Stability against Robust Deviations in the Roommate Problem

Daisuke Hirata* Yusuke Kasuya† Kentaro Tomoeda‡

This Version: July 6, 2021

Abstract

We propose a new solution concept in the roommate problem, based on the "robustness" of deviations (i.e., blocking coalitions). We call a deviation from a matching *robust up to depth k*, if none of the deviators gets worse off than at the original matching after any sequence of at most k subsequent deviations. We say that a matching is *stable against robust deviations (for short, SaRD) up to depth k*, if no deviation from it is robust up to depth k. As a smaller k imposes a stronger requirement for a matching to be SaRD, we investigate the existence of a matching that is SaRD with a minimal depth k. We constructively demonstrate that a SaRD matching always exists for $k = 3$ and establish sufficient conditions for $k = 1$ and 2.

*Hitotsubashi University and UNSW Sydney; d.hirata@r.hit-u.ac.jp
†Kobe University; kasuya@econ.kobe-u.ac.jp
‡University of Technology Sydney; Kentaro.Tomoeda@uts.edu.au

Contents

1 Introduction

In matching models, a matching (or an outcome) is called *stable* if no group of agents can profitably deviate from it by rematching among themselves. Stability has been a central concept in various strands of the literature: It is not just the most popular desideratum in the design of two-sided problems such as school choice (Abdulkadiroğlu and Sönmez, 2003), but also a primary solution concept in one-sided models such as hedonic coalition formation (Bogomolnaia and Jackson, 2002) and network formation (Jackson, 2008).[1] However, it is also well known that a stable matching may not exist in one-sided models. This is true even in the simplest class called the *roommate problem* (Gale and Shapley, 1962), which is a problem to partition finite agents into pairs (roommates) and singletons.[2] Since it is a special case both of coalition formation and of network formation, studying the roommate problem is a natural first step to understand stability in one-sided matching problems.[3]

The purpose of this paper is to propose a new solution concept that weakens stability in the roommate problem. When no matching is stable (i.e., when *any* matching admits some profitable deviations), a natural solution concept would "minimize" remaining deviations in some sense. The simplest way to do so is to treat all possible deviations equally and minimize the number of them (Abraham et al., 2006). Alternatively, one could argue that deviations differ in their "seriousness" and that a matching is "more instable" when it admits "more serious" deviations. From such a perspective, a solution should minimize the "seriousness" of, instead of the number of, the deviations. We take this alternative approach and differentiate deviations from a matching based on their robustness as defined below, although there might be other plausible criteria of "seriousness."

[1] By "one-sided" models, we refer to those where any agent can be matched with any other. This is in contrast to "two-sided" models, where agents are partitioned into two sides and any match is between the two sides.

[2] Moreover, the proportion of preference profiles with no stable matching increases steeply as the number of agents increases (Gusfield and Irving, 1989; Pittel and Irving, 1994).

[3] Indeed, Klaus et al. (2010, p. 2219) write "roommate markets can be considered as an important benchmark for the development of solution concepts for matching, network and coalition formation models."

Specifically, we call a deviation from a matching *robust up to depth k*, if none of the deviators gets worse off than at the original matching they deviate from, after *any* sequence of at most k subsequent deviations. Suppose that a group D of agents deviates from a matching μ and leads to another matching ν. The robustness of this deviation depends on what can happen after ν is once formed. Suppose, for instance, that ν has three possible deviations that lead to $\nu_1, \ldots, \nu_3$, respectively, and each of these ν_i's also has three deviations to $\nu_{i,1}, \ldots, \nu_{i,3}$. Starting from ν, then, each of $\nu_1, \ldots, \nu_3$ is reachable by a single deviation, while $\nu_{1,1}, \ldots, \nu_{3,3}$ are by a sequence of two. In this example, thus, the original deviation by D from μ to ν is robust up to depth 1 (resp. depth 2) if none of D is worse off at any of $\nu_1, \ldots, \nu_3$ (resp. any of $\nu_1, \ldots, \nu_3, \nu_{1,1}, \ldots, \nu_{3,3}$) than at μ.

We have two possible interpretations of the robustness defined above. The first is to assume the agents have max-min preferences and bounded rationality. Upon the decision to form a deviation or not, such agents would search for the worst-case consequence of the deviation subject to a finite depth k of reasoning. In this interpretation, the more sophisticated the agents are, the harder it is for them to agree on a possible deviation. When a deviation is robust up to a large depth k, however, it would be reachable even among extremely risk-averse and highly sophisticated agents. Secondly, but not less importantly, we can interpret the depth k as the length of time. In reality, forming a deviation should take a certain period of time. In the context of business alliances, for example, it should take time to reach an agreement with a new partner or to dissolve an old partnership. Assuming only one deviation can realize per time period, the gain from a deviation lasts for at least k periods, whatever reactions it triggers in the future, if it is robust up to depth k; otherwise, the deviators must accept the risk of potential losses within a shorter time window. With this interpretation, too, it would be natural to argue that a deviation is more-easily agreeable when it is robust up to a larger depth k.

By measuring the robustness of a deviation with its depth k, we seek matchings that only admit minimally robust deviations. We say that a matching is *stable against*

robust deviations (for short, SaRD) *up to depth* k, if no deviation from it is robust up to depth k (or larger).[4] By definition, if a deviation is robust up to some depth k, then it is so up to any smaller depth k'. Therefore, if a matching is SaRD up to some depth k', then it is so up to any larger depth k. That is, the smaller the depth k is, the stronger requirement SaRD up to depth k becomes.

Our main results are on the existence of SaRD matchings up to depth $k = 1, 2$, and 3. Our first two results (Theorems 1–2) identify sufficient conditions for the existence of a matching that is SaRD up to depth 1 and 2. It should be noted that such matchings do not always exist, as we will see later in this introduction. In contrast, our last main result (Theorem 3) establishes the general existence of a SaRD matching up to depth 3. Namely, we can *construct* a matching that is SaRD up to depth 3 for any roommate problem, i.e., for any set of agents and any preference profile.

For the rest of this introduction, we sketch the key ideas underlying our main results in a simple class of examples: Suppose that there are an odd number $n > 1$ of agents, $a_1, \ldots, a_n$. Each a_i's preference is such that only a_{i+1} and a_{i-1} are acceptable (i.e., better than being single) and the former is preferred to the latter, where the subscripts are in modulo n. Notice that this is a typical case where no stable matching exists. Below, we see how the existence of SaRD matchings depends on and varies with the parameter n. We then briefly explain how the ideas in those simple examples extend to the general case.

In this class of problems, any SaRD matching should satisfy two principles: First, at a SaRD matching, each a_i should be either single or matched to one of a_{i+1} and a_{i-1}; otherwise (i.e., if a_i is matched to an unacceptable agent), she can unilaterally deviate to be single and will never be worse off after any array of voluntary deviations. Second, if a_i is single at a SaRD matching, then a_{i+1} cannot be single; otherwise, they can deviate to be matched to each other and will never be worse off than at the initial situation of being single.

When $n \leq 7$, the above two principles pin down the SaRD matchings for each

[4] In what follows, we use the acronym "SaRD" both as an adjective ("S" for stable) and as a noun ("S" for stability).

fixed n, but their depths vary with n. When $n = 3$, 5, and 7, the principles require to match one, two, and three mutually-acceptable pairs, respectively. Note that all such matchings are symmetric up to rotation for each $n \leq 7$. When $n = 3$, they are SaRD up to depth 1. For instance, consider the matching $\mu_{(3)} = \{\{a_1, a_2\}, \{a_3\}\}$, which means that a_1 and a_2 are matched to each other and a_3 is single. From this $\mu_{(3)}$, the unique possible deviation is by $\{a_2, a_3\}$. After they deviate and are matched to each other, there is another deviation by $\{a_3, a_1\}$. If this subsequent deviation realizes, one of the original deviator, a_2, becomes single and worse off than at $\mu_{(3)}$. That is, the unique deviation from $\mu_{(3)}$ is not robust up to depth 1, and hence, $\mu_{(3)}$ is SaRD up to depth 1.

When $n = 5$, the candidate matchings are all SaRD up to depth 2, but none is SaRD up to depth 1. Let us consider $\mu_{(5)} = \{\{a_1, a_2\}, \{a_3, a_4\}, \{a_5\}\}$, where only a_5 is single. Starting from this $\mu_{(5)}$, the unique deviation is by $D = \{a_4, a_5\}$, and thereafter, there is a unique chain of subsequent deviations, first by $D_1 = \{a_2, a_3\}$, then by $D_2 = \{a_5, a_1\}$, and so on. Notice that the initial deviators, a_4 and a_5, remain matched even after D_1 follows. Therefore, the deviation by D from $\mu_{(5)}$ is robust up to depth 1, and $\mu_{(5)}$ is not SaRD up to depth 1. Yet, if D_2 further deviates following D and D_1, then a_4 becomes single while she was originally matched to a_3. Thus, the original deviation by D is not robust up to depth 2, and $\mu_{(5)}$ is SaRD up to depth 2. The case of $n = 7$ is similar: The matchings that match three (mutually-acceptable) pairs are SaRD up to depth 3 but no matching is SaRD up to depth 2.

When $n \geq 9$, the two principles no longer pin down the number of matched pairs for a matching to be SaRD, and there are SaRD matchings up to different depths even for a fixed n. In particular, matching as many pairs as possible may undermine the degree of stability measured by depth k. To see the point, suppose $n = 9$ and consider two matchings,

$$\mu_{(9)} = \{\{a_1, a_2\}, \{a_3, a_4\}, \{a_5, a_6\}, \{a_7, a_8\}, \{a_9\}\}, \text{ and}$$
$$\mu'_{(9)} = \{\{a_1, a_2\}, \{a_3\}, \{a_4, a_5\}, \{a_6\}, \{a_7, a_8\}, \{a_9\}\}.$$

Note that they differ in the number of matched pairs, although both satisfy the two principles. Moreover, they are both SaRD but up to different depths. By similar arguments as above, one can check $\mu_{(9)}$ is SaRD up to depth 4 but not up to depth 3. On the other hand, $\mu'_{(9)}$ is SaRD up to depth 2. Suppose, for example, that $D = \{a_2, a_3\}$ deviates from $\mu'_{(9)}$. Then, $a_2 \in D$ can be single after two subsequent deviations, first by $D_1 = \{a_5, a_6\}$ and then by $D_2 = \{a_3, a_4\}$. One can check in a similar manner that all the other deviations are not robust up to depth 2, and hence, $\mu'_{(9)}$ is SaRD up to depth 2.

The case of $n = 9$ is still special in that any matching with three matched pairs is symmetric to $\mu'_{(9)}$; in general, what is critical in our construction is the number of "consecutive" matched pairs, rather than the total number of matched pairs. For example, suppose $n = 13$ and consider two matchings,

$$\mu_{(13)} = \{\{a_1, a_2\}, \{a_3\}, \{a_4, a_5\}, \{a_6, a_7\}, \{a_8, a_9\}, \{a_{10}\}, \{a_{11}, a_{12}\}, \{a_{13}\}\}, \text{ and}$$
$$\mu'_{(13)} = \{\{a_1, a_2\}, \{a_3, a_4\}, \{a_5\}, \{a_6, a_7\}, \{a_8, a_9\}, \{a_{10}\}, \{a_{11}, a_{12}\}, \{a_{13}\}\},$$

both of which match five pairs. Note that $\mu_{(13)}$ matches three "consecutive" pairs, from $\{a_4, a_5\}$ to $\{a_8, a_9\}$. It should be easy by now to see that as a consequence, the deviation by $\{a_2, a_3\}$ from it is robust up to depth 3; i.e., $\mu_{(13)}$ is not SaRD up to depth 3. In contrast, $\mu'_{(13)}$ matches at most two pairs "consecutively," such as $\{a_1, a_2\}$ and $\{a_3, a_4\}$. As a result, $\mu'_{(13)}$ is SaRD up to depth 3, even though it matches the same number of pairs as $\mu_{(13)}$ does.

While we have focused on the simple cases, they actually capture the essence of the general case. This is because for any set of agents and any preference profile, we can always partition the agents so that within each subset, their preferences form a cycle as in the above examples (Tan, 1991). First, this allows us to parametrize problems with the lengths of preference cycles, as we did above with n. In particular, our Theorems 1–2 state that there is a matching that is SaRD up to depth $k = 1$ and 2, if all cycles consisting of an odd number of agents are sufficiently short. Thus, these theorems

generalize the cases of $n = 3$ and 5 in the above examples. Second, the cycle structure determines (i) whether a pair of agents are "adjacent" to each other, as a_i and a_{i+1} above, and (ii) whether two "adjacent" pairs of agents are "consecutive." Unlike the above examples, "non-adjacent" pairs may be mutually acceptable in the general case, and we do match such pairs in our construction of a SaRD matching up to depth 3 (Theorem 3). Yet, how to match adjacent pairs remains to be the key in certain senses. In particular, we need to carefully control the number of "consecutively-matched" adjacent pairs, as we will elaborate in Section 4.

The rest of the paper is organized as follows: Section 1.1 briefly overviews the related literature. Section 2 introduces our model and key definitions. Section 3 presents the main results, and Section 4 illustrates the key ideas behind them. Section 5 discusses the relationships between our SaRD and other solutions concepts. Appendices A–C provide the proofs. Appendix D discusses the conditions for Tan's (1991) results, which we heavily exploit in our analysis.

1.1 Related Literature

In the literature, a number of studies have defined stability concepts based on chains of deviations and their final outcomes, in a similar spirit with ours. Among others, the most closely related is Barberà and Gerber (2003). They study the hedonic coalition formation, which generalizes the roommate problem, and propose a solution concept called *durability*. We share the spirit with them in distinguishing what we call robust deviations, and actually, in the roommate problem their durability coincides with our SaRD up to a sufficiently large depth k. However, we further differentiate robust deviations across different depths and look for a SaRD matching up to a minimal depth, whereas Barberà and Gerber (2003) treat all deviation chains of any length as equally serious. The set of SaRD matchings up to depth 3 is generally smaller than that of durable matchings, and hence, our concept can be seen as a refinement of durability. Relatedly, Troyan et al. (2020) propose in the school choice problem a solution concept called *essential stability*, which also corresponds to our SaRD with a sufficiently large k.

It should be noted, however, that a stable matching always exists in the school choice problem and their motivation differs from ours.

While we investigate a static model with dynamic arguments as a possible interpretation and motivation, Kadam and Kotowski (2018) and Kotowski (2015) explicitly study a dynamic marriage market, where agents have their preferences over the histories (i.e., sequences) of matched partners. They also define stability concepts for their dynamic setting, but it should be noted that their concepts reduce to the standard stability in the static setting. Also in a dynamic marriage market, Kurino (2019) proposes *credible stability*, which reduces in the static setting to a weaker version of our SaRD up to depth $k = 1$.[5]

Unsolvable roommate problems have long been studied in economics and other related fields, and several more solution concepts have been proposed. These include the maximum stable matchings (Tan, 1990), almost stable matchings (Abraham et al., 2006), P-stable matchings (Inarra et al., 2008), absorbing sets (Iñarra et al., 2013), and Q-stable matchings (Biró et al., 2016). Each of them partially extends the properties of stability to unsolvable problems in a certain direction. In addition, several studies apply other general concepts than stability, such as stochastic stability (Klaus et al., 2010) and farsighted stable sets (Klaus et al., 2011), to the roommate problem. The relation between our SaRD and other solution concepts will be discussed in more detail in Section 5.

2 Preliminaries

A *roommate problem* $(N, \succ)$ consists of a finite set N of agents and a profile $\succ = (\succ_a)_{a \in N}$ of strict preference relations over N. Given agent a's strict preference $\succ_a$, we write $b \succeq_a c$ to denote $[b \succ_a c \text{ or } b = c]$. We say that an agent a is *acceptable to* another agent b if $a \succ_b b$. A matching is a bijection $\mu : N \to N$ satisfying $\mu^2(a) = a$ for all $a \in N$. We also identify a matching with the partition it induces; e.g., when

[5] In Appendix F of the working paper version (Hirata et al., 2020), we formally define this weaker concept and establish its existence.

we write $\mu = \{\{a, b\}, \{c\}\}$, it refers to the matching defined by $\mu(a) = b$, $\mu(b) = a$, and $\mu(c) = c$. 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 $\nu(a) \succeq_a \mu(a)$ holds for all $a \in D$. A matching μ is called *individually rational* if $\mu \succeq_N \mathrm{id}$, where id denotes the identity mapping over N. A matching μ is said to *leave no mutually-acceptable pairs of singles* if

$$[a \succ_b b \text{ and } b \succ_a a] \Longrightarrow [\mu(a) \neq a \text{ or } \mu(b) \neq b],$$

holds for all $a, b \in N$. This can be seen as a mild efficiency property, as a mutually-acceptable pair of singles implies Pareto inefficiency. Let us call a matching *regular* if it is individually rational and leaves no mutually-acceptable pairs of singles.

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 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 (1) $\nu \succ_D \mu$, (2) $\nu(a) \in D$ for any $a \in D$, (3) $\mu(i) \in D \Rightarrow \nu(i) = i$ for any $i \in N - D$, and (4) $\mu(j) \notin D \Rightarrow \nu(j) = \mu(j)$ for any $j \in N - D$.[6] When μ is individually rational and $|D| = 2$, the identity of D pins down the unique matching ν such that (D, ν) can be a deviation from μ. More specifically, for $\nu \rhd_{\{a,b\}} \mu$ to hold given μ is individually rational, ν needs to be such that $\nu(a) = b$, $\nu(b) = a$, $\nu(i) = i$ for all $i \in \{\mu(a), \mu(b)\} - \{a, b\}$, and $\nu(j) = \mu(j)$ for all $j \notin \{a, b, \mu(a), \mu(b)\}$. Although we will not fully specify the associated ν when $|D| = 2$, it should thus cause no confusion. A matching μ is *stable* if there is no deviation (D, ν) from it.

Now we introduce our key concepts. A deviation (D, ν) from μ is called *robust up to depth* $k \in \mathbb{N}$, if $\nu_\kappa \succeq_D \mu$ holds for any sequence of deviations $(D_1, \nu_1), \ldots, (D_\kappa, \nu_\kappa)$ with $\kappa \leq k$ such that

$$\nu_\kappa \rhd_{D_\kappa} \nu_{\kappa-1} \rhd_{D_{\kappa-1}} \cdots \rhd_{D_2} \nu_1 \rhd_{D_1} \nu. \qquad (*)$$

[6] Part (3) of this definition implicitly assumes that the partners of the members of D at μ are left single after the deviation. In Section 2.1.2, we discuss an alternative definition of a deviation that allows for instantaneous rematch among the agents who are left behind by D.

