD in their direction.

here that neither the set of wSaRD matchings nor the bargaining set includes the other. Therefore, constructing a wSaRD matching does not generally imply that it belongs to the bargaining set, and vice versa. For this reason, our construction is related to but different from that of Atay et al. (2021), who provide an algorithm to find a matching in the bargaining set, thereby establishing its non-emptiness. Indeed, the matching they construct is *not* wSaRD in some cases.[4] It is also worth mentioning that thanks to the aforementioned condition for wSaRD and weak stability, our construction involves a smaller number of subcases than Atay et al. (2021), thereby being arguably simpler.

2 Preliminaries

A *roommate problem* $(N, \succ)$ consists of a finite set N of agents and a profile $\succ = (\succ_i)_{i \in N}$ of strict preference relations over N. In what follows, generic agents are denoted by i, j, k, a, b, and so on. Given an agent i's strict preference $\succ_i$, we write $j \succeq_i k$ to denote $[j \succ_i k$ or $j = k]$. We say that an agent i is *acceptable to* another agent j if $i \succ_j j$. A matching is a bijection $\mu : N \to N$ satisfying $\mu^2(i) = i$ for all $i \in N$. Given a subset $D \subseteq N$ of agents and two matchings μ and ν, we write $\nu \succ_D \mu$ if $\nu(a) \succ_a \mu(a)$ holds for all $a \in D$, and similarly, $\nu \succeq_D \mu$ if $\nu(a) \succeq_a \mu(a)$ for all $a \in D$. A matching μ is *individually rational* (henceforth, IR) if $\mu \succeq_N \mathrm{id}$, where id denotes the identity mapping over N. It is *weakly efficient* if there is no other matching ν such that $\nu \succ_N \mu$.

A non-empty subset D of agents, associated with a matching ν, is said to form a *deviation from* another matching μ if they prefer ν to μ and can enforce the change from μ to ν by themselves, in the sense that their new partners are also in D. More precisely, we call (D, ν) a deviation from μ and write $\nu \rhd_D \mu$, if

- $\nu \succ_D \mu$,
- $\nu(D) \equiv \{\nu(a) : a \in D\} = D$, and
- $[\mu(i) \in D \Rightarrow \nu(i) = i]$ for any $i \in N - D$,
- $[\mu(j) \notin D \Rightarrow \nu(j) = \mu(j)]$ for any $j \in N - D$.

A pair $\{a, b\}$ of agents is said to *block* a matching μ, if $a \succ_b \mu(b)$ and $b \succ_a \mu(a)$. Note that when μ is IR, a pair $\{a, b\}$ blocks μ if and only if there is a unique ν such that $(\{a, b\}, \nu)$ is a deviation from μ. A matching μ is *stable* if there is no deviation (D, ν) from it. Equivalently, it is stable if it is IR and has no blocking pair.

SaRD, wSaRD, and the Bargaining Set

Now we define several solution concepts for the roommate problem. All of them weaken stability defined above, in allowing a matching to have deviations as long as

[4] Consider, e.g., their cases 1, 2.1, and 2.2.3.1. A concrete example is available upon request.

they would be "unlikely" to realize. Yet they require different, albeit similar, conditions for a deviation to be deemed "unlikely." First, we define *stability against robust deviations* (henceforth, SaRD) and *weak stability against robust deviations* (henceforth, wSaRD) as follows: We say that a deviation (D, v) from a matching μ is *robust* if there is no (D', v') such that

$$v' \rhd_{D'} v \text{ and } \mu(a) \succ_a v'(a) \text{ for some } a \in D, \tag{$*$}$$

and that a matching is *SaRD* if no deviation from it is robust. Similarly, a deviation (D, v) from μ is *strongly robust* if there is no (D', v') such that

$$v' \rhd_{D'} v \text{ and } \mu(a) \succeq_a v'(a) \text{ for some } a \in D, \tag{$\dagger$}$$

and that a matching is *wSaRD* if no deviation from it is strongly robust. Put differently, a matching μ is SaRD (resp. wSaRD) if for any deviation (D, v) from it, there exists (D', v') satisfying condition ($*$) (resp. satisfying condition ($\dagger$)).

