Journal of Artificial Intelligence Research 80 (2024) 1063-1098 Submitted 08/2023; published 07/2024

On the Convergence of Swap Dynamics to Pareto-Optimal Matchings

Felix Brandt BRANDT@TUM.DE
Technische Universität München
Boltzmannstr. 3, 85748 Munich, Germany

Anaëlle Wilczynski ANAELLE.WILCZYNSKI@CENTRALESUPELEC.FR
MICS, CentraleSupélec, Université Paris-Saclay
9 rue Joliot Curie, 91190 Gif-sur-Yvette, France

Abstract

We study whether Pareto-optimal stable matchings can be reached via pairwise swaps in one-to-one matching markets with initial assignments. We consider housing markets, marriage markets, and roommate markets as well as three different notions of swap rationality. Our main results are as follows. While it can be efficiently determined whether a Pareto-optimal stable matching can be reached when defining swaps via blocking pairs, checking whether this is the case for *all* such sequences is computationally intractable. When defining swaps such that all involved agents need to be better off, even deciding whether a Pareto-optimal stable matching can be reached via *some* sequence is intractable. This confirms and extends a conjecture made by Damamme, Beynier, Chevaleyre, and Maudet (2015) who have shown that convergence to a Pareto-optimal matching is guaranteed in housing markets with single-peaked preferences. We prove that in marriage and roommate markets, single-peakedness is not sufficient for this to hold, but the stronger restriction of one-dimensional Euclidean preferences is.

1. Introduction

One-to-one matchings, where individuals are matched with resources or other individuals, are omnipresent in everyday life. Examples include the job market, assigning offices to workers, pairing students in working groups, and online dating. The formal study of matching procedures is fascinating and has been of increasing interest within the computer science and AI communities, because it leads to challenging mathematical and algorithmic problems while being of immediate practical interest (see, e.g., Manlove (2013), Klaus, Manlove, and Rossi (2016)).

One typically distinguishes between three different types of abstract one-to-one matching settings. In *housing markets* (Shapley & Scarf, 1974), each agent is matched with an object (usually referred to as a house). In *marriage markets* (Gale & Shapley, 1962), agents are partitioned into two groups—say, males and females—and each member of one group is matched with an agent from the other group. Finally, in *roommate markets* (Gale & Shapley, 1962), all agents belong to the same group, and each agent is matched with another agent. By supposing that the agents are rational and want to maximize their satisfaction, individual agreements may naturally occur among them and especially, for realistic reasons, between small groups of agents. An important question, which is common in multi-agent systems, is then whether sequences of such individual agreements can lead to socially optimal outcomes.

In many applications, it is reasonable to assume that there is an initial assignment because agents already live in a house, are engaged in a relationship, and are employed by a company (see, e.g., Abdulkadiroğlu and Sönmez (1999), Morrill (2010)). Under these assumptions, we focus on atomic agreements that require the least coordination: *pairwise swaps*.

In general, we consider three different types of individual rationality for pairwise swaps. In housing markets, there is only one meaningful notion of swap rationality: two agents will only exchange objects if both of them are better off. By contrast, when matching agents with each other, one could require that all four agents involved in a swap, or just two of them, are better off. The latter requirement allows for two kinds of swap rationality: two agents who exchange their match are better off (e.g., a company and its subsidiary exchange employees without asking for their consent), or two agents who decide to form a new pair are better off (e.g., two lovers leave their current partners to be together).

Social optimality in settings with ordinal preferences like that of matching markets is measured in terms of Pareto-optimality. We therefore study whether there exists a sequence of pairwise swaps that results in a Pareto-optimal matching that does not allow for further swaps (and hence is called stable). Whenever all sequences of pairwise swaps are of this kind, we say that the given type of swap dynamics converges. Note that in two-sided matching, Pareto-optimality classically refers to Pareto-optimality according to only one side of the market. Since we focus on general matching markets, we do not restrict Pareto-optimality to a subset of agents but define it for the whole set of involved agents (i.e., both males and females for the specific case of the marriage market).

It turns out that in all three types of matching markets and all three notions of swap rationality, it may not be possible to reach a Pareto-optimal stable matching from the initial assignment. We prove that deciding whether this is the case is NP-hard for two types of swap rationality while it can be solved in polynomial time for swaps based on blocking pairs. However, for all types of swap rationality, checking convergence is co-NP-hard. On the other hand, we show that when preferences are one-dimensional Euclidean—a natural but demanding restriction—swap dynamics for two types of swap rationality will always converge.