When no deviation from it is robust up to depth k, a matching μ is said to be *stable against robust deviations* (henceforce, *SaRD*) up to depth k. By definition, if a deviation is robust up to depth k, then so is it up to any depth $k' < k$. Consequently, if a matching is SaRD up to depth k, then so is it up to any depth $k'' > k$. As we argued for the simple examples in the introduction, any SaRD matching must be individually rational and leave no mutually-acceptable pairs of singles:

Proposition 1. *For any $k \geq 1$, if a matching μ is SaRD up to depth k, then it is regular.*

Proof. The proof is straightforward and is thus omitted. ■

Before we present our results in Section 3, the rest of this section is organized as follows: In Section 2.1, we further discuss our definition of SaRD, addressing possible conceptual concerns. In Section 2.2, we introduce the concepts and results of Tan (1991), which we will heavily rely on in our analysis.

2.1 Discussions of our Concepts and Definitions

2.1.1 Consistency of the Definition of SaRD Matchings

One might argue that our concept of SaRD is inconsistent in that we try to exclude robust deviations while we allow non-robust subsequent deviations in defining robust deviations per se. In response to such a concern, we make two remarks. First, requiring consistency could lead to some subtlety, making it difficult for the solution to be a matching-wise concept. A natural way to require consistency would be to call a deviation "consistently robust" if the original deviators will never be worse-off after any subsequent deviations as long as those subsequent deviations are also "consistently robust." Since this definition is self-referential, the set of all "consistently robust" deviations should be a fixed point of an equation with the variable being a set of deviations; once we solve for it, we can further identify "consistently SaRD" matchings using it. However, such an equation might have multiple fixed points, each corresponding to *a different set of all "consistently robust" deviations*. As a result,

a matching may be "consistently SaRD" according to one well-defined set of "consistently robust" deviations but not according to another. Such multiplicity would be unappealing because, for instance, it makes it difficult to directly compare the degree of stability between two matchings. Comparing matchings would reduce to comparing the sets of "consistently robust" deviations supporting them, but the latter would in turn require something outside our model, such as beliefs of the agents.

Secondly, but not less importantly, we do not claim that a SaRD matching is fully immune to deviations or, in other words, that non-robust deviations would never realize. Instead we would argue, as we did in the introduction, that robust deviations would be more likely to realize than the others and hence, that SaRD matchings would be "less unstable" than the others. And our argument could still apply even if we define "consistently robust" deviations as above: The benefit from such a deviation is guaranteed under the hypothesis that only "consistently robust" deviations can follow. This hypothesis might be true if every agent is sophisticated enough to tell whether a deviation is "consistently robust" or not based on a shared criterion. However, even if an agent herself is sophisticated, she could be unsure if the others are also sophisticated. Further, even if she believes the others to be sophisticated as well, she could be still unsure what criteria of "consistent robustness" they adopt, since there could be multiple of them as argued above. For an agent facing such ambiguities, a deviation would be less secure when it is "consistently robust" than when it is robust in our sense. Our strategy in this study is to eliminate deviations that would be the most secure and likely to realize.

2.1.2 Definition of Deviations

Our definition requires a deviation (D, ν) from μ to satisfy $\nu(i) = i$ if $i \notin D$ and $\mu(i) \in D$. That is, we implicitly assume that the agents who are left behind by D remain single at ν, while one might argue that those agents could instantaneously rematch among themselves. To be concrete, let us call (D, ξ) a *deviation with possible (instantaneous) rematch* from μ and write $\xi \rhd^*_D \mu$ if (1) $\xi \succ_D \mu$, (2) $\xi(a) \in D$ for any

$a \in D$, (3') $\mu(i) \in D \Rightarrow \xi(i) \succeq_i i$ for any $i \in N - D$, and (4) $\mu(j) \notin D \Rightarrow \xi(j) = \mu(j)$ for any $j \in N - D$. We can also define the robustness of a deviation (with rematch) and the SaRD property using $\rhd^*$ instead of $\rhd$.

We make two remarks on such alternative definitions. First, once we fix an initial deviation (with or without rematch) from an original matching, its robustness measured by depth k is independent of whether we allow rematch or not in subsequent deviations, i.e., whether we use $\rhd$ or $\rhd^*$. On the one hand, rematch in subsequent deviations must not increase the depth k. This is because part (3') in the definition of $\rhd^*$ allows $\xi(i) = i$ and hence, the deviations with rematch include those without rematch as a subset. On the other hand, allowing rematch cannot decrease the depth, either. Suppose that an original deviator is worse off after a subsequent deviation *and* rematch. Then before the rematch, she should have been even worse off if she is a part of the rematch and been equally worse off otherwise. Therefore, the rematch in subsequent deviations is irrelevant to our purpose.

Second, allowing rematch for an initial deviation can increase its robustness, but it is tantamount to making a hidden assumption against our spirit. To be more concrete, suppose that $\nu \rhd_D \mu$, $\xi \rhd^*_D \mu$, and $\nu(i) = \xi(i)$ for all $i \in D$. That is, ν and ξ differ only in that the agents in $\mu(D)$ are left single at ν while they are rematched among themselves at ξ. Since $\xi(j) \succeq_j \nu(j)$ for $j \in \mu(D)$, there may exist ξ_1 such that $\xi_1 \rhd_{D_1} \nu$ for some D_1 but not $\xi_1 \rhd_{D_1'} \xi$ for any D_1'. This is why (D, ξ) may be more robust than (D, ν). However, this merely means that a deviation may become more robust if the deviators can enforce a particular way of rematch among those who they leave behind. Unless we presume such enforcement powers, thus, the alternative definitions based on $\rhd^*$ is against our spirit in this study, which is to measure the robustness of a deviation based on worst-case scenarios for the deviators.

2.2 Party Permutation and Stable Partition

In this subsection, we introduce the concepts of a *party permutation* and of a *stable partition* (Tan, 1991), which will be the basis of our analysis. A *permutation* is a bijection

from N to itself. A permutation σ divides N into a finite number of cycles and hence, induces a partition $\mathscr{P}(\sigma)$ of N. Namely, $\{a_1, \ldots, a_n\} \subseteq N$ is a member of $\mathscr{P}(\sigma)$ if $\sigma^m(a_1) = a_{m+1}$ for all $m = 1, 2, \ldots, n-1$ and $\sigma^n(a_1) = a_1$. Throughout the rest of the paper, given a permutation σ over N, we let π denote its inverse σ^{-1} and call a pair (a, b) of agents *adjacent* if $\sigma(a) = b$ or $\sigma(b) = a$. Taking σ and its inverse π as given, we sometimes refer to $\sigma(a)$ and $\pi(a)$ as, respectively, the successor and predecessor of agent a. It should be noted, however, that $a, \sigma(a), \pi(a)$ need not be distinct.[7]

We will focus on the following special class of permutations, which requires each $P \in \mathscr{P}(\sigma)$ to form a preference cycle:

Definition 1. A permutation $\sigma : N \to N$ is called a *semi-party permutation* if for each $P \in \mathscr{P}(\sigma)$, one of the following holds:

- $|P| = 1$,
- $|P| = 2$ and $\sigma(a) \succ_a a$ for each $a \in P$, or
- $|P| \geq 3$ and $\sigma(a) \succ_a \pi(a) \succ_a a$ for each $a \in P$. □

Given a semi-party permutation σ and hence its inverse π, an agent $a \in N$ is said to be *superior* for another agent $b \in N$ when $a \succ_b \pi(b)$. When a is not superior for b (i.e., when $\pi(b) \succeq_b a$), then a is said to be *inferior* for b.[8] With this terminology, we can define an even more special subclass of semi-party permutations as follows:

Definition 2. A semi-party permutation σ is called a *party permutation* if the following holds: for any $a, b \in N$, if a is superior for b, then b is inferior for a. □

When σ is a party permutation, $\mathscr{P}(\sigma)$ is called a *stable partition*, and each of its elements a *party*. Given a party permutation σ, for each $a \in N$, let $P(a)$ denote the party a belongs to; i.e., $a \in P(a) \in \mathscr{P}(\sigma)$. A party in a stable partition $\mathscr{P}(\sigma)$ is called *odd* (resp. *even*) if its cardinality is odd (resp. even). When it is a singleton, we call a party *solitary*. Note that when $\{a\} \in \mathscr{P}(\sigma)$ is a solitary party, $b \neq a$ is acceptable to a if and only if b is superior for a.

[7] If $\{a\} \in \mathscr{P}(\sigma)$, then $a = \sigma(a) = \pi(a)$. If $a \neq \sigma(a)$ but $\{a, \sigma(a)\} \in \mathscr{P}(\sigma)$, then $\sigma(a) = \pi(a)$.

[8] Here we slightly modify Tan's (1991) original definition: when $\{a, b\} \in \mathscr{P}(\sigma)$, a and b are inferior for each other according to our definition, whereas they are neither superior nor inferior for each other according to Tan's. As this does not alter the definition of party permutations below at all, Tan's (1991) results continue to hold with our definition.

Taking a party permutation σ and a regular matching μ as arbitrarily given, we define two symbols to denote subsets of the agents as functions of μ:

$$I^{\circ}_{\mu} := \{i \in N : \pi(i) \succ_i \mu(i)\}, \text{ and}$$
$$A_{\mu} := \{i \in N : i \neq \mu(i) \in \{\pi(i), \sigma(i)\}\}.$$

Strictly speaking, they depend on σ as well as μ, but it should cause no confusion because we will never consider multiple party permutations for a given problem. The former, I°_{μ}, denotes the set of those who are matched to a "strictly" inferior partner (or are single when $i \neq \pi(i)$) at μ. The latter, A_{μ}, is the set of those who are (not single and) matched to an adjacent partner at μ. Note that $i \in I^{\circ}_{\mu}$ is necessary for an agent i to prefer some inferior agent, including $\pi(i)$, to $\mu(i)$. Thus, I°_{μ} can be rephrased to be the set of those who may potentially deviate with an inferior agent. Note also that no solitary party intersects with I°_{μ} as long as μ is regular.[9] Since $\sigma(i) \succeq_i \pi(i)$ always holds for any i by the definition of a (semi-)party permutation, A_{μ} is disjoint from I°_{μ} for any matching μ. We will recurrently use these observations in our analysis.

While the definition of a party permutation might look complicated, Tan (1991) shows that at least one exists for any problem and that odd parties are uniquely identified across all party permutations even when multiple exist:[10]

Theorem (Tan, 1991)**.** *For any roommate problem $(N, \succ)$, at least one party permutation exists. If σ and σ' are both party permutations, then for any $P \subseteq N$ with $|P|$ being odd, $P \in \mathscr{P}(\sigma) \iff P \in \mathscr{P}(\sigma')$.*

For a problem $(N, \succ)$ with a party permutation σ, define $\#(N, \succ) \in \mathbb{N}$ by

$$\#(N, \succ) := \max\Big[\{|P| : P \in \mathscr{P}(\sigma) \text{ and } |P| \text{ is odd } \} \cup \{0\}\Big],$$

which is independent of the choice of σ thanks to the above theorem. Namely,

[9] By definition, $\pi(i) = i$ when $\{i\} \in \mathscr{P}(\sigma)$. In such a case, $\pi(i) \succ_i \mu(i)$ means $\mu(i)$ is unacceptable.

[10] Tan's (1991) original paper assumes preferences are symmetric in a certain sense, whereas we do not in this study. However, his results continue to hold true without the symmetry assumption. See Appendix D for details.

$\#(N, \succ)$ denotes the maximal size of odd parties in $(N, \succ)$ if there exists any, and it is set to zero otherwise. Roughly speaking, $\#(N, \succ)$ is the length of the longest preference cycle among those involving an odd number of agents, since each party is a preference cycle by definition. With this notation, the existence of a stable matching can be characterized as follows:[11]

Theorem (Tan, 1991)**.** *A stable matching exists in a roommate problem* $(N, \succ)$ *if and only if* $\#(N, \succ) \leq 1$.

Before closing this section, let us briefly explain why $\#(N, \succ)$ is critical for the existence of a stable matching, as it would also be helpful in understanding our results. To this end, fix a party permutation σ and suppose $\#(N, \succ) \leq 1$, which means that every party is either even or solitary. Then we can construct a stable matching μ as follows: For each a in an even party, match a to an agent "adjacent" to her with respect to σ, i.e., $\mu(a) \in \{\pi(a), \sigma(a)\}$; for each b in a solitary party, leave b single, i.e., $\mu(b) = b$. The following example illustrates the construction in a simple case.

Example 1. Let $N = \{a_1, \ldots, a_4, b_1\}$ and suppose that a party permutation is given by

$$\sigma = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & b_1 \\ a_2 & a_3 & a_4 & a_1 & b_1 \end{pmatrix},$$

where the right-hand side denotes $\sigma(a_1) = a_2$, $\sigma(a_2) = a_3$, and so on. Note that $\mathscr{P}(\sigma) = \{\{a_1, \ldots, a_4\}, \{b_1\}\}$. Then, the above construction leads to either $\mu_1 = \{\{a_1, a_2\}, \{a_3, a_4\}, \{b_1\}\}$ or $\mu_2 = \{\{a_2, a_3\}, \{a_4, a_1\}, \{b_1\}\}$. □

When $\#(N, \succ) \leq 1$, any matching μ constructed as above is stable. The point here is that $I^\circ_\mu = \{i \in N : \pi(i) \succ_i \mu(i)\}$ is empty. Recall that an agent can potentially deviate from μ with an inferior partner only if she belongs to I°_μ. For a pair of agents to form a deviation from μ when I°_μ is empty, therefore, each must be superior for the other. Since such a pair does not exist by the definition of a party permutation, thus,

[11] As we noted in footnote 10, the following theorem, as well as the previous one, holds without the symmetry of preferences. See Appendix D for details.

μ is stable if I°_{μ} is empty. Conversely, the main problem when $\#(N, \succ) > 1$ is that $I^{\circ}_{\mu'}$ cannot be empty for any matching μ', and this is essentially why a stable matching fails to exist.

3 Main Results

In this section, we present our main results. The first two results are on the existence of a SaRD matching up to depth $k = 1$ and 2:

Theorem 1. *For any roommate problem $(N, \succ)$ such that $\#(N, \succ) \leq 3$, there exists a matching that is SaRD up to depth* 1.

Proof. See Proposition 3 in Appendix A and Proposition 5 in Appendix B. ■

Theorem 2. *For any roommate problem $(N, \succ)$ such that $\#(N, \succ) \leq 5$, there exists a matching that is SaRD up to depth* 2.

Proof. See Proposition 3 in Appendix A and Proposition 5 in Appendix B. ■

We should make a couple of remarks on the conditions in the above two theorems. First, it is computationally feasible to directly check the conditions with respect to $\#(N, \succ)$: For any $(N, \succ)$, one can compute a party permutation in $O(|N|^2)$ time (Tan, 1991; Tan and Hsueh, 1995), and then, it is immediate to check $\#(N, \succ)$. Second, while they are stated based on $\#(N, \succ)$, we can interpret the conditions in terms of the primitives of the model. To do so, remember that by definition, a party is a preference cycle; i.e., if $\{a, \sigma(a), \ldots, \sigma^{2m}(a)\}$ is an odd party in $(N, \succ)$, they form a preference cycle in the sense that $\sigma^{i+1}(a) \succ_{\sigma^i(a)} \sigma^{i-1}(a)$ for all $i \in \mathbb{N}$. Therefore, the limit on the size of such preference cycles is a simple sufficient condition in terms of $\succ$ for its counterpart stated in terms of $\#(N, \succ)$, although it is not necessary.[12]

[12] A preference cycle is not necessarily a party, even though the converse is true. To see this, suppose that $N = \{a_1, \ldots, a_{2m+1}, b_1, \ldots, b_{2m+1}\}$. Assume that $\{a_1, \ldots, a_{2m+1}\}$ and $\{b_1, \ldots, b_{2m+1}\}$ form a preference cycle and that for each $i \in \{1, \ldots, 2m+1\}$, a_i and b_i are the best partner to each other. Then, the unique stable partition is $\{\{a_1, b_1\}, \ldots, \{a_{2m+1}, b_{2m+1}\}\}$ containing no odd party, in spite of the existence of the odd preference cycles.

It should also be noted that the conditions in Theorems 1–2 are tight among those which depend only on $\#(N, \succ)$.[13] That is, for each odd $n > 3$ (resp. odd $n > 5$), we can easily construct a problem $(N, \succ)$ such that $\#(N, \succ) = n$ and no matching is SaRD up to depth 1 (resp. depth 2) as follows. Recall that in the introduction, we have illustrated a problem with 5 agents and the simplest cyclic preferences such that no matching is SaRD up to depth 1. Combining this example with another odd cycle of n agents, we obtain a problem with $\#(N, \succ) = n$ and no SaRD matching up to depth 1. The case for depth 2 is analogous.

The above observation, along with Tan's theorem, might suggest that it becomes harder to guarantee the existence of a SaRD matching up to a fixed depth k as $\#(N, \succ)$ grows larger. In fact, perhaps surprisingly, this is not the case. We can establish a uniform bound for the robustness of possible deviations, which applies to *any* problem $(N, \succ)$, as follows:

Theorem 3. *For any roommate problem $(N, \succ)$, there exists a matching that is SaRD up to depth 3.*

Proof. See Proposition 4 in Appendix A and Proposition 5 in Appendix B. ■

To conclude this section, let us make a remark on the relation among the three theorems from a technical perspective. As we will see in the following section, the construction of a SaRD matching for Theorems 1–2 is substantially simpler than that for Theorem 3. Indeed, the algorithm we provide for Theorem 3 also produces a SaRD matching up to depth 1 (resp. depth 2) when $\#(N, \succ) \leq 3$ (resp. when $\#(N, \succ) \leq 5$). In this respect, thus, Theorems 1–2 could be seen a corollary of (the proof of) Theorem 3. The main message we can draw from Theorems 1–2 would rather be that the main difficulty for the existence of a SaRD matching lies in the cases of $\#(N, \succ) > 5$, which we address in Theorem 3.

[13] In Appendix D of the working paper version (Hirata et al., 2020), we further show that they are almost tight, in a certain sense, among those depending only on σ.

4 Proof Ideas

The full proofs of Theorems 1–3 are two-fold: First, we identify in Appendix A a set of sufficient conditions for a regular matching to be SaRD up to depth $k = 3$. When $\#(N, \succ) \leq 3$ and 5, respectively, the same set of conditions further ensures $k = 1$ and 2. Second, in Appendix B, we provide an algorithm to compute for any problem $(N, \succ)$ a regular matching satisfying all the conditions. As the full proofs are rather complicated, we relegate them to the appendices.