Next, we define the *bargaining set* in our setup. An *objection against* a matching μ is a deviation (D, v) from μ. A *counterobjection* against an objection (D, v) is (D', v') such that

$$\begin{cases} D' - D, D - D', D \cap D' \text{ are all non-empty,} \\ [v'(i) \neq \mu(i) \Rightarrow v'(i) \in D'] \text{ for all } i \in D', \\ v'(j) \succeq_j \mu(j) \text{ for all } j \in D' - D, \text{ and} \\ v'(a) \succeq_a v(a) \text{ for all } a \in D' \cap D. \end{cases} \tag{$\ddagger$}$$

The bargaining set is the set of all matching μ's such that any objection against it has a counterobjection. In the current environment, the bargaining set is closely related to the following concepts: Taking a matching μ as given, a blocking pair $\{a, b\}$ against it is said to be *weak* if there is another blocking pair $\{a', b'\}$ against the same μ such that either $[a' = a \text{ and } b' \succ_a b]$ or $[b' = b \text{ and } a' \succ_b a]$. A matching μ is *weakly stable* if it is IR and all blocking pairs (if any) are weak. Then, the bargaining set is characterized as follows:

Fact (Klijn and Massó, 2003; Atay et al., 2021)**.** *A matching belongs to the bargaining set if and only if it is both weakly efficient and weakly stable.*

SaRD and the bargaining set are similar in that both require the existence of some (D', v') that would preclude a deviation/objection (D, v). The key difference lies in what the agents in $D' - D$ compare: In condition ($*$), all agents in D' compare v' and v,

while in (‡), those in $D' - D$ compare ν' and μ. In other words, SaRD requires (D', ν') be a deviation after (D, ν), whereas it is an alternative to (D, ν) in the definition of the bargaining set. As a consequence, the two concepts may totally disagree in the sense that the set of all SaRD matchings can be disjoint from the bargaining set, as Hirata et al. (2021) have already demonstrated. Once we weaken (∗) to (†), however, the set of wSaRD matchings always overlaps with the bargaining set, as we will establish below as the Theorem. The following simple example highlights these points.

Example 1. Let $N = \{1, 2, 3\}$ and $\succ = (\succ_1, \succ_2, \succ_3)$ be such that $(i+1) \succ_i (i-1) \succ_i i$ for each i in modulo 3. In this problem, there are four possible matchings:

$$\mu_0 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \quad \mu_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \quad \mu_2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}, \quad \text{and} \quad \mu_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix},$$

by which we mean, for instance, μ_1 is such that $\mu_1(1) = 1$, $\mu_1(2) = 3$, and $\mu_1(3) = 2$. It is easy to check that among those four, [1] all but μ_0 is SaRD, [2] only μ_0 is weakly stable (and belongs to the bargaining set), and [3] all the four are wSaRD. □

It should be noted that *not necessarily* all matchings in the bargaining set are wSaRD, although they are in the above example. Since condition (†) compares ν' and ν for all agents in D' just as condition (∗) does, some matchings may belong to the bargaining set while not being wSaRD. The following is such an example.

Example 2. Let $N = \{1, 2, 3, 4, 5\}$ and $\succ = (\succ_1, \dots, \succ_5)$ be such that

$$\begin{aligned} &2 \succ_1 4 \succ_1 \mathbf{1} \succ_1 3 \succ_1 5, && 3 \succ_2 1 \succ_2 \mathbf{2} \succ_2 4 \succ_2 5, \\ &4 \succ_3 2 \succ_3 \mathbf{3} \succ_3 1 \succ_3 5, && 1 \succ_4 3 \succ_4 \mathbf{4} \succ_4 2 \succ_4 5, \text{ and} \\ &\mathbf{5} \succ_5 1 \succ_5 2 \succ_5 3 \succ_5 4. \end{aligned}$$

In this problem, all IR matchings are weakly efficient, since agent 5 cannot be better off than being single. Among them, consider the following two:

$$\mu_0 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix}, \quad \text{and} \quad \mu_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 4 & 3 & 5 \end{pmatrix}.$$