2. Related Work

Given an initial matching, objects can be reassigned in housing markets to obtain a Pareto-optimal matching by considering successive improving cycles of exchanges along agents, following the well-known top-trading cycle (TTC) algorithm (Shapley & Scarf, 1974). However, such a centralized algorithm may be difficult to implement in a distributed manner since an exchange cycle of arbitrary length may be necessary. Therefore, Damamme et al. (2015) investigated the dynamics of individually rational pairwise swaps in housing markets, where two agents are better off by exchanging their objects. Recently, variants of this problem that further restrict the agents' interactions using underlying graph structures have been examined (Gourvès, Lesca, & Wilczynski, 2017; Saffidine & Wilczynski, 2018; Huang & Xiao, 2020).

In marriage and roommate markets, most of the literature focuses on deviations based on blocking pairs, where two agents decide to leave their old partners in order to be matched with each other. Blocking pairs are best known for their role in the definition of stability (Gale & Shapley, 1962), but some papers also studied the dynamics of blocking pair swaps (Roth & Vande Vate, 1990; Abeledo & Rothblum, 1995). A specific blocking-pair dynamics has, for example, been studied by Salonen and Salonen (2018) in the context of the college admission problem (an extension of marriage markets where colleges can be assigned to more than one student): they iteratively match the mutually best pairs to form a stable matching. Among other properties, they have also studied Pareto-optimality for one side of the market and preference restrictions such as single-peaked or single-crossing preferences. However, in contrast to their work, we do not focus on the properties of a specific centralized mechanism which carefully chooses the agents to match but investigate the properties of the distributed process of swap dynamics.

One of the type of swaps we investigate (blocking pair swaps) has already been studied by Knuth (1976) and can make the old partners, who are matched with each other, worse off. It is thus related to the breakmarriage operation (McVitie & Wilson, 1971) for finding alternative stable matchings in marriage markets. However, the breakmarriage operation differs significantly from blocking pair swaps since only one of the four involved agents may be better off in the new matching.

The notion of exchange stability, where two agents agree to exchange their partners has been investigated in both roommate markets (Alcalde, 1994; Cechlárová, 2002) and marriage markets (Cechlárová & Manlove, 2005). To the best of our knowledge, exchange rational swaps have not been studied in the context of dynamics that reach Pareto-optimal matchings.

In contrast to our definition of Pareto-optimality, some papers on swap dynamics have investigated matchings that are Pareto-optimal *among all reachable matchings* (Gourvès et al., 2017; Aziz, 2019). Other types of dynamics that have been considered in matching markets include pairwise swaps without local rationality constraints (Aziz, 2019), Pareto improvements (Morrill, 2010; Aziz, Brandt, & Harrenstein, 2013), local dynamics based on underlying graphs (Hoefer, 2013; Hoefer, Vaz, & Wagner, 2018), and exchanges among more than two agents (Aziz & Goldwaser, 2017; Aziz, 2019).

Note that we focus on the standard definition of Pareto-optimality where no agent can be better off without making another one worse off, with no distinction among the agents, even in the marriage market. This choice is made for not favoring any type of agents and for serving our purpose to study general one-to-one matching markets and see the connections between them. By contrast, the majority of papers in marriage markets (or in its college admission extension) focus on Pareto-optimality according to one side of the market. Under such an assumption, an incompatibility may notably arise in marriage markets between blocking pair stability and Pareto-optimality for one side of the market (see, e.g., Abdulkadiroğlu, Che, Pathak, Roth, and Tercieux (2017)), which does not occur for our definition of Pareto-optimality. Indeed, our notion of Pareto-optimality in marriage markets, with no distinction among agents, is weaker than Pareto-optimality for one side of the market.

Perhaps closest to our work is a result by Damamme et al. (2015) who proved that swap dynamics always converge to a Pareto-optimal matching in housing markets when the preferences of the agents are single-peaked. They left open the computational problem of deciding whether a Pareto-optimal stable matching can be reached for unrestricted preferences and conjectured this problem to be intractable. We solve this problem and extend it to marriage and roommate markets. Moreover, we prove that their convergence result for housing markets under single-peaked preferences does not extend to marriage and roommate markets, but can be restored when restricting preferences even further.

3. The Model

We are given a set N of agents $\{1, \dots, n\}$ and a set O of objects $\{o_1, \dots, o_n\}$ such that $|N| = |O| = n$. Each agent $i \in N$ has strict ordinal preferences, represented by a linear order $\succ_i$, over a set A_i of alternatives to be matched with. In the matching markets we consider, A_i is either a subset of the set of agents N or the set of all objects O. A tuple of preference relations $\succ = (\succ_1, \dots, \succ_n)$ is called a *preference profile*.