Through the rest of this section, we instead provide a sketch of the sufficiency part, abstracting away from the construction. We introduce our key conditions (Properties 1–5) one by one, explaining what roles they play in bounding the robustness of a possible deviation. To simplify our arguments, we assume for the rest of this section that the deviation is by a pair of agents, while we allow for deviations by more than two agents in the full proofs. More specifically, the rest of this section is organized as follows: Section 4.1 explains our basic strategy and introduces a critical condition that makes it work. The next two subsections describe what conditions will be useful and why, depending on if the deviating pair is non-adjacent (Section 4.2) or adjacent (Section 4.3). Section 4.4 briefly discusses what further complicates our full proofs, where the deviation is not restricted to be pairwise. Lastly, Section 4.5 describes the tensions among our conditions, which we need to take care of in the construction part of our full proofs.

Each of our conditions, Properties 1–5, refers to a party permutation σ, either directly or indirectly. It should be noted that the choice of the party permutation can be arbitrary, whereas it should be fixed across those conditions. For the rest of this section, thus, we arbitrarily fix a party permutation σ for a given problem $(N, \succ)$, and all of the properties should be read as referring to this fixed σ.

4.1 Basic Strategy

In establishing that a matching μ is SaRD up to depth k, our basic strategy is as follows. When $(\{a, b\}, \nu)$ is a deviation from a regular matching μ, the pair of the deviators should be matched at ν and at least one of them is not single at μ.[14] Thus, it is without loss of generality to assume $\mu(a) \neq a$. To bound the robustness of the deviation by $\{a, b\}$, we identify a (shortest) chain of subsequent deviations $(D_1, \nu_1), \ldots, (D_k, \nu_k)$ such that $a \notin D_1, \ldots, D_k$ and $b \in D_k$. Such a chain implies that a ends up being single at ν_k while she was matched to an acceptable partner at μ; thus, $(\{a, b\}, \nu)$ is not robust up to depth k.

Note that the above task becomes easier if b chooses a wider range of agents over a. In particular, it becomes much simpler when a is inferior for b than otherwise. In general, however, it is *not* without loss of generality to assume a is inferior for b, even though no pair of agents are mutually superior. This is because we have already assumed $\mu(a) \neq a$ and hence, we cannot freely swap their roles in the case of $\mu(b) = b$. In what follows, we thus restrict our attention to μ satisfying the following property, so as to make it without loss of generality to assume *both* $\mu(a) \neq a$ and that a is inferior for b, as stated in Lemma 1 below.

Property 1. *For any $a, b \in N$, if a is superior for b and $\mu(b) = b$, then $\mu(a) \succ_a b$.* □

Lemma 1. *Let μ be a regular matching satisfying Property 1, and suppose that $\nu \succ_E \mu$ where $E = \{a, b\}$ and $\nu(a) = b$. Then, at least one of the following holds: (i) a is an inferior agent for b, $\mu(a) \neq a$, and $b \in I^\circ_\mu$; and (ii) b is an inferior agent for a, $\mu(b) \neq b$, and $a \in I^\circ_\mu$, where $I^\circ_\mu = \{i \in N : \pi(i) \succ_i \mu(i)\}$.*

Proof. We first show at least one of the following holds: [i] a is inferior for b and $\mu(a) \neq a$, and [ii] b is inferior for a and $\mu(b) \neq b$. Note that we cannot have both a being superior for b and $\mu(b) = b$, by Property 1 and the assumption of $b \succ_a \mu(a)$. Symmetrically, we cannot have both b being superior for a and $\mu(a) = a$. Therefore, [i] and [ii] simultaneously fail only if a and b are mutually superior to each other or

[14] If a and b get better off by being single, μ is not individually rational. If both a and b are single at μ but are matched to each other at ν, then μ leaves a mutually-acceptable pair of singles.

if they are both single at μ. However, the former contradicts the definition of a party permutation, and the latter is incompatible with μ's regularity.

If a is inferior for b (i.e., $\pi(b) \succeq_b a$), then the assumption of $a = \nu(b) \succ_b \mu(b)$ implies $\pi(b) \succ_b \mu(b)$. That is, $b \in I^\circ_\mu$ if a is inferior for b. Symmetrically, $a \in I^\circ_\mu$ if b is inferior for a. Combined with the conclusion of the previous paragraph, these complete the proof. ■

4.2 Deviation by a Non-Adjacent Pair

Suppose that μ is a regular matching satisfying Property 1 and a pair $\{a, b\}$ deviates from it, resulting in a new matching ν. Also assume, without loss of generality by Lemma 1, $\mu(a) \neq a$ and that a is inferior for b. Recall that $P(b)$ should be non-solitary for the acceptable agent a to be inferior for b, and hence, $\pi(b) \neq b$. In this subsection, suppose further that a and b are *not* adjacent to each other. Since a is inferior for b (i.e., $\pi(b) \succeq_b a$), the non-adjacency implies that b strictly prefers $\pi(b) \neq a$ to a. This is helpful in bounding the robustness of $(\{a, b\}, \nu)$.

Specifically, if b and $\pi(b)$ agree to deviate with each other from ν, their deviation makes a single and thereby makes the original $(\{a, b\}, \nu)$ not robust up to depth 1. Although b should prefer $\pi(b)$ to $a = \nu(b)$ as seen above, the assumptions we made so far on μ are not enough to ensure $\pi(b)$ also prefers b to $\nu(\pi(b))$. Thus, we now require an additional property of the original matching μ:

Property 2. *For any $b \in I^\circ_\mu$, $\mu(\pi(b))$ is inferior for $\pi(b)$.* □

While it is a restriction for μ rather than ν, Property 2 ensures that $\pi(b)$ prefers b to $\nu(\pi(b))$, *under our assumption that the original deviation is pairwise*: Since the deviation is by $\{a, b\}$ and $\pi(b) \notin \{a, b\}$, we can conclude that $\pi(b)$ is not a part of the deviation from μ. Then, $\nu(\pi(b))$ should be equal to $\pi(b)$ if $\mu(\pi(b)) = a$ and to $\mu(\pi(b))$ otherwise. (Note that $\mu(\pi(b)) = b$ is impossible, because it is equivalent to $\pi(b) = \mu(b)$; if it holds, b should not have deviated with a, who is inferior for b.) In either case, Property 2 entails that $\pi(b)$ prefers b to $\nu(\pi(b))$, because b is by definition superior

for $\pi(b)$. That is, we can construct a deviation $(\{b, \pi(b)\}, \nu_1)$ from ν so that $\nu_1(a) = a$, whenever μ meets Property 2 in addition to regularity and Property 1. We summarize our conclusion in this subsection as follows:

Claim 1. *Let $\{a, b\}$ be a non-adjacent pair of agents and suppose that $(\{a, b\}, \nu)$ is a deviation from a regular matching μ satisfying Properties 1–2. Then, the deviation $(\{a, b\}, \nu)$ is not robust up to depth* 1.

4.3 Deviation by an Adjacent Pair

When the deviators are an adjacent pair $\{a, b\}$ with $a = \pi(b) = \nu(b)$, it is impossible to make a worse off by matching b with her predecessor $\pi(b)$. In this case, thus, we alternatively bound the robustness of the deviation from the other direction. Namely, we seek a (shortest) sequence of subsequent deviations that eventually matches b with her successor $\sigma(b)$, thereby bounding the robustness of the deviation by $\{a, b\}$. And it is here that how to match consecutive adjacent pairs is critical. The following example recaps the points we have illustrated in the introduction:

Example 2. Consider the following class of problems: $N := \{a_1, \ldots, a_n\}$ and each a_i's preference is such that $a_{i+1} \succ_{a_i} a_{i-1} \succ_{a_i} a_i$ and all the others are unacceptable, where the subscripts are in modulo n. Note that the party permutation is given by $\sigma(a_i) := a_{i+1}$ for all $a_i \in N$. First, suppose $n = 7$ and consider a matching

$$\mu_{(7)} := \Big\{\{a_1\}, \{a_2, a_3\}, \{a_4, a_5\}, \{a_6, a_7\}\Big\},$$

which is regular and meets Properties 1–2. Note that $\mu_{(7)}$ has a (unique) deviation $(\{a_7, a_1\}, \nu)$, which is by an adjacent pair. Starting from ν, where only a_6 is single, we can construct ν_1, ν_2, and ν_3, by sequentially matching $\{a_5, a_6\}$, $\{a_3, a_4\}$, and $\{a_1, a_2\}$ in this order. We then have $\nu_3 \rhd_{\{a_1,a_2\}} \nu_2 \rhd_{\{a_3,a_4\}} \nu_1 \rhd_{\{a_5,a_6\}} \nu$, where a_7 is single at ν_3. That is to say, the original deviation from $\mu_{(7)}$ by $\{a_7, a_1\}$ is not robust up to depth 3.

Next, suppose $n = 11$ and consider

$$\mu_{(11)} := \Big\{\{a_1\}, \{a_2, a_3\}, \{a_4\}, \{a_5, a_6\}, \{a_7\}, \{a_8, a_9\}, \{a_{10}, a_{11}\}\Big\},$$

which is regular and meets Properties 1–2. Note that it possesses three deviations by an adjacent pair: those by $\{a_{11}, a_1\}$, $\{a_3, a_4\}$, and $\{a_6, a_7\}$. Among them, the one by $\{a_{11}, a_1\}$ is not robust up to depth 2, because after they deviate, there are two subsequent deviations, first by $\{a_3, a_4\}$ and then by $\{a_1, a_2\}$ that leaves a_{11} single. For a similar reason, the deviation by $\{a_3, a_4\}$ is not robust up to depth 2, either. In contrast, the deviation by $\{a_6, a_7\}$ is robust up to depth 2, while not up to depth 3. This is because we need three subsequent deviations, by $\{a_{11}, a_1\}$, $\{a_9, a_{10}\}$, and $\{a_7, a_8\}$, so as to match a_7 to her successor a_8 leaving a_6 single. □

The key lesson we can extract from the above example is the following:

Observation. Let μ be a regular matching satisfying Properties 1–2 and $(\{a, b\}, \nu)$ a deviation from μ such that $a = \pi(b)$. These imply $\mu(a) \neq a$ by Lemma 1. Suppose further that there exists $k \in \mathbb{N}$ such that

- $\sigma^{2\kappa-1}(b)$ is matched to $\sigma^{2\kappa}(b)$ at ν for each κ with $1 \leq \kappa < k$, and
- $\sigma^{2k-1}(b)$ is either single or matched to an inferior partner at ν.

Then, there are k subsequent deviations by $\left\{\sigma^{2k-1}(b), \sigma^{2k-2}(b)\right\}, \ldots, \{\sigma(b), b\}$, which eventually leave a single. Thus, the deviation $(\{a, b\}, \nu)$ is not robust up to depth k if k satisfies the two conditions above. □

In order to bound the robustness of a deviation by adjacent $\{a, b\}$, thus, the key is to ensure that a sufficiently small k satisfies the two conditions in the above Observation. Before we proceed, we introduce the following property, which necessitates that all adjacent deviating pairs should belong to an odd party and thereby simplifies the subsequent arguments. Note that whenever a stable matching exists, a regular matching is stable if it satisfies this property, as we explained at the end of Section 2.2.

Property 3. *For each even party $P \in \mathscr{P}(\sigma)$, we have $P \subseteq A_\mu$, where $A_\mu = \{i \in N : i \neq \mu(i) \in \{\sigma(i), \pi(i)\}\}$ is the union of all adjacent pairs matched at μ.* □

The virtue of this property is in ensuring that no adjacent pair in an even party agrees to deviate from μ. It simply says that μ matches every even party P into $|P|/2$ adjacent pairs. Then, for any adjacent pair $\{a, b\}$ with $a = \pi(b)$ within an even party, either they are matched at μ or b is matched to $\sigma(b)$. Since b prefers $\sigma(b)$ to $a \equiv \pi(b)$ by the definition of a party permutation, no such $\{a, b\}$ can form a deviation.

Throughout the rest of this section, we assume that μ satisfies Properties 1–3 and consider the case of an adjacent pair (from an odd party) deviating from ν. And it is here that the sizes of odd parties become relevant. Specifically, we divide odd parties into the "large" and "small," depending on whether they contain more than seven agents. In each case, we will introduce an additional property to control the robustness of deviations by adjacent pairs.

4.3.1 "Small" Odd Parties

When the deviators are from a "small" odd party, it is easy to guarantee the existence of $k \leq 3$ in the above Observation. More specifically, the following simple property will turn out to be sufficient for our purpose:

Property 4. *For each odd party $P \in \mathscr{P}(\sigma)$ such that $|P| \leq 7$, we have $|P - A_\mu| = 1$, where $A_\mu = \{i \in N : i \neq \mu(i) \in \{\sigma(i), \pi(i)\}\}$ is the union of all adjacent pairs matched at μ.* □

This property requires μ to match as many adjacent pairs as possible within each odd party P with $|P| \leq 7$. Unlike Property 3 for even parties, however, there has to be a "residual" agent $b \in P$ who is matched to neither $\pi(b)$ nor $\sigma(b)$. Although μ may match b to an agent outside P, we cannot guarantee that $\mu(b)$ is preferred to $\pi(b)$; i.e., we cannot preclude the possibility of the deviation by $a := \pi(b)$ and b, who are adjacent to each other.

In light of the above Observation, however, it is not a big problem when $|P| \leq 7$. To see the point, suppose that μ is a regular matching satisfying Property 1–4, $(\{a, b\}, \nu)$ is a deviation from μ such that $a = \pi(b)$, and that $P \supsetneq \{a, b\}$ is an odd party of seven agents. Note that by the definition of a party permutation, b would not deviate with $a \equiv \pi(b)$ if she is matched to $\sigma(b)$ at μ. Property 4 thus requires $P - A_\mu = \{b\}$,

and hence, μ must match the other six agents into three adjacent pairs, $\{\sigma(b), \sigma^2(b)\}$, $\{\sigma^3(b), \sigma^4(b)\}$, and $\{\sigma^5(b), \sigma^6(b)\}$. Among these three, the first two adjacent pairs remain matched after the deviation by $\{a, b\}$, whereas $\sigma^5(b)$ is left single because $\sigma^6(b) = a$ by the assumption of $|P| = 7$. Therefore, we can apply the Observation with $k = 3$ and thereby conclude that the deviation by $\{a, b\}$ is not robust up to depth $k = 3$.

When $|P| = 3$ and 5, we can confirm that a deviation by an adjacent pair $\{a, b\} \subsetneq P$ is not robust up to depth 1 and 2, respectively, following almost the same arguments as in the previous paragraph. This is essentially why we have a smaller k in Theorems 1–2 for the cases of $\#(N, \succ) = 3$ and 5. We can then summarize our arguments in this subsection as follows:

Claim 2. *Let $\{a, b\}$ be an adjacent pair of agents in an odd party P with $|P| \leq 7$ and suppose that $(\{a, b\}, \nu)$ is a deviation from a regular matching μ safisfying Properties 1–4. Then, the deviation $(\{a, b\}, \nu)$ is not robust up to depth 3. If $|P| = 3$ and 5, respectively, it is not robust up to depth 1 and 2.*

4.3.2 "Large" Odd Parties

What remains to be considered is a deviation $(\{a, b\}, \nu)$ when $\{a, b\}$ is an adjacent pair from an odd party of nine or more agents. To cap its robustness by $k = 3$, our Observation above would suggest that we should avoid the case where both

- two adjacent pairs $\{\sigma(b), \sigma^2(b)\}$ and $\{\sigma^3(b), \sigma^4(b)\}$ are matched at ν, and
- $\sigma^5(b)$ is matched to some superior partner at ν.

Furthermore, since we are restricting our attention to a pairwise deviation, it is without loss of generality to replace ν in the previous sentence with μ: As $\{a, b\}$ is the only pair that is matched at ν but not at μ, the three pairs, $\{\sigma(b), \sigma^2(b)\}$, $\{\sigma^3(b), \sigma^4(b)\}$, and $\{\sigma^5(b), \nu(\sigma^5(b))\}$ are matched at ν only if they are so at μ. Thus, what we need μ to satisfy, in addition to the other properties, is the following:

Property 5. *For each odd party $P \in \mathscr{P}(\sigma)$ such that $|P| > 7$, μ satisfies the following: If $b \in P \cap I^\circ_\mu$, $\mu(\sigma(b)) = \sigma^2(b)$, and $\mu(\sigma^3(b)) = \sigma^4(b)$, then we have both $\sigma^5(b) \in I^\circ_\mu$ and*

$\sigma^6(b) \notin I^\circ_\mu$, *where* $I^\circ_\mu = \{i \in N : \pi(i) \succ_i \mu(i)\}$ *is the set of those who potentially deviate with an inferior agent.* □

The aim of Property 5 is simply to circumvent the difficulty we specified above. If we have all $\mu(\sigma(b)) = \sigma^2(b)$, $\mu\left(\sigma^3(b)\right) = \sigma^4(b)$, and $\sigma^5(b) \in I^\circ_\mu$, then we can conclude through the Observation that the deviation by $a = \pi(b)$ and b is not robust up to depth 3. The second conclusion, $\sigma^6(b) \notin I^\circ_\mu$, is not necessary as long as we focus on pairwise deviations. In the general case, however, we need to preclude the possibility that $\sigma^5(b)$ and $\sigma^6(b)$ also deviate and are newly matched to each other at ν. The purpose of $\sigma^6(b) \notin I^\circ_\mu$ is to guarantee she prefers her partner at μ to $\sigma^5(b)$.