Notice that $(\{1, 2, 3, 4\}, \mu_1)$ is a deviation from μ_0 and that μ_1 is a stable matching. Therefore, μ_0 is neither SaRD nor wSaRD. Nevertheless, it is weakly stable, and hence, it belongs to the bargaining set. □

Party Permutation and Related Concepts

Before we present our results, we need to introduce the definitions of a party permutation (Tan, 1991) and related concepts. Given a roommate problem $(N, \succ)$, a bijection σ from N to itself is called a *party permutation* if it meets both of the following conditions for each $i \in N$, where π denotes its inverse σ^{-1}:

- Either $\sigma(i) = \pi(i) = i$, $\sigma(i) = \pi(i) \succ_i i$, or $\sigma(i) \succ_i \pi(i) \succ_i i$.
- For any $j \in N$, if $j \succ_i \pi(i)$, then $\pi(j) \succ_j i$.

Tan (1991) shows that for any problem $(N, \succ)$ at least one party permutation exists.[5] While multiple party permutations may exist for a given problem, the choice among them can be arbitrary for our analysis. In what follows, thus, we take a party permutation σ as given and fixed.

Given the party permutation σ and its inverse π, several more definitions are in order: First, agent i is said to be *inferior* (resp. *superior*) for agent j if $\pi(j) \succeq_j i$ (resp. if $\pi(j) \prec_j i$). By the definition of a party permutation, no pair of agents are mutually superior for each other, while mutually-inferior pairs are possible. Second, a subset of agents is called a *party* if it constitutes a cycle with respect to the party permutation σ. That is, $P \subseteq N$ is a party if $P = \left\{\sigma(i), \sigma^2(i), \ldots, \sigma^{|P|}(i)\right\}$ for each $i \in P$. Note that the set of all parties, denoted by $\mathscr{P}$, is a partition of N. In what follows, for each agent i, let $P(i)$ denote the party that agent i belongs to. A party is called *odd* (resp. *even*) if its cardinality is odd (resp. even).

3 Results

Our main result, the Theorem below, is to establish that the bargaining set *always* contains a wSaRD matching, by constructing such a matching in each of four exhaustive cases. While the details differ across the cases, the key in our proof is to construct a matching μ satisfying the following condition: For each party $P \in \mathscr{P}$,

$$\begin{cases} \text{if } P \text{ is even, then } P \cap I^\circ_\mu = \varnothing, \text{ and} \\ \text{if } P \text{ is odd, either } P \cap I^\circ_\mu = \varnothing \text{ or } P \subseteq I^\circ_\mu, \end{cases} \tag{§}$$

where

$$I^\circ_\mu := \{i \in N : \pi(i) \succ_i \mu(i)\}.$$

[5] While Tan (1991) makes certain assumptions that we do not make in this study, his results continue to hold without them. See appendix D of Hirata et al. (2021) for details.

Note that when $\{i\} \in \mathscr{P}$ is a singleton party (which is a special case of an odd party), $\{i\} \subseteq I^\circ_\mu$ means $i \succ_i \mu(i)$ and hence that μ is not IR. When combined with IR, thus, condition (§) effectively requires $P \cap I^\circ_\mu = \varnothing$ for any singleton party P. To begin, we show that this condition, together with IR, is sufficient for wSaRD.

Lemma 1. *An IR matching μ is wSaRD if it satisfies condition (§) for each party $P \in \mathscr{P}$.*

Proof. Let μ be an IR matching satisfying (§) for each $P \in \mathscr{P}$, and suppose, towards a contradiction, that it is *not* wSaRD. Namely, there is some deviation (D, ν) from it such that

$$\nu'(i) \succ_i \mu(i) \text{ for any deviation } (D', \nu') \text{ from } \nu \text{ and for any } i \in D. \tag{1}$$

Since μ is assumed to be IR, so is ν. Let $D_S := \{a \in D : \nu(a) \succ_a \pi(a)\}$ denote the set of those who deviate to be matched to a superior partner, and define $D_I := D - D_S$. As no pairs of agents are mutually-superior by the definition of a party permutation, $\nu(D_S) \equiv \{\nu(a) : a \in D_S\}$ must be a subset of D_I. Note also that $D_I \subseteq I^\circ_\mu$ by definitions, since $a \in D_I \subseteq D$ implies both $\pi(a) \succeq_a \nu(a)$ and $\nu(a) \succ_a \mu(a)$. For any $a \in D_I$, thus, condition (§) implies that $P(a)$ should be a non-singleton odd party and be a subset of I°_μ.