3.1 Matching Markets

In this article, we are considering three different settings where the goal is to match the agents either with objects—like in housing markets—or with other agents—like in marriage or roommate markets. In all cases, we assume that there is an initial matching. More formally,

- a *housing market* consists of a preference profile where $A_i = O$ for all $i \in N$, and an initial endowment given as a bijection $\mu : N \to O$,
- a *marriage market* consists of a preference profile with even n where $N = W \cup M$, $W \cap M = \emptyset$ and $|W| = |M| = n/2$, with $A_i = M$ for all $i \in W$ and $A_i = W$ for all $i \in M$, and an initial matching given as a bijection $\mu : W \to M$, and
- a *roommate market* consists of a preference profile with even n and $A_i = N \setminus \{i\}$ for all $i \in N$, and an initial matching given as an involution $\mu : N \to N$ such that $\mu(i) \neq i$ for all $i \in N$.

In marriage markets, we will sometimes denote the inverse function μ^{-1} of matching μ by μ for the sake of simplicity. We refer to a general matching market as a tuple $(N, (A_i)_{i \in N}, \succ, \mu^0)$ where μ^0 denotes the initial matching. A possible matched pair in a matching market is a pair $\{i, x\}$ such that $i \in N$ and $x \in A_i$. A matched pair in a matching μ is a pair $\{i, x\}$ such that $\mu(i) = x$.

When allowing for indifferences as well as unacceptabilities in the preferences (see, e.g., Gusfield and Irving (1989) for marriage markets, Irving and Manlove (2002) for roommate markets, and Alcalde-Unzu and Molis (2011) for housing markets), the three settings form a hierarchy: housing markets are marriage markets where the "objects" are indifferent between all agents, and marriage markets are roommate markets where all agents of the same type are considered unacceptable. In this paper, however, we do not make either assumption and therefore, these inclusion relationships do not hold.

The key question studied in this paper is whether Pareto-optimal matchings can result from a sequence of local modifications starting from the initial matching. A matching is *Pareto-optimal* if there is no other matching μ' such that for every agent $i \in N$, $\mu'(i) \succeq_i \mu(i)$ and for at least one agent $j \in N$, $\mu'(j) \succ_j \mu(j)$.

Note that checking whether a given matching μ is Pareto-optimal can be done in polynomial time for all matching markets under consideration since we are considering strict preferences (see, e.g., Abraham, Cechlárová, Manlove, and Mehlhorn (2005) for housing markets and Abraham and Manlove (2004) for roommate markets). Indeed, it suffices to use an adaptation of the TTC algorithm: We construct a graph $G = (N \cup (A_i)_{i \in N}, R \cup B)$ over the agents and alternatives. There is a red edge in R between agent $i \in N$ and alternative $a \in A_i$ if $a \succ_i \mu(i)$ and, in case $A_i \subseteq N$, $i \succ_a \mu(a)$. There is a blue edge in B between agent $i \in N$ and alternative $a \in A_i$ if $\mu(i) = a$. Matching μ is Pareto-optimal iff there is no alternating cycle in G, i.e., a cycle which alternates between blue edges and red edges (this is also known as a μ-augmenting cycle in the matching literature). An alternating cycle can be detected in polynomial time (see, e.g., Gabow, Kaplan, and Tarjan (2001)).

3.2 Preference Restrictions

We consider three restricted preference domains in this article: single-peaked preferences (Black, 1948), globally-ranked preferences (Abraham, Levavi, Manlove, & O'Malley, 2008) and their common subdomain of one-dimensional Euclidean preferences (Coombs, 1950). A preference profile $\succ$ is *single-peaked* if there exists a linear order $>$ over the alternatives in $A := \bigcup_{i \in N} A_i$ such that for each agent i in N and each triple of alternatives $x, y, z \in A_i$ with $x > y > z$ or $z > y > x$, $x \succ_i y$ implies $y \succ_i z$. A preference profile $\succ$ is *globally-ranked* (we also speak about *correlated markets* (Ackermann, Goldberg, Mirrokni, Röglin, & Vöcking, 2008)) if there exists a global order $\rhd$ over all possible matched pairs $\{\{i, x\} : i \in N \text{ and } x \in A_i\}$ in the matching market such that for every agent $i \in N$ and any two alternatives $x, y \in A_i$, $x \succ_i y$ iff $\{i, x\} \rhd \{i, y\}$. Globally-ranked preferences impose no restriction in a housing market (the agents are matched with objects which do not express preference) but may capture in other markets the idea that each pair of agents generates an absolute profit, and thus, each agent prefers the agents with whom she can get a better profit. A preference profile $\succ$ is *one-dimensional Euclidean (1-Euclidean)* if there exists an embedding $E : N \cup O \to \mathbb{R}$ on the real line such that for every agent $i \in N$ and any two alternatives $x, y \in A_i$, $x \succ_i y$ iff $|E(i) - E(x)| < |E(i) - E(y)|$. More generally, a preference profile $\succ$ is *d-dimensional Euclidean (d-Euclidean)* if there exists an embedding $E : N \cup O \to \mathbb{R}^d$ such that for every agent $i \in N$ and any two alternatives $x, y \in A_i$, $x \succ_i y$ iff $\|E(i) - E(x)\| < \|E(i) - E(y)\|$. Clearly, a d-Euclidean preference profile is also d'-Euclidean, for any $d' > d$.