Now let us turn to why Property 5 is enough, even though it imposes no restriction unless two adjacent pairs, $\{\sigma(b), \sigma^2(b)\}$ and $\{\sigma^3(b), \sigma^4(b)\}$, are matched at μ. To fix the idea, let $\{a, b\}$ with $a = \pi(b)$ be an adjacent pair in a large odd party, and suppose that they deviate from a regular matching μ satisfying Properties 1–5. Also assume $\mu(\sigma(b)) = \sigma^2(b)$ and $\mu\left(\sigma^3(b)\right) \neq \sigma^4(b)$. Then, $\sigma(b)$ and $\sigma^2(b)$ should remain matched to each other at ν, because the largeness implies they are disjoint from $\{a, b\}$. If $\nu(\sigma^3(b))$ is inferior for $\sigma^3(b)$, the argument is simple. We can form subsequent deviations by $\{\sigma^2(b), \sigma^3(b)\}$ and $\{b, \sigma(b)\}$ as in the Observation; thus, the original deviation by $\{a, b\}$ is not robust up to depth 2. When $\nu(\sigma^3(b))$ is superior for $\sigma^3(b)$, our Observation is not directly applicable as $\sigma^3(b)$ does not agree to deviate with $\sigma^2(b)$ from ν. Nonetheless, we can still apply it indirectly, with the help of our technique in Section 4.2. The following example illustrates this point:

Example 3. Let $N := \{a_1, a_2, \ldots, a_9\}$ and $\sigma : N \to N$ be such that $\sigma(a_i) := a_{i+1}$ for all $a_i \in N$, where the subscripts are in modulo 9. Also define $\succ$ as follows:

- For each $i \notin \{6, 9\}$, $\succ_{a_i}$ is given by $a_{i+1} \succ_{a_i} a_{i-1} \succ_{a_i} a_i$, and
- $\succ_{a_6}$ and $\succ_{a_9}$ are given by $a_9 \succ_{a_6} a_7 \succ_{a_6} a_5 \succ_{a_6} a_6$ and $a_1 \succ_{a_9} a_8 \succ_{a_9} a_6 \succ_{a_9} a_9$,

where all the unlisted agents are unacceptable. Compared to Example 2, this problem differs only in that a_6 and a_9 are mutually acceptable. Yet, σ remains to be the unique

party permutation for $(N, \succ)$. Now consider two matchings,

$$\mu := \Big\{\{a_1, a_2\}, \{a_3\}, \{a_4, a_5\}, \{a_6, a_9\}, \{a_7, a_8\}\Big\}, \text{ and}$$
$$\nu := \Big\{\{a_1\}, \{a_2, a_3\}, \{a_4, a_5\}, \{a_6, a_9\}, \{a_7, a_8\}\Big\}.$$

Note that μ is a regular matching satisfying all the properties above and that $(\{a_2, a_3\}, \nu)$ is a deviation from μ. Since $a_6 \equiv \sigma^3(a_3)$ is matched to a_9, who is superior for a_6, $\{a_5, a_6\}$ does not constitute a deviation directly from ν. However, a_6 becomes single and agrees to deviate with a_5 once $\{a_8, a_9\}$ deviates from ν. Therefore, we can make a_2 single after three subsequent deviations, respectively by $\{a_8, a_9\}$, $\{a_5, a_6\}$, and $\{a_3, a_4\}$ in this order. That is, the deviation $(\{a_2, a_3\}, \nu)$ from μ is not robust up to depth 3. □

The key in Example 3 is that we can form a deviation from ν by matching a_9 with her predecessor a_8 and thereby make $a_6 \equiv \sigma^3(a_3)$ single. In fact, this is not a mere coincidence but a consequence of our properties. Let us return to the general case and suppose again that $(\{a, b\}, \nu)$ with $a = \pi(b)$ is a deviation from a regular μ satisfying Properties 1–5. Also assume $\mu(\sigma(b)) = \sigma^2(b)$, $\mu\left(\sigma^3(b)\right) \neq \sigma^4(b)$, and that $\nu\left(\sigma^3(b)\right)$ is superior for $\sigma^3(b)$. Since $\sigma^3(b)$ does not belong to the set of the deviators, $\{a, b\}$, these assumptions are compatible only if $\sigma^3(b)$ is matched to the same partner, say c, at μ and ν. Notice that $\sigma^3(b)$ should be inferior for c, as c is superior for $\sigma^3(b)$ by assumption and no pair of agents are mutually superior by definition. It then follows that $c \in I^\circ_\mu$, since $\sigma^3(b) \neq \pi(c)$ by the assumption of $\mu\left(\sigma^3(b)\right) \neq \sigma^4(b)$. This in turn implies via Property 2 that $\pi(c)$ is not matched to a superior partner at μ. Actually, the same should be true also at ν, since $\pi(c)$ cannot be either a or b.[15] Therefore, c and $\pi(c)$ should necessarily agree to form a deviation from ν, after which $\sigma^3(b)$ becomes single and we can apply the logic of the above Observation.

[15]Formally, we can confirm $\pi(c) \notin \{a, b\}$ as follows: If she were a, her successor $c \equiv \mu\left(\sigma^3(b)\right)$ should be b. If so, however, $\sigma^3(b)$ must be single at ν, which would be a contradiction. If $\pi(c)$ were equal to b, $\sigma(b) = c \in I^\circ_\mu$ should follow, but this would contradict the assumption of $\mu(\sigma(b)) = \sigma^2(b)$.

What we have seen so far is that the deviation by adjacent $\{a, b\}$ with $a = \pi(b)$ is not robust up to (at most) depth 3 when $\mu(\sigma(b)) = \sigma^2(b)$ and $\mu(\sigma^3(b)) \neq \sigma^4(b)$. Following the same line of arguments, we can also confirm that it is not robust up to (at most) depth 2 when we instead have $\mu(\sigma(b)) \neq \sigma^2(b)$: If $\mu(\sigma(b))$ is inferior for $\sigma(b)$, we match $\{b, \sigma(b)\}$ so as to make a single. If $\mu(\sigma(b)) = \nu(\sigma(b))$ is superior, we first make $\sigma(b)$ single by matching $\mu(\sigma(b))$ with her predecessor, and then, we match $\{b, \sigma(b)\}$ to leave a single. To summarize this subsection, we have confirmed the following:

Claim 3. *Let $\{a, b\}$ be an adjacent pair of agents in an odd party P with $|P| > 7$ and suppose that $(\{a, b\}, \nu)$ is a deviation from a regular matching μ safisfying Properties 1–5. Then, the deviation $(\{a, b\}, \nu)$ is not robust up to depth 3.*

4.4 Complications with Non-Pairwise Deviations

In the full proofs of Theorems 1–3, we allow for coalitional deviations (i.e., (D, ν) with $|D| > 2$), yet our basic strategy remains the same as in the case of pairwise deviations: We pick a pair $\{a, b\} \subseteq D$ such that $\nu(a) = b$, $\mu(a) \neq a$, and a is inferior for b; and we find a shortest sequence of subsequent deviations at the end of which a becomes single. More specifically, we divide the possible deviations into two classes based on a criterion that reduces to whether $D = \{a, b\}$ is adjacent or not when the deviation is pairwise. In one of the two cases, $\{b, \pi(b)\}$ forms a deviation directly from ν, while in the other, we identify a sequence of three or less subsequent deviations that eventually involve $\{b, \sigma(b)\}$. In those respects, the full proofs for the sufficient conditions are parallel to our arguments above for pairwise deviations.

The difficulty in such generalizations largely lies in the gap between μ and ν. For instance, suppose that (D, ν) is a deviation from a regular μ satisfying all the properties and that $\{a, b\} \subseteq D$ is a non-adjacent pair satisfying the suppositions in the previous paragraph. If the deviation is pairwise, i.e., if $D = \{a, b\}$, it ensures that $\pi(b)$ is not a deviator and hence, $\nu(\pi(b))$ is no better than $\mu(\pi(b))$ for $\pi(b)$.[16] This helps to connect

[16]Note that $\pi(b) = b$ is impossible; if so, a should be unacceptable for b since the former is inferior

the restriction imposed by Property 2 on μ to our arguments of ν, as we did in Section 4.2. When $D \supsetneq \{a, b\}$, in contrast, we need to take into account the possibility of $\pi(b) \in D$. In particular, it is possible that $\pi(b)$ is matched to a superior partner at ν, even though Property 2 guarantees that $\mu(\pi(b))$ is inferior. Generally speaking, allowing coalitional deviations broadens the possible gap between μ and ν, and as a consequence, we have to scrutinize a larger number of subcases. To partially fill in the broader gap, we will introduce another condition, which is unnecessary in the case of pairwise deviations, in addition to Properties 1–5.

4.5 Tensions among the Sufficient Conditions

Before concluding this section, we briefly explain the difficulty in the construction part of our proofs of Theorems 1–3, from which we have hitherto abstracted away. As we emphasized in the introduction, one of the keys in our construction is how to match (consecutive) adjacent pairs, and more formally, it is embodied in Properties 3–5. Then, one might expect a two-step procedure that (i) fixes all the adjacent pairs to be matched first and (ii) matches non-adjacent pairs among the remaining afterwards. In fact, however, our algorithm in Appendix B needs to be more complicated so as to circumvent the difficulty illustrated in the following example.

Example 4. Let $N := \{a_1, a_2, \ldots, a_{11}\}$ and $\sigma : N \to N$ be such that $\sigma(a_i) := a_{i+1}$ for all $a_i \in N$, where the subscripts are in modulo 11. Also define $\succ$ as follows:

- For each $i \notin \{3, 6, 11\}$, $\succ_{a_i}$ is given by $a_{i+1} \succ_{a_i} a_{i-1} \succ_{a_i} a_i$, and
- the preferences for the other three agents are given by

$$\succ_{a_3} : a_4 \succ_{a_3} a_2 \succ_{a_3} a_{11} \succ_{a_3} a_6 \succ_{a_3} a_3,$$
$$\succ_{a_6} : a_7 \succ_{a_6} a_5 \succ_{a_6} a_3 \succ_{a_6} a_6, \qquad \text{and}$$
$$\succ_{a_{11}} : a_3 \succ_{a_{11}} a_1 \succ_{a_{11}} a_{10} \succ_{a_{11}} a_{11},$$

where all the unlisted agents are unacceptable. Note that σ is the unique party permu-

for the latter.

tation for $(N, \succ)$. In this problem, no regular matching μ both satisfies Properties 1–5 and matches the following four adjacent pairs: $\{a_1, a_2\}$, $\{a_4, a_5\}$, $\{a_7, a_8\}$, and $\{a_9, a_{10}\}$. By regularity, such μ would need to match either $\{a_3, a_6\}$ or $\{a_3, a_{11}\}$ among the remaining three agents. On the one hand, if μ matches $\{a_3, a_6\}$ leaving a_{11} single, then it violates Property 1, since a_3 is superior for a_{11} and prefers a_{11} to a_6. On the other hand, if μ matches $\{a_3, a_{11}\}$ leaving a_6 single, then $a_6 \in I^\circ_\mu$ but $\sigma^5(a_6) \equiv a_{11} \notin I^\circ_\mu$; i.e., Property 5 fails. Nevertheless, there is a regular matching that satisfies all the properties. An example is $\mu^* := \{\{a_1\}\{a_2, a_3\}, \{a_4\}, \{a_5, a_6\}, \{a_7\}, \{a_8, a_9\}, \{a_{10}, a_{11}\}\}$, which matches different four adjacent pairs from above. □

In matching adjacent pairs, we need to carefully keep the compatibility of our sufficient conditions. In the above example, the two patterns of four adjacent pairs, $\{\{a_1, a_2\}, \{a_4, a_5\}, \{a_7, a_8\}, \{a_9, a_{10}\}\}$ and $\{\{a_2, a_3\}, \{a_5, a_6\}, \{a_8, a_9\}, \{a_{10}, a_{11}\}\}$, are symmetric up to rotation.[17] Namely, we cannot differentiate them from the information contained in the party permutation σ, while only one of them makes Properties 1 and 5 compatible. In other words, the information contained in σ is insufficient to match adjacent pairs avoiding the tension among the desired conditions. In our construction presented in Appendix B, thus, we carefully match non-adjacent pairs based on more detailed information of preferences, prior to fixing all the adjacent pairs to be matched.

5 Relation to Other Solution Concepts

5.1 Bargaining Set

Particularly with depth $k = 1$, our definition of SaRD matchings might remind readers of the bargaining set in cooperative game theory. In our definition, a deviation is robust if there is no further deviation that makes an original deviator worse off, and a matching is SaRD if there is no robust deviation. In cooperative games, an objection

[17] More specifically, σ maps $\{a_1, a_2\}$, $\{a_4, a_5\}$, $\{a_7, a_8\}$, and $\{a_9, a_{10}\}$ to $\{a_2, a_3\}$, $\{a_5, a_6\}$,$\{a_8, a_9\}$, and $\{a_{10}, a_{11}\}$, respectively.

is justified if it has no counterobjection, and an imputation is in the bargaining set if it has no justified objection. By definitions, our SaRD is a weakening of stability, whereas the bargaining set is a superset of the core, which is equivalent to the set of stable matchings in matching models. Given those similarities, it would be natural to ask how the SaRD matchings relate to the bargaining set.

To closely compare the two concepts, let us formally define Zhou's (1994) bargaining set in our setup.[18] An *objection* against a matching μ is a deviation (D, ν) from μ. A *counterobjection* against an objection (D, ν) is a pair (D', ν') such that

- $D' - D, D - D', D \cap D'$ are all non-empty,
- for all $i \in D'$, $\nu'(i) \neq \mu(i)$ implies $\nu'(i) \in D'$, and
- $\nu'(a) \succeq_a \mu(a)$ for all $a \in D' - D$ and $\nu'(b) \succeq_b \nu(b)$ for all $b \in D \cap D'$.

The similarity between our SaRD matchings and the bargaining set lies in that both require the existence of some (D', ν') that precludes a deviation (D, ν) from (or, an objection against) μ.

The key distinction, however, exists in the reference points with which (D', ν') is compared. On the one hand, in our definition of SaRD matchings, (D', ν') is a deviation from ν and hence, all the agents in D' compare ν and ν'. On the other hand, in the definition of the bargaining set, the agents in $D' - D$ compare μ and ν'.[19] Consequently, the (set of) SaRD matchings and bargaining set are logically independent as we show by examples below:

Example 5 (The SaRD matchings are not included in the Bargaining Set)**.** Let $N := \{a_1, a_2, a_3\}$ and $\succ_{a_i}$ be such that $a_{i+1} \succ_{a_i} a_{i-1} \succ_{a_i} a_i$ for each $a_i \in N$, where the subscripts are in modulo 3. In this problem, it is easy to check that $\mu = \{\{a_1, a_2\}, \{a_3\}\}$ is SaRD up to depth 1: $(D, \nu) = (\{a_2, a_3\}, \{\{a_1\}, \{a_2, a_3\}\})$ is the only deviation from μ, and this is not robust up to depth 1 as $\nu' \rhd_{\{a_1, a_3\}} \nu$ and agent $a_2 \in D$ gets strictly worse off at ν' than at μ, where $\nu' = \{\{a_1, a_3\}, \{a_2\}\}$. However, this μ is not in the bargaining

[18] For a more standard definition and characterization of Zhou's bargaining set in matching problems, see Klijn and Massó (2003) and Atay et al. (2019). Our definition below is equivalent to theirs.

[19] While there exist a number of different definitions of a bargaining set (e.g., Aumann and Maschler, 1964; Mas-Colell, 1989) all of those we are aware of commonly require that a counterobjecction to be an objection against the original allocation (i.e., contain some comparison between ν' and μ). Hence, our point here should apply to the general concept of bargaining sets, not only to the one by Zhou (1994).

set, because $\nu'(a_1) = a_3 \not\succeq_{a_1} a_2 = \mu(a_1)$ and hence, $(\{a_1, a_3\}, \nu')$ is not qualified to be a counterobjection against $(\{a_2, a_3\}, \nu)$. □

Example 6 (The SaRD matchings do not include the Bargaining Set)**.** Let $N := \{m_1, m_2, w_1, w_2, w_3\}$ and $\succ$ be such that

$$w_1 \succ_{m_1} w_2 \succ_{m_1} w_3 \succ_{m_1} m_1 \succ_{m_1} m_2, \quad w_2 \succ_{m_2} w_1 \succ_{m_2} w_3 \succ_{m_2} m_2 \succ_{m_2} m_1,$$
$$m_2 \succ_{w_1} m_1 \succ_{w_1} w_1 \succ_{w_1} w_2 \succ_{w_1} w_3, \quad m_1 \succ_{w_2} m_2 \succ_{w_2} w_2 \succ_{w_2} w_1 \succ_{w_2} w_3, \quad \text{and}$$
$$w_3 \succ_{w_3} m_1 \succ_{w_3} m_2 \succ_{w_3} w_1 \succ_{w_3} w_2.$$

This problem is a marriage problem with $M = \{m_1, m_2\}$ and $W = \{w_1, w_2, w_3\}$ being the sets of men and women. It is easy to verify that $\mu = \{\{m_1\}, \{m_2\}, \{w_1\}, \{w_2\}, \{w_3\}\}$ is in Zhou's (1994) bargaining set.[20] However, Proposition 1 implies that this μ is not SaRD up to any depth k, as it leaves mutually-acceptable pairs of singles. □

5.2 Farsightedly Stable Set

Our concept of SaRD might also remind readers of the farsighted stable set à la Harsanyi (1974), as condition $(*)$ in the definition of robust deviations on page 8 might appear to resemble indirect dominance in the definition of stable sets.[21] In relation to the farsighted stable set, we make three remarks here: First, the stable set is a set solution whereas ours is a pointwise (i.e., matching-wise) concept. Moreover, Klaus et al. (2011) establish in the roommate problem that a singleton is a farsighted stable set if and only if its unique element is a stable matching.[22] Therefore, although focusing on singletons can be a possible way to compare a set solution with a point solution, such an approach is not helpful to overcome the general non-existence of a stable matching in our setup.

[20] In this environment, Zhou's bargaining set is the set of all matchings that are both weakly stable and weakly Pareto efficient (Klijn and Massó, 2003; Atay et al., 2019). A matching μ is *weakly stable* if for any pairwise deviation $(\{a, b\}, \nu)$ from it, there exists either (i) a' such that $a' \succ_b a$ and $b \succ_{a'} \mu(a')$ or (ii) b' such that $b' \succ_a b$ and $a \succ_{b'} \mu(b')$.

[21] For the formal definitions of farsighted stable sets, see also Chwe (1994) and Ray and Vohra (2015).

[22] See also Ehlers (2007) and Mauleon et al. (2011) for related results in the marriage problem.

Second, it should be noted that we can obtain exactly the same set of results even if we introduce "farsightedness" into our definitions. Specifically, let us say that a deviation (D, ν) from a matching μ is *farsightedly-robust* up to depth k, if $\nu_\kappa \succeq_D \mu$ for any sequence of deviations $(D_1, \nu_1), \ldots, (D_\kappa, \nu_\kappa)$ with $\kappa \leq k$ that satisfies $\nu_\kappa \succeq_{D_\lambda} \nu_{\lambda-1}$ for all $\lambda \in \{1, \ldots, \kappa\}$ (with $\nu_0 := \nu$) in addition to the original requirement $(*)$. Such a definition could be seen "farsighted" as the agents in D_λ also compare the final outcome ν_κ with the situation before they deviate, $\nu_{\lambda-1}$, while they only compare ν_λ and $\nu_{\lambda-1}$ in our original definitions. We can also define *farsightedly-SaRD* matchings up to depth k based on this farsighted-robustness. Notice that taking a depth k as fixed, farsighted-robustness is implied by and thus weaker than the original robustness, since the former considers only a subset of subsequent deviations that the latter does. Consequently, the farsighted-SaRD is stronger than the original SaRD. However, our existence results continue to hold with those alternative definitions: This is because whenever we consider a sequence of deviations, no agent deviates more than once along the sequence; that is, when we conclude that an original deviation is not robust up to depth k, it is also shown to be not farsightedly-robust up to depth k.