First, we can derive a contradiction unless the following holds true:

$$\sigma(a), \pi(a) \in D_S \text{ for each } a \in D_I. \tag{2}$$

To see why, arbitrarily fix $a \in D_I$. Since $P(a) \subseteq I^\circ_\mu$ is a non-singleton odd party, then, $a, \sigma(a), \pi(a)$ are all distinct. Then, a contradiction arises if $\sigma(a) \notin D_S$: If so, $\sigma(a)$ should strictly prefer a to $\nu(\sigma(a))$, since $\sigma(a) \in I^\circ_\mu - D_S$. It follows from $a \in D_I$ that a also strictly prefers $\sigma(a)$ to $\nu(a)$. That is, a and $\sigma(a)$ could form a deviation from ν, leading to a new matching ν' such that $\nu' \rhd_{\{a,\sigma(a)\}} \nu$ and $\nu'(\nu(a)) = \nu(a)$. This would be a contradiction to (1) with i being $\nu(a)$.[6] The case of $\pi(a) \notin D_S$ is analogous: If so, a and $\pi(a)$ would form a deviation from ν leaving $\nu(a)$ single, which would contradict (1) with i being $\nu(a)$.

Second, we can still derive a contradiction even if (2) holds true. Suppose that P is a party such that $P \cap D_I$ is non-empty. Recall that such P should be non-singleton odd. Applying (2) for each $a \in P \cap D_I$, hence, we should obtain $|P \cap D_S| > |P \cap D_I|$. That is, for every party $P \in \mathscr{P}$, either $P \cap D_I = \varnothing$ or $|P \cap D_S| > |P \cap D_I|$. Moreover, D_I cannot be empty, since $\nu(D_S) \subseteq D_I$ as noted above. Summing up $|P \cap D_S|$ and $|P \cap D_I|$ across

[6] Note that $\nu(a) \in D$ by $a \in D$ and that $\mu(\nu(a)) \succeq_{\nu(a)} \nu(a)$ follows from IR.

all parties, thus, we should have $|D_S| > |D_I|$. However, this is a contradiction, because $|D_S| \leq |D_I|$ also follows from $v(D_S) \subseteq D_I$ and $v(\cdot)$ being a bijection. ■

With any given instance of the roommate problem, it is straightforward to construct an IR matching meeting condition (§). If we set aside the bargaining set, thus, the existence of a wSaRD matching immediately follows from Lemma 1 alone.

Corollary. *For any roommate problem $(N, \succ)$, there is a wSaRD matching.*

Proof. Given the party permutation σ, we can easily construct a matching μ such that

- the even parties are matched into "adjacent" pairs, i.e., $\mu(i) \in \{\pi(a), \sigma(a)\}$ for any $i \in N$ such that $P(i)$ is even; and
- the odd parties, including the singletons, are left fully unmatched, i.e., $\mu(j) = j$ for any $j \in N$ such that $P(j)$ is odd.

Any such μ is wSaRD, since it is IR and satisfies condition (§) for every party $P \in \mathscr{P}$ ■

Next, we demonstrate that the same condition is also sufficient for weak stability.