Lastly, several recent studies (Ray and Vohra, 2019; Dutta and Vohra, 2017; Dutta and Vartiainen, 2020) propose new concepts of farsighted stable sets that incorporate dynamic consistency à la subgame perfection. Among them, the one by Dutta and Vartiainen (2020), history-dependent rational-expectation farsighted stable set (HREFS), is particularly relevant to the roommate problem, as it always exists in any finite game. In Appendix G of the working paper version (Hirata et al., 2020), however, we provide a class of examples where the set of all individually rational matchings forms an HREFS. At least without further refinements, thus, the HREFS may be too inclusive and not necessarily useful in the context of the roommate problem.

5.3 P-stable Matching

Inarra et al. (2008) propose the following concept of $\mathscr{P}$-stable matchings, which is closely related to absorbing sets and stochastic stability in the roommate problem

(Iñarra et al., 2013; Klaus et al., 2011):

Definition 3. Given a stable partition $\mathscr{P} = \mathscr{P}(\sigma)$, a matching μ is said to be $\mathscr{P}$-stable if $|P - A_\mu| \leq 1$ for all $P \in \mathscr{P}$ and $\mu(b) = b$ for all $b \notin A_\mu$. □

Note that $\mathscr{P}$-stable matchings differ in two respects from the SaRD matchings we construct in this paper. First, a $\mathscr{P}$-stable matching always matches as many adjacent pairs as possible, whereas we do not in our construction. As a result, even when it is SaRD up to some depth, its depth is generally greater than what we construct. Second, it may not be regular, since it does not match any non-adjacent pair. By Proposition 1, thus, it may not be SaRD up to any depth. However, we can always convert a $\mathscr{P}$-stable matching into a SaRD matching by eliminating mutually-acceptable pair of singles in a particular way, and its depth can be characterized as follows:

Proposition 2. *Suppose that $\#(N, \succ) = 2k + 1$ for some $k \in \mathbb{N}$. Then, for any $\mathscr{P}$-stable matching μ', there exists a matching μ that is SaRD up to depth k and "includes" μ' in the sense that $\mu'(a) = b \neq a$ implies $\mu(a) = b$ for all $a, b \in N$.*

Proof. See Appendix C. ■

Combined with the results of Inarra et al. (2008, Thoerem 1) and Iñarra et al. (2013, Theorem 1), this proposition implies that when $\#(N, \succ) = 2k + 1$, the set of all SaRD matchings up to depth k is reachable by random paths of myopic deviations. More specifically, for any matching μ in $(N, \succ)$ with $\#(N, \succ) = 2k + 1$, there exist some $(D_1, \nu_1), \ldots, (D_n, \nu_n)$ such that $\nu_n \rhd_{D_n} \nu_{n-1} \ldots \nu_1 \rhd_{D_1} \mu$ and ν_n is SaRD up to depth k.[23] It should be also noted, however, that the same is not always true for SaRD matchings up to depth 3 when $\#(N, \succ) > 7$, since the matchings we construct for depth 3 may not include a $\mathscr{P}$-stable matching. For instance, suppose that $N = \{a_1, \ldots, a_9\}$, and that for each $i \in N$, let $\succ_{a_i}$ be such that $a_{i+1} \succ_{a_i} a_{i-1}$ and all the other agents are unacceptable, where the subscripts are in modulo 9. Fix a $\mathscr{P}$-stable matching, say

[23] In the more general coalition formation game, Barberà and Gerber (2003, Theorem 2.1) show that the set of durable coalition structures are reachable by myopic deviations. Since durability coincides with SaRD up to a sufficiently large k, the above claim can be seen as a refinement of their theorem in the special case of roommate problems.

$\mu = \{\{a_1, a_2\}, \{a_3, a_4\}, \{a_5, a_6\}, \{a_7, a_8\}, \{a_9\}\}$, as the initial matching. Then after *any* sequence $(D_1, \nu_1), \ldots, (D_n, \nu_n)$ of myopic deviations such that $\nu_n \rhd_{D_n} \ldots \nu_1 \rhd_{D_1} \mu$, the resulting matching ν_n is one of the $\mathscr{P}$-stable matchings, which are SaRD up to depth 4 but not depth 3.

Acknowledgments

We thank an Advisory Editor and two referees for extremely careful reading and helpful suggestions. We are particularly grateful to Benjamin Balzer for detailed comments. We also 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 helpful comments. Shinpei Noguchi and Yusuke Yamaguchi provided excellent research assistance. Hirata and Kasuya gratefully acknowledge financial support from JSPS KAKENHI (#16K17081 and #20K13452).

References

ABDULKADIROĞLU, A. AND T. SÖNMEZ (2003): "School Choice: A Mechanism Design Approach," *American Economic Review*, 93, 729–747.

ABRAHAM, D. J., P. BIRÓ, AND F. MANLOVE DAVID (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 (2019): "A Bargaining Set for Roommate Problems," *mimeo*.

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.

BOGOMOLNAIA, A. AND M. O. JACKSON (2002): "The Stability of Hedonic Coalition Structures," *Games and Economic Behavior*, 38, 201–230.

CHWE, M. S.-Y. (1994): "Farsighted Coalitional Stability," *Journal of Economic Theory*, 63, 299–325.

DUTTA, B. AND H. VARTIAINEN (2020): "Coalition Formation and History Dependence," *Theoretical Economics*, 15, 159–197.

DUTTA, B. AND R. VOHRA (2017): "Rational Expectations and Farsighted Stability," *Theoretical Economics*, 12, 1191–1227.

EHLERS, L. (2007): "Von Neuman-Morgenstern Stable Sets in Matching Problems," *Journal of Economic Theory*, 134, 537–547.

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

GUSFIELD, D. AND R. W. IRVING (1989): *The Stable Marriage Problem: Structure and Algorithms*, MIT Press.

HARSANYI, J. C. (1974): "An Equilibrium-Point Interpretation of Stable Sets and a Proposed Alternative Definition," *Management Science*, 20, 1472–1495.

HIRATA, D., Y. KASUYA, AND K. TOMOEDA (2020): "Stability against Robust Deviations in the Roommate Problem," *mimeo*.

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.

JACKSON, M. O. (2008): *Social and Economic Networks*, Princeton University Press.

KADAM, S. V. AND M. H. KOTOWSKI (2018): "Multi-Period Maching," *International Economic Review 59*, 1927–1947.

KLAUS, B., F. KLIJN, AND M. WALZL (2010): "Stochastic Stability for Roommate Markets," *Journal of Economic Theory*, 145, 2218–2240.

——— (2011): "Farsighted Stability for Roommate Markets," *Journal of Public Economic Theory 13*, 921–933.

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

KOTOWSKI, M. H. (2015): "A Note on Stability in One-to-One, Multi-Period Matching Markets," *mimeo*.

KURINO, M. (2019): "Credibility, Efficiency, and Stability: A Theory of Dynamic Matching Markets," *Japanese Economic*, Review, forthcoming.

MAS-COLELL, A. (1989): "An Equivalence Theorem for a Bargaining Set," *Journal of Mathematical Economics*, 18, 129–139.

MAULEON, A., V. J. VANNETELBOSCH, AND W. VERGOTE (2011): "Von Neumann-Morgenstern Farsightedly Stable Sets in Two-Sided Matching," *Theoretical Economics*, 6, 499–521.

PITTEL, B. G. AND R. W. IRVING (1994): "An Upper Bound for the Solvability Probability of a Random Stable Roommates Instance," *Random Structures and Algorithms*, 5, 465–486.

RAY, D. AND R. VOHRA (2015): "The Farsighted Stable Set," *Econometrica*, 83, 977–1011.

——— (2019): "Maximality in the Farsighted Stable Set," *Econometrica 87*, 1763–1779.

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.

TAN, J. J. M. AND Y.-C. HSUEH (1995): "A Generalization of the Stable Matching Problem," *Discrete Applied Mathematics*, 59, 87–102.

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.

A Conditions for the Existence in Theorems 1–3

In this appendix, we identify a set of sufficient conditions for a regular matching to be SaRD up to depth 3, which also suffice for depth 1 and 2 when $\#(N, \succ)$ is 3 and 5. When we restricted deviations to be pairwise in Section 4, we have seen the following five properties constitute such a set of conditions. They continue to be a part of the conditions for the general case, although we will add one last property below.

Properties for a matching to be SaRD up to depth 3 (restated). Take a problem $(N, \succ)$, a party permutation σ, and a matching μ as given and fixed. Letting $I^{\circ}_{\mu} = \{i \in N : \pi(i) \succ_i \mu(i)\}$ and $A_{\mu} = \{i \in N : i \neq \mu(i) \in \{\sigma(i), \pi(i)\}\}$, the five properties are given as follows:

1. For any $a, b \in N$, if a is superior for b and $\mu(b) = b$, then $\mu(a) \succ_a b$.
2. For any $b \in I^{\circ}_{\mu}$, $\mu(\pi(b))$ is inferior for $\pi(b)$.

3. For each even party $P \in \mathscr{P}(\sigma)$, we have $P \subseteq A_\mu$.
4. For each odd party $P \in \mathscr{P}(\sigma)$ such that $|P| \leq 7$, we have $|P - A_\mu| = 1$.
5. For each odd party $P \in \mathscr{P}(\sigma)$ such that $|P| > 7$, the following is satisfied: For any $b \in P \cap I_\mu^\circ$ such that $\mu(\sigma(b)) = \sigma^2(b)$ and $\mu(\sigma^3(b)) = \sigma^4(b)$, we have both $\sigma^5(b) \in I_\mu^\circ$ and $\sigma^6(b) \notin I_\mu^\circ$. □

Remember that in Section 4, we divided the case depending on whether the pair of deviators are adjacent or not. When the deviation is by a non-adjacent pair $\{a, b\}$, we ensured it is not robust up to depth 1, by forming the subsequent deviation by $\{\pi(b), b\}$. To generalize this idea, we introduce the following definition.

Definition 4. Taking an arbitrary regular matching μ and a deviation (D, ν) from it as given, we say that $\{a, b\} \subseteq D$ is a *convenient pair of deviators* if all of the following hold: $\{a, b\} \notin \mathscr{P}(\sigma)$, $\mu(a) \neq a$, $\nu(a) = b$, a is inferior for b, and $\nu(\pi(b))$ is inferior for $\pi(b)$. □

When a deviation (D, ν) involves a convenient pair $\{a, b\}$, it is not robust up to depth 1 as we state as a lemma below, because b and $\pi(b)$ form a deviation from ν. Note that when $\{a, b\}$ is a convenient pair of deviators, they cannot be adjacent to each other: The qualification of $\{a, b\} \notin \mathscr{P}(\sigma)$ precludes the possibility of $a = \pi(b) = \sigma(b)$. Then, a cannot be inferior for b if $a = \sigma(b) \neq \pi(b)$. Moreover, if $a = \pi(b) \neq \sigma(b)$, then, $\nu(\pi(b)) = b = \sigma(a)$ cannot be inferior for $\pi(b) = a$. The following lemma thus generalizes Claim 1 in Section 4.2 in some sense. It should be noted, however, the former holds solely by the definition of a convenient pair, whereas the latter assumes Properties 1–2. Claim 1 essentially says that non-adjacency implies convenience (and equivalently, non-convenience implies adjacency) when $D = \{a, b\}$ and μ meets the two properties.

Lemma 2. *If a deviation (D, ν) from a regular matching μ contains a convenient pair $\{a, b\}$ of deviators, then it is not robust up to depth* 1.

Proof. The proof is immediate and thus omitted. ■

The following lemma states that we can find a convenient pair of deviators *unless* the deviation has a special structure. When the deviation is pairwise, the conclusion of this lemma reduces to the deviators being adjacent to each other. Thus, it generalizes the way we divided the pairwise deviations in Section 4.

Lemma 3. *Suppose $\nu \rhd_D \mu$, where μ is a regular matching satisfying Properties 1–3, and that no pair $\{a, b\} \subseteq D$ is convenient. Then, $D_S = \nu(D_I) = \pi(D_I)$, where $D_S := \{i \in D : \nu(i) \in D$ is superior for $i\}$ and $D_I := D - D_S$.*

Proof. See Section A.1 below. ■

In general, a deviation may not contain a convenient pair even though all the newly-matched pairs of deviators are non-adjacent. The following is an example of such a deviation.

Example 7. Let $N := \{a_1, \ldots, a_5, b_1, \ldots, b_5\}$ and $\sigma : N \to N$ be such that $\sigma(a_i) = a_{i+1}$ and $\sigma(b_i) = b_{i+1}$ for each i, where the subscripts are in modulo 5. Also define $\succ$ as follows: For each $i \in \{1, 2, 3\}$, $\succ_{a_i}$ and $\succ_{b_i}$ are such that $a_{i+1} \succ_{a_i} a_{i-1} \succ_{a_i} a_i$ and $b_{i+1} \succ_{b_i} b_{i-1} \succ_{b_i} b_i$. For the other four agents, the preferences are given by

$$\succ_{a_4} : b_5 \succ_{a_4} a_5 \succ_{a_4} a_3 \succ_{a_4} a_4, \qquad \succ_{a_5} : a_1 \succ_{a_5} a_4 \succ_{a_5} b_4 \succ_{a_5} a_5,$$
$$\succ_{b_4} : a_5 \succ_{b_4} b_5 \succ_{b_4} b_3 \succ_{b_4} b_4, \text{ and } \succ_{b_5} : b_1 \succ_{b_5} b_4 \succ_{b_5} a_4 \succ_{b_5} b_5.$$

In any case, all the unlisted agents are unacceptable. Note that σ is the unique party permutation for $(N, \succ)$, and there are two non-adjacent mutually-acceptable pairs, $\{a_4, b_5\}$ and $\{b_4, a_5\}$. Now, consider the following two matchings,

$$\mu := \Big\{\{a_1, a_2\}, \{a_3, a_4\}, \{a_5\}, \{b_1, b_2\}, \{b_3, b_4\}, \{b_5\}\Big\}, \text{ and}$$
$$\nu := \Big\{\{a_1, a_2\}, \{a_3\}, \{a_4, b_5\}, \{b_1, b_2\}, \{b_3\}, \{b_4, a_5\}\Big\}.$$

Notice that μ is a regular matching satisfying all the properties and that (D, ν) is a deviation from μ with $D = \{a_4, a_5, b_4, b_5\}$. It is easy to see neither $\{a_4, b_5\}$ nor $\{b_4, a_5\}$

is convenient. The conclusion of Lemma 3, $D_S = \nu(D_I) = \pi(D_I)$, holds with $D_S = \{a_4, b_4\}$ and $D_I = \{a_5, b_5\}$. □

Now, let us introduce our last property, which is unnecessary when the deviation is pairwise as in Section 4. Its purpose is to preclude the possibility of four consecutive agents being deviators. When no pair of deviators is convenient and four consecutive agents are all deviators, by Lemma 3, there must exist b among the four such that $b, \sigma^2(b) \in D_I$ and $\pi(b), \sigma(b) \in D_S$. The next property precludes the possibility of $b, \sigma^2(b) \in D_I$, because D_I is a subset of I°_μ by definitions.

Property 6. *For any $b \in I^\circ_\mu = \{i \in N : \pi(i) \succ_i \mu(i)\}$, we have $\sigma^2(b) \notin I^\circ_\mu$.* □

The rest of this appendix establishes the following two propositions, which generalize Claims 2–3 in Section 4.3. We present their proofs in A.2 and A.3, respectively, after we prove Lemma 3 in A.1.

Proposition 3. *Suppose that μ is a regular matching satisfying Properties 1–4. If $\#(N, \succ)$ is no greater than* 3, 5, *and* 7, *respectively, no deviation from μ is robust up to depth* 1, 2, *and* 3.

Proof. See Section A.2 below. ■

Proposition 4. *Suppose that μ is a regular matching satisfying Properties 1–6. Then, no deviation from μ is robust up to depth* 3.

Proof. See Section A.3 below. ■

A.1 Proof of Lemma 3

Since each deviator should be matched to another at ν and no pair of agents are mutually superior, we have $\nu(D_S) \subseteq D_I$ and hence $|D_S| \leq |D_I|$. It thus suffices to establish $D_S \supseteq \nu(D_I)$ and $D_S \subseteq \pi(D_I)$ for the following reason: Since ν is a bijection, $D_S \supseteq \nu(D_I)$ is equivalent to $\nu(D_S) \supseteq (\nu \circ \nu)(D_I) \equiv D_I$. Combined with $\nu(D_S) \subseteq D_I$, it thus entails $\nu(D_S) = D_I$. As $\nu(D_S) = D_I$ implies $|D_S| = |D_I|$, then, $D_S \subseteq \pi(D_I)$ implies $D_S = \pi(D_I)$.

First, to establish $D_S \subseteq \pi(D_I)$, arbitrarily fix $a^0 \in D_S$. By the definition of D_S, $b^0 := \nu(a^0)$ is superior for a^0 and is a member of D_I. Property 1 implies $\mu(a^0) \neq a^0$, and hence, $\{a^0, b^0\}$ satisfies all the requirements to be a convenient pair except $\nu\left(\pi(b^0)\right)$ being inferior for $\pi(b^0)$.[24] To meet the supposition of the lemma, thus, $\nu\left(\pi(b^0)\right)$ must be superior for $\pi(b^0)$. We then have $\pi(b^0) \in D_S$, because $b^0 \in D_I$ implies $b^0 \in I_\mu^\circ$ and hence, $\mu\left(\pi(b^0)\right)$ is inferior for $\pi(b^0)$ by Property 2. Now recursively define $(a^t, b^t) := \left(\pi(b^{t-1}), \nu(a^t)\right)$ for each $t \in \mathbb{N}$. What we have shown from $a^0 \in D_S$ is $a^1 \in D_S$, which implies $b^1 \in D_I$. By repeatedly applying the arguments, we can obtain $a^t \in D_S$ and $b^t \in D_I$ for each t. Since the number of agents is finite, there must exist some T such that $a^T = a^0$, or equivalently, $a^0 = \pi\left(b^{T-1}\right)$. As $b^{T-1} \in D_I$, it follows that $a^0 \in \pi(D_I)$. Since a^0 is an arbitrary member of D_S, we can conclude that $D_S \subseteq \pi(D_I)$.

To show $D_S \supseteq \nu(D_I)$ and thereby complete the proof, next suppose towards a contradiction that there are $\alpha, \beta \in D_I$ such that $\nu(\alpha) = \beta$. Since the original matching μ is assumed to be regular, at least one of $\mu(\alpha) \neq \alpha$ and $\mu(\beta) \neq \beta$ should hold; without loss of generality, assume $\mu(\alpha) \neq \alpha$. Note that $\{\alpha, \beta\} \notin \mathscr{P}(\sigma)$, as otherwise Property 3 requires they be matched at μ. For $\{\alpha, \beta\}$ not to be a convenient pair of deviators, then, $\pi(\beta)$ must be matched to a superior partner. Then, $\pi(\beta)$ must be a member of D_S, because $\beta \in D_I$ entails $\beta \in I_\mu^\circ$, and hence, $\mu\left(\pi(\beta)\right)$ should be inferior for $\pi(\beta)$ by Property 2. Now apply the arguments in the previous paragraph starting from $a^0 = \pi(\beta)$: That is, $a^t = (\pi \circ \nu)^t a^0$ is a member of D_S for each t, and $a^T = a^0$ for some T. However, these imply that $a^{T-1} \in D_S$ and $a^{T-1} = \nu(\beta) = \alpha$, which contradicts to the assumption of $\alpha \in D_I$. ■

A.2 Proof of Proposition 3

Suppose that (D, ν) is a deviation from a regular matching μ satisfying Properties 1–4. By Lemmas 2–3, we can restrict our attention to the case of $D_S = \nu(D_I) = \pi(D_I)$, where $D_S = \{i \in D : \nu(i) \succ_i \pi(i)\}$ and $D_I = D - D_S$. Recall also that $D_I \subseteq I_\mu^\circ = \{j \in$

[24] Note that $\{a^0, b^0\}$ cannot be a party; if $\{a^0, b^0\} \in \mathscr{P}(\sigma)$, they are mutually inferior to each other, and hence, $\nu(a^0) = b^0$ and $a^0 \in D_S$ are incompatible.

$N : \pi(j) \succ_j \mu(j)\}$ by definitions.

Now arbitrarily fix $b \in D_I \subseteq I^\circ_\mu$ and let $a := \nu(b)$. Since we consider the case of $D_S = \nu(D_I)$, we have $a \in D_S$, which means b is superior for a. Then $\mu(a) \neq a$ should follow, as otherwise Property 1 would require $\mu(b) \succ_b a$. Note also that $P(b)$ is odd, as Property 3 requires that no even party should intersect with I°_μ. It suffices to show that (D, ν) is not robust up to depth 1, 2, and 3, respectively if $|P(b)| = 3$, 5, and 7.

Suppose $|P(b)| = 7$ and hence, $P(b) = \{\pi(b), b, \sigma(b), \ldots, \sigma^5(b)\}$. Since $b \in I^\circ_\mu$ implies $b \notin A_\mu$, Property 4 requires that three adjacent pairs, $\{\sigma(b), \sigma^2(b)\}$, $\{\sigma^3(b), \sigma^4(b)\}$, and $\{\sigma^5(b), \pi(b)\}$ should be matched at μ. Since D_I is a subset of I°_μ, b should be the unique member of $P(b) \cap D_I$. Under the assumption of $D_S = \pi(D_I)$, this in turn implies $P(b) \cap D_S = \{\pi(b)\}$. Since $\sigma^5(b) \neq \pi(b)$ by the assumption of $|P(b)| = 7$, $\sigma^5(b)$ cannot be a member of D and should be single at ν. By matching $\{\sigma^4(b), \sigma^5(b)\}$, $\{\sigma^2(b), \sigma^3(b)\}$, and $\{b, \sigma(b)\}$, thus, we can construct ν_1, ν_2, and ν_3 so that

$$\nu_3 \rhd_{\{b,\sigma(b)\}} \nu_2 \rhd_{\{\sigma^2(b),\sigma^3(b)\}} \nu_1 \rhd_{\{\sigma^4(b),\sigma^5(b)\}} \nu.$$

Since $a \equiv \nu(b) \in D$ is single at ν_3, the original deviation (D, ν) is not robust up to depth 3. The cases for $|P(b)| = 3$ and 5 are similar and thus omitted. □

A.3 Proof of Proposition 4

Suppose that (D, ν) is a deviation from a regular matching μ satisfying Properties 1–6. By Lemmas 2–3, we can restrict our attention to the case of $D_S = \nu(D_I) = \pi(D_I)$, where $D_S = \{i \in D : \nu(i) \succ_i \pi(i)\}$ and $D_I = D - D_S$. Recall that by definitions, $D_I \subseteq I^\circ_\mu$ and any deviator should belong to a non-solitary odd party.[25]

To begin with, fix an agent $b \in D_I$ such that $\sigma^3(b) \notin D_S$. We should be able to find such b for the following reason: Suppose that there is no such b. Since $D_S = \pi(D_I)$, then, $\sigma^4(b') \in D_I$ holds for all $b' \in D_I$. This, however, is a contradiction since

[25]Under the assumption of $D_S = \pi(D_I)$, if a party intersects with D, then it should intersect with both D_I and D_S, which are disjoint to each other by definitions. Thus, no solitary party can be a part of D. Any even party is disjoint from D because it is so from D_I by Property 3.

$P(b')$ is odd for any $b' \in D_I$ by Property 3.[26] Taking b with $\sigma^3(b) \notin D_S$ as given, let $a := \nu(b) \in D_S$, $m \in \mathbb{N}$ be such that $|P(b)| = 2m+1$, and $c_j := \sigma^j(b)$ for each $j \in \{1, \ldots, 2m\}$. Remember that $a \in D_S$ means $b = \nu(a)$ is superior for a and thus implies $\mu(a) \neq a$ by Property 1. To establish the non-robustness of (D, ν) up to depth κ, it suffices to construct a sequence of κ further deviations such that $\nu_\kappa \rhd_{D_\kappa} \cdots \nu_1 \rhd_{D_1} \nu$, $a \notin D_1 \cup \cdots \cup D_\kappa$, and $b \in D_\kappa$.

If $\nu(c_1)$ is inferior for c_1, then (D, ν) is not robust up to depth 1 since we can construct ν_1 by immediately matching b and c_1 so that $\nu_1 \rhd_{\{b, c_1\}} \nu$. For the rest of the proof, thus, we investigate two cases assuming $\nu(c_1)$ is superior for c_1. Notice that in any case, $a \notin \{c_1, \nu(c_1)\}$.[27]

Case 1: $\nu(c_1) \neq c_2$ is superior for c_1. We first show $c_1 \notin D$ as follows. Towards a contradiction, suppose otherwise. Since $\nu(c_1)$ is assumed to be superior, then, $c_1 \in D_S$ and hence, $c_2 \in D_I$ by $D_S = \pi(D_I)$. This, however, contradicts Property 6, because $b \in D_I$ and $c_2 \equiv \sigma^2(b) \in D_I$ respectively imply $b \in I^\circ_\mu$ and $\sigma^2(b) \in I^\circ_\mu$. We should thus have $c_1 \notin D$, which also necessitates $\nu(c_1) = \mu(c_1) \notin D$ because $\nu(c_1) \neq c_1$.

Next we define $d := \nu(c_1) = \mu(c_1)$ and show that $\pi(d)$ is matched to an inferior partner or is single at ν. To begin, remember that c_1 is inferior for d, as no pair is mutually superior. Since $d \neq c_2 \equiv \sigma(c_1)$ by assumption, it follows that $d \in I^\circ_\mu$. Property 2 then guarantees that $\mu(\pi(d))$ is inferior for $\pi(d)$. Moreover, $\pi(d)$ is not a member of D_S; otherwise, $d \in D_I$ follows from $D_S = \pi(D_I)$, but we have already confirmed $d \equiv \nu(c_1) \notin D$. Therefore, $\pi(d) \in D$ is possible only with $\pi(d) \in D_I$. Consequently, $\nu(\pi(d))$ is inferior for $\pi(d)$ no matter if $\pi(d) \in D$ or not. Since $\{\pi(d), d\}$ cannot be a two-agent party, d is superior for $\pi(d)$ by the definition of a party permutation.[28] It thus follows that $\pi(d)$ prefers d to $\nu(\pi(d))$.

Given the above observations, we can construct ν_1 and ν_2 by matching $\{\pi(d), d\}$

[26] Given P is an odd party, $|P| \bmod 4$ must be either 1 or 3. In either case, $\{\sigma^{4t}(b) : t \in \mathbb{N}\} = P$ holds for any $b \in P$. If $b \in D_I$ for some $b \in P$ and $b' \in D_I \Rightarrow \sigma^4(b') \in D_I$ for all $b' \in P$, thus, $P = \{\sigma^{4t}(b) : t \in \mathbb{N}\} \subseteq D_I$ should follow. This, however, contradicts $D_S = \pi(D_I)$.

[27] First, $a \neq c_1$ because b is assumed to be superior for a whereas b is inferior for $c_1 \equiv \sigma(b)$. Second, $a \neq c_1$ because $a = \nu(b)$ by assumption and $b \neq c_1$ by definition.

[28] If $\{\pi(d), d\} \in \mathscr{P}(\sigma)$, Property 3 would require they be matched at μ; however, it would contradict the definition that $d := \nu(c_1) = \mu(c_1)$.

and $\{b, c_1\}$, respectively, so that $\nu_2 \rhd_{\{b,c_1\}} \nu_1 \rhd_{\{\pi(d),d\}} \nu$. Since $a \equiv \nu(b)$ is single at ν_2, the original deviation (D, ν) is not robust up to depth 2.

Case 2: $\nu(c_1) = c_2$. Note that this case arises only when $\mu(c_1) = c_2$, as Property 6 guarantees $c_2 \notin I^\circ_\mu$, which is equivalent to $\mu(c_2) \succeq_{c_2} c_1$. Note further that $|P(b)| \geq 5$ is also necessary; if $|P(b)| = 3$, then c_2 should be equal to $\pi(b) \in D_S$, but this contradicts $\nu(c_2) = c_1$ being inferior for c_2. Therefore, $c_3 \equiv \sigma^3(b) \neq b$ in this case. Note that this also entails $a \notin \{c_3, \nu(c_3)\}$.[29] If $\nu(c_3)$ is inferior for c_3, then ν is not robust up to depth 2, because we can construct ν_1 and ν_2 by respectively matching $\{c_2, c_3\}$ and $\{b, c_1\}$, so that $\nu_2 \rhd_{\{b,c_1\}} \nu_1 \rhd_{\{c_2,c_3\}} \nu$.

For the rest of the proof, we consider the subcase where $\nu(c_3)$ is superior for c_3. As we have chosen b so that $c_3 \equiv \sigma^3(b) \notin D_S$, we should have $c_3 \notin D$ and hence $\nu(c_3) = \mu(c_3)$. To simplify the notation, in what follows, let e denote $\nu(c_3) = \mu(c_3)$. Note that $a \notin \{c_3, e\}$, since $a \in D$ by assumption. First, suppose $e \neq c_4$. Then, by the same arguments as when we showed $\nu(\pi(d))$ is inferior for $\pi(d)$ in Case 1, $\nu(\pi(e))$ must be inferior for $\pi(e)$. Since this implies $\pi(e) \neq a$, we can then construct ν_1, ν_2, and ν_3, by respectively matching $\{\pi(e), e\}$, $\{c_2, c_3\}$, and $\{b, c_1\}$, so that $\nu_3 \rhd_{\{b,c_1\}} \nu_2 \rhd_{\{c_2,c_3\}} \nu_1 \rhd_{\{\pi(e),e\}} \nu$. That is, the original deviation (D, ν) is not robust up to depth 3.

To complete the proof, our last subcase to consider is $e = c_4$. Remember that we have shown $P(b) \geq 5$. Then, $e = c_4$ requires $c_4 \neq \pi(b)$ and hence $|P(b)| \geq 7$, since $c_3, e \notin D$ as argued above while $\pi(b) \in D$ by assumption. That is, we have $c_5 \equiv \sigma^5(b) \notin \{b, \sigma(b), \ldots, \sigma^4(b)\}$. Now, we prove that $\nu(c_5)$ cannot be superior for c_5 for the following reasons:

- If $|P(b)| = 7$, Property 4 requires μ should match three adjacent pairs, $\{c_1, c_2\}$, $\{c_3, c_4\}$, and $\{c_5, c_6\}$. Since $D_I \subseteq I^\circ_\mu$ and $D_S = \pi(D_I)$, it follows that b and $c_6 \equiv \pi(b)$ are the unique member of $P \cap D_I$ and of $P \cap D_S$, respectively. That is, c_5 is not a deviator and is left, by $c_6 \equiv \pi(b)$, to be single at ν.
- If $P(b) > 7$, Property 5 requires $c_5 \in I^\circ_\mu$ and $c_6 \notin I^\circ_\mu$. For $\nu(c_5)$ to be superior for

[29] First, $a \neq c_3$ follows from the assumption of $\sigma^3(b) \notin D_S$. Second, $a = \nu(b)$ and $b \neq c_3$ imply $a \neq \nu(c_3)$.

c_5, thus, $c_5 \in D_S$ is necessary. Under the assumption of $D_S = \pi(D_I)$, however, this requires $c_6 \in D_I$, which is incompatible with $c_6 \notin I^\circ_\mu$.

Given $\nu(c_5) \neq c_4$ is inferior for c_5, we have $a \neq c_5$. We can thus construct ν_1, ν_2, and ν_3, by matching $\{c_4, c_5\}$, $\{c_2, c_3\}$, and $\{b, c_1\}$, respectively, so that $\nu_3 \rhd_{\{b,c_1\}} \nu_2 \rhd_{\{c_2,c_3\}} \nu_1 \rhd_{\{c_4,c_5\}} \nu$. Since $a \equiv \nu(b)$ is single at ν_3, the original deviation (D, ν) is not robust up to depth 3. ■

B Construction of a SaRD Matching for Theorems 1–3

This appendix presents an algorithm to compute, for any problem $(N, \succ)$, a regular matching that satisfies Properties 1–6. We first provide an overview of the algorithm in B.1. We then fully specify the algorithm in B.2. Lastly, we establish in B.3 that its outcome is always regular and satisfies all the properties (Proposition 5). Combined with the results of Appendix A, this constitutes the proofs for Theorems 1–3. It should also be noted that the outcome of our algorithm is always a stable matching whenever there is any.

B.1 Overview of the Algorithm

In this subsection, we overview our algorithm and briefly explain how it ensures regularity and Properties 1–6. Our algorithm takes a problem $(N, \succ)$ and a party permutation σ as arbitrarily given, and it computes a matching μ as its output through six phases. It matches pairs step by step, and it never resolves any pair it once has matched. The algorithm contains some arbitrary choices of agents and pairs, and its outcome generally varies with those choices. However, *any* outcome of the algorithm meets our goals even if it is not unique.

In Phases 1–2 of the algorithm, we focus on even parties and "small" odd parties, and we form as many adjacent pairs of agents (with respect to the given σ) as possible. In Phase 1, we match every agent a in each even party to an agent adjacent to her, i.e., to either $\sigma(a)$ or $\pi(a)$. In Phase 2, we form $\frac{|P|-1}{2}$ adjacent pairs for each odd party

P whose size $|P|$ is less than or equal to seven, leaving exactly one arbitrary agent unmatched. Remember that these are exactly what Properties 3–4 require. Phase 1 also ensures that at the final outcome μ, the even parties should be disjoint from I_μ°. Thus, Properties 2 and 6 are vacuous for the agents in an even party. Further, if a small odd party contains $b \in I_\mu^\circ$, she should be the one who is left unmatched in Phase 2; thus, $\pi(b)$ and $\sigma^2(b)$ should be matched, respectively, to $\pi^2(b)$ and $\sigma(b)$. That is, Properties 2 and 6 will also hold for $b \in I_\mu^\circ$ within the small odd parties.

In Phase 3, we match adjacent pairs in each odd party whose size is a multiple of three. Specifically, we match n adjacent pairs for a party of $3n$ agents, by skipping exactly one agent between any two adjacent pairs. Those parties can contain multiple $b \in I_\mu^\circ$ in the end, depending on the outcomes of the subsequent phases. Yet, it is still true that if b is left unmatched at the end of this phase, $\pi(b)$ and $\sigma^2(b)$ should be matched, respectively, to $\pi^2(b)$ and $\sigma(b)$. Thus, Properties 2 and 6 should be satisfied for $b \in I_\mu^\circ$ from a party of $3n$ agents. Property 5 is also vacuous for such parties, as this phase does not consecutively match two adjacent pairs.

Phase 4 is the main innovation of this algorithm. Contrary to the previous phases, here we match non-adjacent (and possibly across-party) pairs, as well as adjacent ones. To begin, we arbitrarily order and label all the agents who are yet to be matched as $x_1, \ldots, x_T$. Each step t of this phase then proceeds roughly as follows: Let Y_t be the set of those who are still unmatched, are mutually acceptable with x_t, and perceive x_t as superior.[30] If x_t is already matched during an earlier step or if Y_t is empty, then we proceed to the next step without matching any pair. Otherwise, we match x_t to her most preferred agent y_t among Y_t. When x_t and y_t are matched, we then check if we can match any adjacent pairs in $P(x_t)$ and $P(y_t)$. If we can, we do so in a systematic way. In particular, we "exhaust" adjacent pairs in the following sense: If an adjacent pair in party P is formed at step t of this phase, then P contains no adjacent pair both of whom remain unmatched after this step. Thus, no adjacent pair will be matched in the same party afterwards.

[30] More precisely, we exclude $\pi(x_t)$ and $\pi^2(x_t)$ from Y_t. Since x_t is inferior for $\sigma(x_t)$ by definition, none of Y_t is adjacent to x_t.