Lemma 2. *An IR matching μ is weakly stable if it satisfies condition (§) for each party $P \in \mathscr{P}$.*

Proof. Let μ be an IR matching satisfying (§) and $\{a, b\}$ an arbitrary blocking pair against μ. Without loss of generality (by the definition of a party permutation), assume b is inferior for a; i.e., $\pi(a) \succeq_a b$. As we have seen in the first paragraph of the proof of Lemma 1, $P(a)$ should be a non-singleton odd party and be a subset of I°_μ. The first observation implies $\sigma(a) \succ_a b$, since $\sigma(a) \succ_a \pi(a)$ by the definition of a party permutation. The second means that $\sigma(a)$ should also prefer $a \equiv \pi(\sigma(a))$ to $\mu(\sigma(a))$. Combining these together, we can conclude that $\{a, b\}$ is a weak blocking pair. ■

Combining Lemmas 1–2 with the characterization of the bargaining set (the Fact above), our goal reduces to the construction of a matching satisfying IR, weak efficiency, and condition (§) for each party P. By constructing such a matching for each possible case, we establish the following:

Theorem. *For any roommate problem $(N, \succ)$, there is a matching that is wSaRD and belongs to the bargaining set.*

Proof. By Lemmas 1–2 and the Fact above, it suffices to establish the existence of a matching that is IR, weakly efficient, and satisfying condition (§) for each party. To begin, let $\mathcal{M}$ be the set of IR matching μ's such that for all $i, j \in N$,

$$\begin{cases} [P(i) \text{ is even}] \Rightarrow [\mu(i) \succ_i \pi(i) \text{ and } P(\mu(i)) \text{ is odd}], & \text{and} \\ [P(j) \text{ is odd}] \Rightarrow [\text{either } \mu(j) = j \text{ or } P(\mu(j)) \text{ is even}]. \end{cases}$$

By definition, any matching in $\mathcal{M}$ meets condition (§) for each party $P \in \mathscr{P}$. For the moment, suppose $\mathcal{M}$ is non-empty. Then, it contains some μ^* that is maximal within $\mathcal{M}$ with respect to $\succ_{N_e}$, where $N_e := \{i \in N : P(i) \text{ is even}\}$. In other words, $\mu^* \in \mathcal{M}$ is such that for any $\mu' \in \mathcal{M}$, there is $i \in N$ such that $P(i)$ is even and $\mu^*(i) \succeq_i \mu'(i)$. Any such μ^* must be weakly efficient for the following reason: Towards a contradiction, suppose $\mu' \succ_N \mu^*$ for some matching μ'. By assumptions, any $i \in N_e$ should be matched to a superior partner at μ'. Since no pairs are mutually superior, it follows that $\mu'(i) \notin N_e$ for all $i \in N_e$. We can then construct another matching $\mu'' \in \mathcal{M}$ such that $\mu'' \succ_{N_e} \mu^*$, by defining $\mu''(i) := \mu'(i)$ if either i or $\mu'(i)$ is in N_e and $\mu''(i) := i$ otherwise; however, this is a contradiction to the maximality of μ^*. Therefore, whenever $\mathcal{M}$ is non-empty, at least one of its elements is weakly efficient, in addition to being IR and satisfying condition (§). For the rest of this proof, we consider the case where $\mathcal{M}$ is empty, and we divide it into three subcases.

Case 1: There is at least one even party. In this case, consider a matching μ that we constructed in the proof of the Corollary; that is, μ is such that

- the even parties are matched into "adjacent" pairs, i.e., $\mu(i) \in \{\pi(a), \sigma(a)\}$ for any $i \in N$ such that $P(i)$ is even; and
- the odd parties, including the singletons, are left fully unmatched, i.e., $\mu(j) = j$ for any $j \in N$ such that $P(j)$ is odd.

Any such μ is IR and meets (§) for each party $P \in \mathscr{P}$. Hence, it suffices to demonstrate its weak efficiency. Towards a contradiction, suppose $\mu' \succ_N \mu$ for some μ'. For any $i \in N_e$, then, $\mu'(i) \succ_i \pi(i)$ by the construction of μ. Since no pairs of agents is mutually superior for each other, $P(\mu'(i))$ should be odd for each $i \in N_e$. That is, μ' satisfies the first requirement for belonging to $\mathcal{M}$. We can then construct $\mu'' \in \mathcal{M}$ by letting $\mu''(i) := \mu'(i)$ if either i or $\mu'(i)$ is in N_e and $\mu''(i) := i$ otherwise. This contradicts our assumption of $\mathcal{M}$ being empty.