Phase 4 is designed to simultaneously guarantee all of Properties 1, 2, 5, and 6 for the relevant parties. First, the sequential matching of non-adjacent $\{x_t, y_t\}$'s ensures that the final outcome μ meets Property 1. To see why, suppose that a is superior for b, b is left unmatched at the final outcome μ, and that b is acceptable for a.[31] Since the first supposition implies $\pi(a) \succeq_a b$, $\mu(a) \succ_a b$ would fail to hold only if $\pi(a) \succ_a \mu(a)$. (Note that $\mu(a) = \pi(a) = b$ is impossible by the second supposition.) Furthermore, with the above suppositions, our algorithm ensures that $\pi(a) \succ_a \mu(a)$ occurs only if $a = x_t$ is matched to y_t at some step t of Phase 4; if so, a should prefer $\mu(a) = y_t$ to b by construction, because $b \in Y_t$.[32]

Second, the systematic matching of adjacent pairs guarantees Properties 2, 5, and 6 for the parties relevant in this phase, while we defer the details to the full description in B.2. The reasons for Properties 2 and 6 are similar to those in the previous phases. The key for Property 5 is how to match adjacent pairs "around" x_t and y_t. If x_t is matched to y_t at some step t of this phase, then she belongs to I°_μ at the final outcome. We thus need to ensure $\sigma^5(x_t) \in I^\circ_\mu$ whenever $\mu(\sigma(x_t)) = \sigma^2(x_t)$ and $\mu\left(\sigma^3(x_t)\right) = \sigma^4(x_t)$. Actually, we match adjacent pairs so that the two equations *never* simultaneously hold and thereby make the requirement vacuous. Similarly, we also make sure that $\pi^5(y_t) \notin I^\circ_\mu$ whenever $\mu(\pi(y_t)) = \pi^2(y_t)$ and $\mu\left(\pi^3(y_t)\right) = \pi^4(y_t)$ both hold. Consequently the requirement for $\sigma^5(b)$ is satisfied when we substitute $b = \pi^5(y_t)$ and $\sigma^5(b) = y_t$, even though $y_t \notin I^\circ_\mu$. Our construction also warrants the requirement for $\sigma^6(b)$ in Property 5.

In Phase 5, we consider the parties from which no agent has been matched by Phase 4, and we form adjacent pairs so as to meet Properties 2, 5, and 6. In particular, we match them so that both (i) no more than two consecutive adjacent pairs are matched and (ii) no pair of the remaining agents are adjacent. As we will establish as Lemma 4 below, an agent will necessarily be a member of I°_μ at the final outcome, if she is unmatched at the end of this phase. Together with the patterns of adjacent pairs Phases

[31] If b is unacceptable for a, then $\mu(a) \succ_a b$ follows from the individual rationality of the final outcome.

[32] Strictly speaking, we exclude $\pi(a)$ and $\pi^2(a)$ from Y_t as we mentioned in footnote 30. We thus need to treat the case of $b \in \{\pi(a), \pi^2(a)\}$ separately.

1–5 match, this feature ensures all the properties referring to I°_{μ}, Properties 2–6, at the final outcome. It should also be noted that we need to run Phases 1–5 in this order so as to obtain this feature.[33]

In Phase 6, lastly, we match those who are still unmatched so that no mutually-acceptable pair of singles is left. As we have argued above, Property 1 is guaranteed by Phase 4, and how we match (non-adjacent) pairs in this phase is irrelevant to Properties 2–6. Hence, our task in this phase is just to satisfy regularity, and we can exhaust mutually-acceptable pairs in an arbitrary way.

B.2 Description of the Algorithm

Taking a problem $(N, \succ)$ and a party permutation σ as given, construct a matching μ as follows. To simplify the description, we write "define $\mu(a) := b$," when it should read as "define $\mu(a) := b$ and $\mu(b) := a$." The whole procedure is divided into six phases.

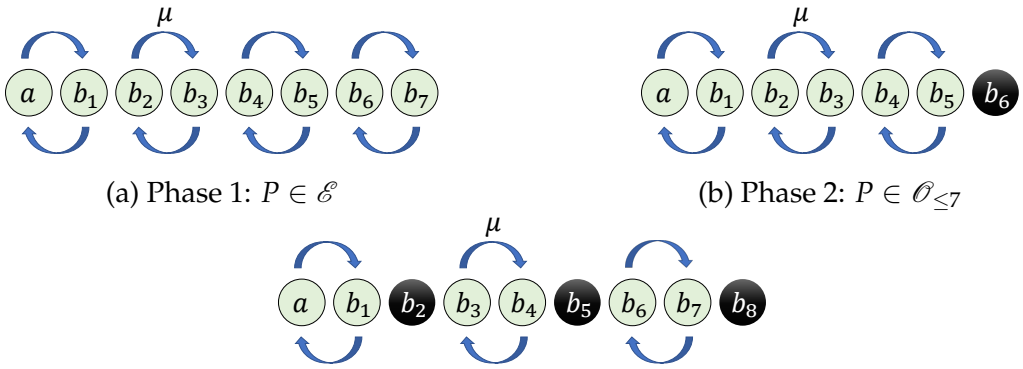
B.2.1 Phase 1 of the Algorithm

Let $\mathscr{E}$ be the family of even parties; i.e., $\mathscr{E} := \{P \in \mathscr{P}(\sigma) : |P| \text{ is even}\}$. For each $P \in \mathscr{E}$, arbitrarily take $a \in P$ and define $\mu\left(\sigma^{2j-2}(a)\right) := \sigma^{2j-1}(a)$ for each $j \in \left\{1, \ldots, \frac{|P|}{2}\right\}$ as illustrated in Figure 1 (a).

B.2.2 Phase 2 of the Algorithm

Let $\mathscr{O}_{\leq 7}$ be the family of non-solitary odd parties whose sizes are less than or equal to seven; i.e., $\mathscr{O}_{\leq 7} := \{P \in \mathscr{P}(\sigma) : |P| \in \{3,5,7\}\}$. For each $P \in \mathscr{O}_{\leq 7}$, arbitrarily take $a \in P$ and define $\mu\left(\sigma^{2j-2}(a)\right) := \sigma^{2j-1}(a)$ for each $j \in \left\{1, \ldots, \frac{|P|-1}{2}\right\}$ as illustrated in Figure 1 (b). Note that $\mu\left(\sigma^{|P|-1}(a)\right)$ is undefined.

[33] As seen more clearly below, Phases 1–3 and 5 might look similar in that all of them match adjacent pairs only, whereas Phase 4 is quite different. It would thus be natural to ask what happens if we run Phase 5 *before* Phase 4 and thereby fix all the adjacent pairs before entering the complicated phase. It turns out, however, such an alternative algorithm does not work, mainly because it does not guarantee the feature we highlight here.



(a) Phase 1: $P \in \mathscr{E}$

(b) Phase 2: $P \in \mathscr{O}_{\leq 7}$

(c) Phase 3: $P \in \mathscr{O}_{3\times}$

Figure 1: Matching during Phases 1–3 of the Algorithm. For each j, b_j represents $\sigma^j(a)$. Each arrow between two agents means they are matched, and the agents represented by black circles are not matched in these phases.

B.2.3 Phase 3 of the Algorithm

Let $\mathscr{O}_{3\times}$ be the family of odd parties whose sizes are a multiple of three and greater than three; i.e., $\mathscr{O}_{3\times} := \{P \in \mathscr{P}(\sigma) : |P| = 3n \text{ for some odd } n \geq 3\}$. For each $P \in \mathscr{O}_{3\times}$, arbitrarily take $a \in P$ and define $\mu\left(\sigma^{3j}(a)\right) := \sigma^{3j+1}(a)$ for each $j \in \left\{1, \ldots, \frac{|P|}{3}\right\}$ as illustrated in Figure 1 (c). Note that for each j, we leave $\mu\left(\sigma^{3j+2}(a)\right)$ undefined.

B.2.4 Phase 4 of the Algorithm

Let $U_0 \subseteq N$ be the set of agents who are unmatched yet and $\mathscr{U}_0$ the family of parties none from which is matched yet.[34] Specifically, $P \in \mathscr{U}_0$ if and only if its cardinality is odd, greater than seven, and not a multiple of three. In what follows, U_t and $\mathscr{U}_t$ will be, respectively, the set of agents who are unmatched by step t of this phase and the family of parties no agent from which is matched by step t.

Arbitrarily order the members of U_0 as $x_1, \ldots, x_T$, where $T := |U_0|$, and iterate the following step for $t = 1, \ldots, T$.

[34]Remember that $a \in U_0$ does not necessarily imply $P(a) \in \mathscr{U}_0$ since $P(a)$ may be in $\mathscr{O}_{\leq 7} \cup \mathscr{O}_{3\times}$.

Step $t = 1, \ldots, T$ of Phase 4:

If $x_t \notin U_{t-1}$, proceed to step $t+1$ with $U_t = U_{t-1}$ and $\mathscr{U}_t = \mathscr{U}_{t-1}$. Otherwise, define

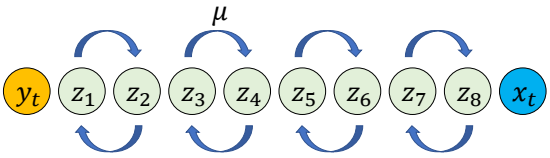
$$Y_t := \left\{ y \in U_{t-1} - \{\pi(x_t), \pi^2(x_t)\} : x_t \text{ is superior for } y \text{ and } y \text{ is acceptable for } x_t \right\}.$$

If Y_t is empty, then proceed to step $t+1$ with $U_t = U_{t-1}$ and $\mathscr{U}_t = \mathscr{U}_{t-1}$.[35] Otherwise, let $y_t \in Y_t$ denote the best agent for x_t among those in Y_t; that is, $y \in Y_t \Rightarrow y_t \succeq_{x_t} y$. Define $\mu(x_t) := y_t$ and $\mathscr{U}_t = \mathscr{U}_{t-1} - \{P(x_t), P(y_t)\}$. If $\mathscr{U}_t = \mathscr{U}_{t-1}$, proceed to step $t+1$ with $U_t = U_{t-1} - \{x_t, y_t\}$. Otherwise, we match adjacent pairs in $P(x_t)$ and/or $P(y_t)$ as we specify below. Note that in either case, we "exhaust" adjacent pairs in the relevant party; i.e., if both of an adjacent pair have been unmatched by the end of this step t, then they belong to some $P \in \mathscr{U}_t$.

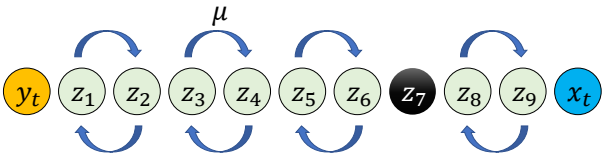
Case 1: $P(x_t) = P(y_t) \in \mathscr{U}_{t-1}$. In this case, there exist $q, r \leq |P(x_t)|$ such that $\sigma^{q+1}(y_t) = x_t$ and $\sigma^{r+1}(x_t) = y_t$. It should also be noted that $q \geq 2$ by the definition of Y_t. Match adjacent pairs in $P(x_t) = P(y_t)$ as follows:

- Matching among $\sigma(y_t), \ldots, \sigma^q(y_t)$:
  If $q = 2m$ for some $m \in \mathbb{N}$, then $\mu\left(\sigma^{2j-1}(y_t)\right) := \sigma^{2j}(y_t)$ for each $j \in \{1, \ldots, m\}$. If $q = 2m+1$ for some $m \in \mathbb{N}$, then $\mu\left(\sigma^{2j-1}(y_t)\right) := \sigma^{2j}(y_t)$ for each $j \in \{1, \ldots, m-1\}$, and $\mu(\left(\sigma^{2m}(y_t)\right) := \sigma^{2m+1}(y_t)$, leaving $\mu\left(\sigma^{2m-1}(y_t)\right)$ undefined. Figure 2 (a)–(b) illustrate the matching in these cases.
- Matching among $\sigma(x_t), \ldots, \sigma^r(x_t)$:
  If $r = 3n$ or $3n+1$ for some $n \in \mathbb{N}$, then, let $\mu\left(\sigma^{3j'-1}(x_t)\right) := \sigma^{3j'}(x_t)$ for each $j' \in \{1, \ldots, n\}$. Notice that $\mu\left(\sigma^{3n+1}(x_t)\right)$ is undefined when $r = 3n+1$. Similarly, we leave $\mu\left(\sigma(x_t)\right)$ undefined if $r = 1$. If $r = 3n+2$ for some $n \in \mathbb{N} \cup \{0\}$, then, let $\mu\left(\sigma^{3j'-2}(x_t)\right) := \sigma^{3j'-1}(x_t)$ for each $j' \in \{1, \ldots, n+1\}$. Figure 2 (c)–(e) illustrate the matching in these cases.

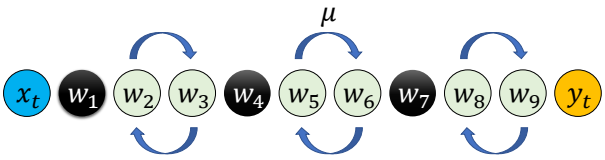
[35] Remember that when $\{x_t\} \in \mathscr{P}(\sigma)$ is a solitary party, y is acceptable for x_t if and only if y is superior for x_t. Since this implies x_t is inferior for y by the definition of a party permutation, in such a case, Y_t must be empty.



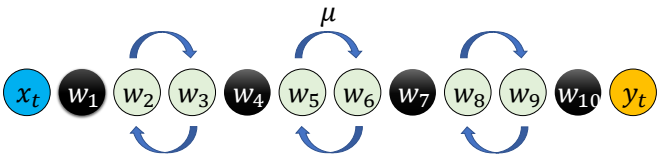
(a) Case of q being even



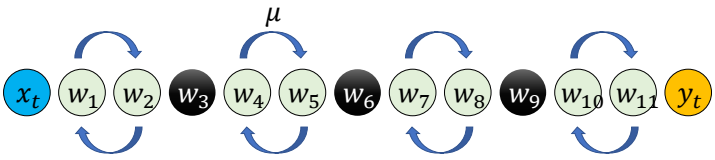
(b) Case of q being odd



(c) Case of $r = 3n$ for some $n \in \mathbb{N}$

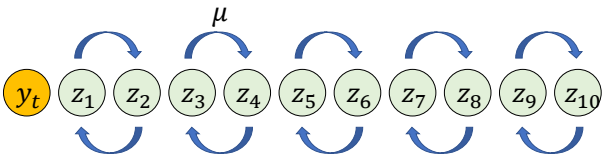


(d) Case of $r = 3n + 1$ for some $n \in \mathbb{N} \cup \{0\}$

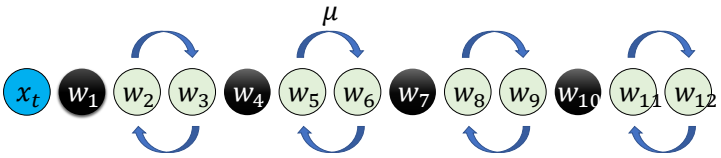


(e) Case of $r = 3n + 2$ for some $n \in \mathbb{N} \cup \{0\}$

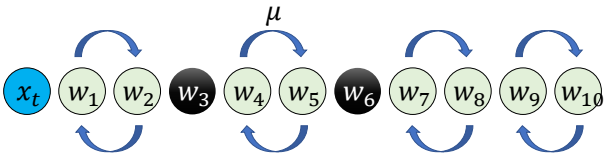
Figure 2: Matching in Case 1 of Phase 4. For each j, z_j and w_j denote $\sigma^j(y_t)$ and $\sigma^j(x_t)$, respectively. Each arrow between two agents means they are matched, and the agents represented by black circles are not matched in this step.



(a) Matching of $P(y_t)$



(b) Matching of $P(x_t)$ with $|P(x_t)| = 3n+1$ for some $n \in \mathbb{N}$



(c) Matching of $P(x_t)$ with $|P(x_t)| = 3n+2$ for some $n \in \mathbb{N}$

Figure 3: Matching of the agents in $P(x_t), P(y_t) \in \mathscr{U}_{t-1}$ in Case 2 of Phase 4. For each j, z_j and w_j denote, respectively, $\sigma^j(y_t)$ and $\sigma^j(x_t)$. Each arrow between two agents means they are matched, and the agents represented by black circles are not matched in this step.

Let $U_t := U_{t-1} - M_t$, where M_t is the set of agents matched in this step, including x_t and y_t, and proceed to step $t+1$.

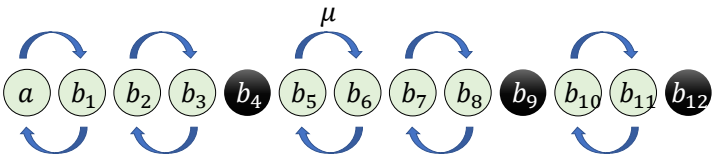
Case 2: $P(x_t) \neq P(y_t)$. In this case, match the members of $P(x_t)$ and $P(y_t)$, respectively, if $P(x_t) \in \mathscr{U}_{t-1}$ and if $P(y_t) \in \mathscr{U}_{t-1}$ as follows:

- Matching among $P(y_t) \in \mathscr{U}_{t-1}$:

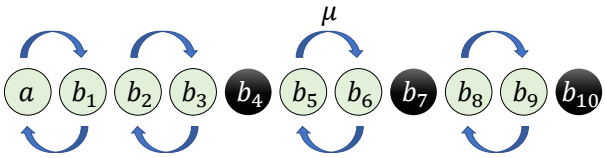
  If $P(y_t) \in \mathscr{U}_{t-1}$, define $\mu\left(\sigma^{2j-1}(y_t)\right) := \sigma^{2j}(y_t)$ for each $j \in \left\{1, \ldots, \frac{|P(y_t)|-1}{2}\right\}$ as illustrated in Figure 3 (a).

- Matching among $P(x_t) \in \mathscr{U}_{t-1}$:

  If $P(x_t) \in \mathscr{U}_{t-1}$, then $|P(x_t)| = 3n+1$ or $3n+2$ for some $n \in \mathbb{N}$, as $\mathscr{U}_{t-1} \subset \mathscr{U}_0$ is disjoint from $\mathscr{O}_{3\times}$. In the former case, define $\mu\left(\sigma^{3j'-1}(x_t)\right) := \sigma^{3j'}(x_t)$ for each $j' \in \{1, \ldots, n\}$. In the latter, let $\mu\left(\sigma^{3j'-2}(x_t)\right) := \sigma^{3j'-1}(x_t)$ for each $j' \in \{1, \ldots, n\}$ and $\mu\left(\sigma^{3n}(x_t)\right) := \sigma^{3n+1}(x_t)$. Figures 3 (b)–(c) illustrate the matching in these cases.



(a) $P \in \mathscr{U}_T$ with $|P| = 3n + 1$ for some $n \in \mathbb{N}$



(b) $P \in \mathscr{U}_T$ with $|P| = 3n + 2$ for some $n \in \mathbb{N}$

Figure 4: Matching during Phase 5. For each j, b_j denotes $\sigma^j(a)$. Each arrow between two agents means they are matched, and the agents represented by black circles are not matched in this Phase.

Let $U_t := U_{t-1} - M_t$, where M_t is the set of agents matched in this step, including x_t and y_t, and proceed to step $t + 1$.