Case 2: All parties are odd, and there is a mutually-acceptable pair $\{a, b\}$ of agents such that $P(a) \neq P(b)$ and $a \succ_b \pi(b)$. Let $\{a, b\}$ be such a pair. It is without loss of generality to assume a is the best partner for b within $N - P(b)$; i.e., $[P(i) \neq P(b) \Rightarrow a \succeq_b i]$ for any $i \in N$. Define μ to be the unique matching such that

- a and b are matched to each other, i.e., $\mu(a) = b$;
- the other agents in party $P(b)$ are matched into $\frac{P(b)-1}{2}$ pairs, i.e., $\mu(i) \in \{\pi(i), \sigma(i)\}$ for all $i \in P(b) - \{b\}$; and
- all the other agents are left single, i.e., $\mu(j) = j$ for all $j \notin \{a\} \cup P(b)$.

By definition, $P(b) \cap I^\circ_\mu = \varnothing$ while any other party $P \neq P(b)$, which is odd by assumption, is either disjoint from I°_μ (if P is a singleton) or is a subset of I°_μ (otherwise). Therefore, this μ meets condition (§) for each party $P \in \mathscr{P}$. Since it is also clear that μ is IR, hence, it suffices to demonstrate its weak efficiency. Towards a contradiction, suppose $\mu' \succ_N \mu$ for some matching μ'. By our assumptions on $\{a, b\}$, $\mu'(b)$ must be superior for b and be a member of $P(b)$. By the construction of μ, the latter further implies $b \succ_{\mu'(b)} \pi(\mu'(b))$. That is, b and $\mu'(b)$ should be mutually superior for each other, but this contradicts the definition of a party permutation.

Case 3: All parties are odd, and for any mutually-acceptable pair $\{i, j\}$ of agents, $P(i) \neq P(j) \Rightarrow \pi(i) \succ_i j$. In this case, arbitrarily fix a party $P^* \in \mathscr{P}$ and consider an IR matching μ such that

- for any $i \in N$ such that $\mu(i) \neq i$, $[i \in P^* \Leftrightarrow \mu(i) \notin P^*]$; and
- for any blocking pair $\{a, b\}$ against μ, either $\{a, b\} \subseteq P^*$ or $\{a, b\} \subseteq N - P^*$.

Notice that we can always construct such a matching. To see why, define $\sqsupset = (\sqsupset_i)_{i \in N}$ to be a preference profile that we can obtain from $\succ$ by making all the pairs within P^* and those within $N - P^*$ mutually unacceptable. More precisely, let $\sqsupset$ be such that

- $j \sqsupseteq_i k \sqsupset_i i \Leftrightarrow [j, k \notin P^* \text{ and } j \succeq_i k \succ_i i]$, for any $i \in P^*$ and any $j, k \in N$; and
- $j \sqsupseteq_i k \sqsupset_i i \Leftrightarrow [j, k \in P^* \text{ and } j \succeq_i k \succ_i i]$, for any $i \notin P^*$ and any $j, k \in N$.

Then, an IR matching μ satisfies the above two conditions in the original $(N, \succ)$ if and only if it is a stable matching in $(N, \sqsupset)$. Further, $(N, \sqsupset)$ is a marriage problem with the two "sides" of agents being P^* and $N - P^*$. We can thus construct a stable matching of $(N, \sqsupset)$ via the deferred acceptance algorithm.

Further, any such μ meets condition (§) for each party $P \in \mathscr{P}$ for the following reasons: First, suppose $\{i\} \in \mathscr{P}$ is a singleton party. By the assumption of this case, then, agent i does not constitute a mutually-acceptable pair with any other agent j. Since μ is constructed to be IR, $\mu(i) = i$ and hence $\{i\} \cap I^\circ_\mu = \varnothing$. Second, let $P \in \mathscr{P}$ be a non-singleton party, which is odd by the assumption of this case, and arbitrary fix its member $i \in P$. If $\mu(i) = i$, then $i \in I^\circ_\mu$ by definitions. If $\mu(i) \neq i$, then i and $\mu(i)$ should be mutually acceptable by the IR of μ and should belong to different parties by the first requirement for μ. By the assumption of this case, these together imply $i \in I^\circ_\mu$. Since i is arbitrary, $i \in I^\circ_\mu$ holds for any $i \in P$, and thus, $P \subseteq I^\circ_\mu$.