B.2.5 Phase 5 of the Algorithm

Remember that $\mathscr{U}_T$ is the family of odd parties no member from which has been matched yet. Recall also that that $|P|$ is not a multiple of three for any $P \in \mathscr{U}_T$. For each $P \in \mathscr{U}_T$, fix an arbitrary member $a \in P$ and match adjacent pairs in the following way, as illustrated in Figure 4:

- If $|P| = 3n + 1$ for some $n \in \mathbb{N}$, then, define $\mu(a) := \sigma(a)$, $\mu\left(\sigma^2(a)\right) := \sigma^3(a)$, $\mu\left(\sigma^5(a)\right) := \sigma^6(a)$, and $\mu\left(\sigma^{3j-2}(a)\right) := \sigma^{3j-1}(a)$ for each $j \in \{3, \ldots, n\}$.
- If $|P| = 3n + 2$ for some $n \in \mathbb{N}$, then define $\mu(a) := \sigma(a)$, $\mu\left(\sigma^2(a)\right) := \sigma^3(a)$, and $\mu\left(\sigma^{3j-1}(a)\right) := \sigma^{3j}(a)$ for each $j \in \{2, \ldots, n\}$.

B.2.6 Phase 6 of the Algorithm

Let R_0 be the set of those who still remain unmatched, and arbitrarily order its members as $r_1, \ldots, r_{|R_0|}$. Iterate the following step for $\tau = 1, \ldots, |R_0| + 1$. We will have defined $\mu(a)$ for each a after these steps, and the algorithm is complete.

- Step $\tau = 1, \ldots, |R_0|$ of Phase 6:

  If $r_\tau \in R_{\tau-1}$ and there exists some $r_i \in R_{\tau-1}$ who is mutually acceptable with r_τ, then define $\mu(r_\tau) := r_i$ and proceed to step $\tau + 1$ with $R_\tau := R_\tau - \{r_\tau, r_i\}$.[36]
- Step $|R_0| + 1$ of Phase 6:

  For any $r \in R_{|R_0|}$, i.e., for any agent not matched yet, define $\mu(r) := r$.

B.3 Properties of the Algorithm

In this subsection, we formally establish the properties of the outcomes of the above algorithm. To begin with, we prove the following lemma, which we highlighted in B.1:

Lemma 4. *Let R_0 be the set of the agents who remain unmatched at the beginning of Phase 6. For any $a, b \in R_0$ with $a \neq b$, either of the following statements holds: [1] they are not mutually acceptable, and [2] they are inferior for each other.*

Proof. Let $a, b \in R_0$ with $a \neq b$. We then have $a \neq \pi(b)$ because there is no adjacent pair among R_0. In addition, we also have $a \neq \pi^2(b)$, because as one can confirm from Figures 1–4, it follows from $b \in R_0$ that $\pi^2(b)$ is matched by the end of Phase 5.[37]

Towards a contradiction, suppose now that a and b are mutually acceptable and b is superior for a. For $b \in R_0$, there should exist step t of Phase 4 such that $x_t = b \in U_{t-1}$. Since $a \notin \{\pi(b), \pi^2(b)\}$, our assumptions entail $a \in Y_t$. However, this in turn implies that $x_t = b$ should have been matched to y_t during Phase 4, which contradicts the original assumption of $b \in R_0$. As a and b are symmetric, the proof is complete. ■

Now we are ready to prove the following proposition. Together with the results in Appendix A, it completes the proofs of Theorems 1–3.

Proposition 5. *Let μ be an outcome of the algorithm we specify in B.2. For any problem $(N, \succ)$, then, μ is regular and satisfies Properties 1–6 (with respect to the party permutation σ*

[36] In general multiple members of $R_{\tau-1}$ may be mutually acceptable with r_τ. Even if so, the choice of r_i can be arbitrary.

[37] More specifically, $b \in R_0$ should be left unmatched when the algorithm matches adjacent pairs in $P(b)$ during one of Phases 2–5. That is, b should correspond to a black circle in some of those Figures. It is then easy to check that $\pi^2(b)$ is always matched into an adjacent pair in any of those Figures.

fixed at the beginning of the algorithm). Consequently, μ is stable whenever $(N, \succ)$ possesses a stable matching.[38]

Proof of Regularity. It is immediate to check that μ is individually rational as we only match mutually-acceptable pairs during the algorithm. It leaves no mutually-acceptable pairs of singles because of Phase 6. ■

Proof of Property 1. We establish $\mu(a) \succ_a b$ assuming that a is superior for b and $\mu(b) = b$. Note that b should be inferior for a, i.e., $\pi(a) \succeq_a b$, because no pair is mutually superior. Suppose further that b is acceptable for a, as otherwise $\mu(a) \succ_a b$ immediately follows from individual rationality. Then, Lemma 4 necessitates $a \notin R_0$; i.e., a should be matched to $\mu(a) \neq a$ by the end of Phase 5.

If a is a part of an adjacent pair at μ, it is easy to see $\mu(a) \succ_a b$: If $\mu(a) = \pi(a)$, then $\mu(a) \succ_a b$ holds, because $\mu(b) = b$ implies $b \neq \mu(a) = \pi(a)$, and $\pi(a) \succeq_a b$ as noted above. If $\mu(a) = \sigma(a) \neq \pi(a)$, then $\mu(a) \succ_a b$ follows from $\sigma(a) \succ_a \pi(a)$, which is a part of the definition of a party permutation.

What remains to check is the case where a is matched to $\mu(a) \notin \{\pi(a), \sigma(a)\}$ during Phase 4. If $a = y_t$ is matched to x_t in some step t during Phase 4, $\mu(a) = x_t$ is superior for $a = y_t$ and hence, $\mu(a) \succ_a b$ holds. If $a = x_t$ is matched to y_t in some step t during Phase 4, our assumptions imply $b \in Y_t$.[39] It thus follows that $\mu(a) = y_t \succ_a b$, because y_t is chosen to satisfy $y_t \succ_a y$ for any $y \in Y_t - \{y_t\}$. ■

Proof of Property 2. Suppose $b \in I^{\circ}_{\mu}$, which implies $P(b)$ is non-solitary and odd. There are two cases: (i) $P(b) \in \mathscr{U}_{t-1}$ and $b = x_t$ is matched to y_t at some step t of Phase 4 and (ii) b is left unmatched when one of Phases 2–5 matches adjacent pairs in $P(b)$, although b may be matched to $\mu(b)$ afterwards. In the first case, one can confirm, with Figure 2 (a)–(b) and Figure 3 (b)–(c), that $\pi(x_t)$ is always matched to $\pi^2(x_t)$. In

[38] Remember that when a stable matching exists (i.e., if $\#(N, \succ) \leq 1$), individual rationality and Property 3 imply stability, as we illustrated in Section 2.2.

[39] In this case, $b \notin \{\pi(a), \pi^2(a)\}$ holds for the following reason: As we assume $\mu(b) = b$, it suffices to confirm that neither $\pi(a)$ nor $\pi^2(a)$ is single at μ, which is clearly true if $\mu(\pi(a)) = \pi^2(a)$. Given $a = x_t$ is matched to y_t during Phase 4, $\mu(\pi(a)) = \pi^2(a)$ fails only if $\pi(a) = x_{t'}$ is matched to $y_{t'}$ in an earlier step $t' < t$. Moreover, for both a and $\pi(a)$ to remain unmatched until step t', we must have $P(a) \in \mathscr{U}_{t'-1}$ and hence, $\pi^2(a)$ should also be matched in step t' (to $\pi^3(a)$).

the second case, b should correspond to some black circle in one of Figures 1–4. With those Figures, one can verify that either [1] $\pi(b)$ is matched to $\pi^2(b)$ or [2] $\pi(b) = x_{t'}$ at some step t' of Phase 4 and is matched to $y_{t'}$. In any case, $\mu(\pi(b))$ is inferior for $\pi(b)$. ■

Proof of Properties 3–4. These two properties are immediate from the constructions in Phases 1–2. ■

Proof of Property 5. We derive $\sigma^5(b) \in I^\circ_\mu$ and $\sigma^6(b) \notin I^\circ_\mu$ from $b \in I^\circ_\mu$, $\mu(\sigma(b)) = \sigma^2(b)$, $\mu\left(\sigma^3(b)\right) = \sigma^4(b)$, and $|P(b)| = 2m + 1$ with $m > 3$. Note that $|P(b)|$ is not a multiple of three, because otherwise all the adjacent pairs in $P(b)$ are formed during Phase 3, which never matches two consecutive adjacent pairs. Therefore, no agent in $P(b)$ is matched by the end of Phase 3, i.e., $P(b) \in \mathscr{U}_0$. We divide the case into two depending on whether any agent in $P(b)$ is matched during Phase 4.

First, suppose that t is the first step of Phase 4 such that agents in $P(b)$ are matched. Note that $b = y_t$ is impossible as $\mu(y_t) = x_t$ is superior for y_t by definition. Further, $b = x_t$ is also impossible, because $\mu(\sigma(x_t)) = \sigma^2(x_t)$ and $\mu\left(\sigma^3(x_t)\right) = \sigma^4(x_t)$ never simultaneously hold, as one can confirm with Figure 2 (c)–(e) and Figure 3 (b)–(c). Thus, b must be left unmatched at this step t. More specifically, the only possibility consistent with $\mu(\sigma(b)) = \sigma^2(b)$ and $\mu\left(\sigma^3(b)\right) = \sigma^4(b)$ is the case illustrated in Figure 3 (c), where $P(b) = P(x_t) \neq P(y_t)$, $P(b) = 3n + 2$ for some n, and $\sigma^5(b) = x_t$.[40] In such a case, $\sigma^5(b) = x_t$ is matched to y_t, and $\sigma^6(b) = \sigma(x_t)$ is matched to her successor. These respectively imply $\sigma^5(b) \in I^\circ_\mu$ and $\sigma^6(b) \notin I^\circ_\mu$, as desired.

Next, suppose that no agent from $P(b)$ is matched during Phase 4, i.e., $P(b) \in \mathscr{U}_T$. By Lemma 4, a member of $P(b)$ also belongs to I°_μ if and only if she is not matched into an adjacent pair during Phase 5. With Figure 4, it is then easy to confirm that $b \in I^\circ_\mu$, $\mu(\sigma(b)) = \sigma^2(b)$, and $\mu\left(\sigma^3(b)\right) = \sigma^4(b)$ jointly imply that $\sigma^5(b) \in I^\circ_\mu$ and $\sigma^6(b) \notin I^\circ_\mu$. ■

[40] Since $b \in I^\circ_\mu$ but $b \neq x_t$, b should correspond to one of the black circles in Figures 2–3. Among them, only $b = w_6$ in Figure 3 (c) is consistent with the assumption of $\mu(\sigma(b)) = \sigma^2(b)$ and $\mu\left(\sigma^3(b)\right) = \sigma^4(b)$.

Proof of Property 6. Suppose $b \in I^{\circ}_{\mu}$. Remember that there are two possibilities: (i) $P(b) \in \mathscr{U}_{t-1}$ and $b = x_t$ is matched to y_t at some step t of Phase 4 and (ii) b is left unmatched when one of Phases 2–5 matches adjacent pairs in $P(b)$, although b may be matched to $\mu(b)$ afterwards. In both cases, one can confirm with Figures 1–4 that $\sigma^2(b)$ is always matched to either $\sigma(b)$ or $\sigma^3(b)$. Hence, $\sigma^2(b)$ never belongs to I°_{μ}. ■

C Proof of Proposition 2

Suppose $\#(N, \succ) = 2k + 1$ and fix an arbitrary $\mathscr{P}$-stable matching μ'. Construct another matching μ that includes μ' by the following procedure:

- First, for each a such that $\mu'(a) \neq a$, let $\mu(a) := \mu'(a)$.
- Next, run Phase 4 of our algorithm in Appendix B with $U_0 := \{a : \mu'(a) = a\}$.
- Lastly, run Phase 6 of our algorithm in Appendix B.

Then, by similar arguments to those in Appendix A, one can confirm that μ is SaRD up to depth k.[41] ■

D Generalization of Tan's (1991) Theorems

Tan (1991) originally establishes his results under two additional assumptions we do not impose in this paper: (i) the number of agents is even and (ii) the preferences are symmetric in the sense that $a \succ_b b \Leftrightarrow b \succ_a a$ for all $a, b \in N$. It is well known as a folk knowledge that his results continue to hold without those assumptions, but to the best of our knowledge, none in the literature has provided an explicit proof for such extensions. For completeness of the paper and for possible future reference, this appendix demonstrates why Tan's (1991) results hold in the from we presented in Section 2.2 without the additional assumptions.

Note that the condition for the existence of a stable matching is straightforward once we generalize the existence of a party permutation and the uniqueness of odd parties. When there is a party permutation without any non-solitary odd party, we can

[41] For details, refer to the working paper version (Hirata et al., 2020).

construct a stable matching from a party permutation, by matching all the members of even parties into adjacent pairs, as we demonstrated in Section 2.2. Conversely, when a stable matching $\mu : N \to N$ exists, $\sigma = \mu$ constitutes a party permutation such that all odd parties (if any) are a singleton. Therefore, a stable matching exists if and only if $\#(N, \succ) \leq 1$. In what follows, thus, we generalize the existence of a party permutation and the uniqueness of odd parties to an arbitrarily given problem.

To begin with, consider a problem $(N, \succ)$ such that $N = \{a_1, \ldots, a_{2n-1}\}$ while $\succ = (\succ_{a_1}, \ldots, \succ_{a_{2n-1}})$ is a symmetric preference profile. So as to apply Tan's original result, construct another problem $(N', \succ')$ by adding a dummy agent as follows: $N' = \{a_0, a_1, \ldots a_{2n-1}\}$ and $\succ' = (\succ'_{a_0}, \succ'_{a_1} \ldots, \succ'_{a_{2n-1}})$, where

- the dummy agent's preference $\succ'_{a_0}$ is such that $a_0 \succ'_{a_0} b$ for all $b \in N$, and
- for each $a \in N$, her preference $\succ'_a$ is such that $b \succ'_a a_0$ for all $b \in N$ and that $c \succ'_a d \Leftrightarrow c \succ_a d$ for all $c, d \in N$.

Note that $|N'| = 2n$ is even and that $\succ'$ is symmetric since the original $\succ$ is. By Tan (1991, Theorem 3.3), therefore, a party permutation exists and odd parties are uniquely identified at $(N', \succ')$. To confirm that the same is true also at the original $(N, \succ)$, we establish a one-to-one correspondence between party permutations for $(N, \succ)$ and for $(N', \succ')$. Notice that at any party permutation for $(N', \succ')$, the dummy agent must form a solitary party. Thus, if σ' is a party permutation for $(N', \succ')$, the restriction of σ' to N continues to meet all the requirements to be a party permutation for $(N, \succ)$. Conversely, when $\sigma|_N : N \to N$ is a party permutation for $(N, \succ)$, we can construct a party permutation $\sigma : N' \to N'$ for $(N', \succ')$ by extending $\sigma|_N$ with $\sigma(a_0) = a_0$. With these observations, it is immediate to see that the existence of a party permutation and the uniqueness of odd parties are inherited from $(N', \succ')$ to $(N, \succ)$.

Next suppose $\succ = (\succ_a)_{a \in N}$ is asymmetric, while $|N|$ is either even or odd. Let $\succ^* = (\succ^*_a)_{a \in N}$ be a "symmetrization" of $\succ$ such that

- $b \succ^*_a a \Leftrightarrow [b \succ_a a$ and $a \succ_b b]$ for any $a, b \in N$, and
- $b \succ^*_a c \Leftrightarrow b \succ_a c$ for any $a, b, c \in N$ such that $b, c \succ^*_a a$.[42]

[42] A symmetrization of $\succ$ is not unique because the ranking among unacceptable agents is not pinned

The first condition requires that a pair of agents are mutually acceptable at $\succ^*$ if and only if they are at $\succ$. Combined with the second condition, it follows that a's ranking between b and c remains unchanged if both $\{a, b\}$ and $\{a, c\}$ are mutually acceptable. Recall that we have already established that Tan's theorem holds for $(N, \succ^*)$. Hence, it suffices to confirm that a permutation σ over N is a party permutation for $(N, \succ)$ if and only if it is so for $(N, \succ^*)$. In doing so, we consider the two directions separately.

First, let σ^* be a party permutation for the symmetrized $(N, \succ^*)$ and π^* its inverse. By the definition of $\succ^*$, it follows from $\sigma^*(a) \succ_a^* a$ and $\sigma^*(a) \succ_a^* \pi(a) \succ_a^* a$, respectively, that $\sigma^*(a) \succ_a a$ and $\sigma^*(a) \succ_a \pi(a) \succ_a a$. Hence, σ^* is a semi-party permutation for $(N, \succ)$. Towards a contradiction, now suppose that a and b are superior to each other with respect to $\succ$; that is, $a \succ_b \pi^*(b)$ and $b \succ_a \pi^*(a)$. Note that $\pi^*(b)$ is acceptable for b at $\succ_b$, as she is so at $\succ_b^*$. The supposition of $a \succ_b \pi^*(b)$ thus implies that a is acceptable for b at $\succ_b$. Combined with the symmetric arguments, a and b must be mutually acceptable at $\succ$. Then, the suppositions of $a \succ_b \pi^*(b)$ and $b \succ_a \pi^*(a)$ imply the same rankings continue to hold at $\succ^*$. This, however, contradicts the original assumption that σ^* is a party permutation with respect to $\succ^*$. Therefore, no such a and b should exist, and σ^* is a party permutation for $(N, \succ)$.

Second, suppose that σ is a party permutation for the original problem $(N, \succ)$. Recall that the transformation from $\succ$ to $\succ^*$ keep the set of mutually-acceptable pairs unchanged. For any member a of a non-solitary party, hence, $(a, \sigma(a))$ and $(a, \pi(a))$ are a mutually acceptable pair not only with $\succ$ but also with $\succ^*$. Therefore, $\sigma(a) \succ_a^* a$ and $\sigma(a) \succ_a^* \pi(a) \succ_a^* a$ follow, respectively, from $\sigma(a) \succ_a a$ and $\sigma(a) \succ_a \pi(a) \succ_a a$. That is to say, σ is a semi-party permutation for $(N, \succ^*)$. To complete the proof, let $a, b \in N$ be such that $a \succ_b^* \pi(b)$. Since b and $\pi(b)$ are mutually acceptable with respect to $\succ$, this necessitates $a \succ_b \pi(b)$. Then $\pi(a) \succeq_a b$ follows from the assumption that σ is a party permutation for $(N, \succ)$. As a and $\pi(a)$ are also mutually acceptable, this entails $\pi(a) \succeq_a^* b$.

To summarize, we have demonstrated that a party permutation exists and the odd

down. However, the following argument is independent of the choice of $\succ^*$.

parties are uniquely identified for any number of agents and any preference profiles, as we stated in Section 2.2.