What remains to be shown is the weak efficiency of μ. Towards a contradiction, suppose $\mu' \succ_N \mu$ for some μ'. Since μ is IR, no agent is single at μ'. For any $i \in N$, thus, $\{i, \mu'(i)\}$ should be a blocking pair against μ; by the second requirement for μ, it is a subset of P^* or of $N - P^*$. In particular, each $j \in P^*$ must be matched to another member of P^* at μ'. However, this is impossible because $|P^*|$ is odd by assumption. ■

References

Abraham, D. J., P. Biró, and D. F. Manlove (2006): ""Almost Stable" Matchings in the Roommates Problem," in *Approximation and Online Algorithms: Third International Workshop, WAOA 2005*, ed. by T. Erlebach and G. Persiano, Springer Berlin Heidelberg, 1–14.

Atay, A., A. Mauleon, and V. Vannetelbosch (2021): "A Bargaining Set for Roommate Problems," *Journal of Mathematical Economics*, 94, Article 102465.

Aumann, R. J. and M. Maschler (1964): "The Bargaining Set for Cooperative Games," in *Advances in Game Theory*, ed. by M. Dresher, L. S. Shapley, and A. W. Tucker, Princeton University Press, Princeton, 443–476.

Barberà, S. and A. Gerber (2003): "On Coalition Formation: Durable Coalition Structures," *Mathematical Social Sciences*, 45, 185–203.

Biró, P., E. Iñarra, and E. Molis (2016): "A new solution concept for the roommate problem: $\mathscr{Q}$-stable matchings," *Mathematical Social Sciences*, 79, 74–82.

Gale, D. and L. S. Shapley (1962): "College Admissions and the Stability of Marriage," *American Mathematical Monthly*, 69, 9–15.

Hirata, D., Y. Kasuya, and K. Tomoeda (2020): "Stability against Robust Deviations in the Roommate Problem," *Available at SSRN 3552365*.

——— (2021): "Stability against Robust Deviations in the Roommate Problem," *Games and Economic Behavior*, 130, 474–498.

Iñarra, E., C. Larrea, and E. Molis (2013): "Absorbing sets in roommate problems," *Games and Economic Behavior*, 81, 165–178.

Inarra, E., C. Larrea, and E. Molis (2008): "Random Paths to P-Stability in the Roommate Problem," *International Journal of Game Theory*, 36, 461–471.

Kasuya, Y. and K. Tomoeda (2012): "Credible Stability in the Roommate Problem," *mimeo*.

Klaus, B., F. Klijn, and M. Walzl (2010): "Stochastic Stability for Roommate Markets," *Journal of Economic Theory*, 145, 2218–2240.

Klijn, F. and J. Massó (2003): "Weak Stability and a Bargaining Set for the Marriage Model," *Games and Economic Behavior*, 42, 91–100.

Kurino, M. (2020): "Credibility, Efficiency, and Stability: A Theory of Dynamic Matching Markets," *Japanese Economic Review*, 71, 135–165.

Tan, J. J. M. (1990): "A Maximum Stable Matching for the Roommate Problem," *BIT*, 29, 631–640.

——— (1991): "A Necessary and Sufficient Condition for the Existence of a Complete Stable Matching," *Journal of Algorithms*, 12, 154–178.

Troyan, P., D. Delacrétaz, and A. Kloosterman (2020): "Essentially Stable Matchings," *Games and Economic Bahavior*, 120, 370–390.

Zhou, L. (1994): "A New Bargaining Set of an N-Person Game and Endogeneous Coalition Formation," *Games and Economic Behavior*, 6, 512–526.