It is well known that one-dimensional Euclidean preferences form a subdomain of single-peaked preferences (see, e.g., Coombs (1964)), because every 1-Euclidean preference profile is singled-peaked with respect to the linear order $>$ given by $x > y$ iff $E(x) > E(y)$. Below, we illustrate with a minimal counterexample that a single-peaked preference profile may not be 1-Euclidean.

Observation 1. *The 1-Euclidean preference domain is strictly contained in the single-peaked preference domain.*

Example 1. *Consider an instance with four agents. Each agent $i \in N$ has preferences over the same set of alternatives $A_i = O = \{o_1, o_2, o_3, o_4\}$, which are given below.*

$$\begin{array}{lccccccc} 1: & o_1 & \succ & o_2 & \succ & o_3 & \succ & o_4 \\ 2: & o_4 & \succ & o_3 & \succ & o_2 & \succ & o_1 \\ 3: & o_2 & \succ & o_3 & \succ & o_4 & \succ & o_1 \\ 4: & o_3 & \succ & o_2 & \succ & o_1 & \succ & o_4 \end{array}$$

Observe that this preference profile is single-peaked only with respect to the linear order $o_1 < o_2 < o_3 < o_4$ (or its reverse order) because of the preferences of Agents 1 and 2. Suppose that this preference profile is 1-Euclidean with respect to an embedding E on the real line. Then, without loss of generality, we can assume that $E(o_1) < E(o_2) < E(o_3) < E(o_4)$. Since $o_2 \succ_3 o_3 \succ_3 o_1$, we have $E(o_1) < E(3) < E(o_3)$, and since $o_3 \succ_4 o_2 \succ_4 o_4$, we have $E(o_2) < E(4) < E(o_4)$. Moreover, by the fact that Agent 3 prefers o_2 to o_3 and Agent 4 prefers o_3 to o_2, it must hold that $E(3) < E(4)$. However, $o_4 \succ_3 o_1$, therefore $E(o_4) - E(3) < E(3) - E(o_1)$. It follows that $E(o_4) - E(4) < E(4) - E(o_1)$, implying that Agent 4 prefers o_4 to o_1, a contradiction.

Note that this counterexample is minimal with respect to the number of agents. For three agents, we build the embedding based on the single-peaked axis over alternatives with equal distance between consecutive alternatives. Then, if the most preferred alternative of agent i is the extreme-left (resp., extreme-right) one in the axis, then we embed agent i on the left (resp., on the right) of this alternative. Otherwise, we place agent i in the embedding between her most preferred alternative and her second most preferred one and closer to her most preferred alternative.

Moreover, one-dimensional Euclidean preferences, and more generally d-Euclidean preferences, form a subdomain of globally-ranked preferences: from a d-Euclidean preference profile, a global ranking over all possible pairs can be extracted by sorting all pairs according to the Euclidean distance on the embedding E between the two partners.[1] We illustrate below with a minimal counterexample that a globally-ranked preference profile may not be 1-Euclidean.

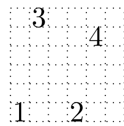
Observation 2. *The 1-Euclidean preference domain is strictly contained in the 2-Euclidean and the globally-ranked preference domains.*

Example 2. *Consider a roommate market with four agents. The preferences of the agents are given below.*

$$\begin{array}{lccccc} 1: & 2 & \succ & 3 & \succ & 4 \\ 2: & 1 & \succ & 4 & \succ & 3 \\ 3: & 4 & \succ & 1 & \succ & 2 \\ 4: & 3 & \succ & 2 & \succ & 1 \end{array}$$

Observe that this preference profile is globally-ranked with respect to global order $\{1,2\} \rhd \{3,4\} \rhd \{1,3\} \rhd \{2,4\} \rhd \{2,3\} \rhd \{1,4\}$ over all possible matched pairs in the market, and 2-Euclidean with respect to, e.g., the following planar embedding.

1. Note that Abraham et al. (2008) have already observed that one-dimensional Euclidean preferences are globally-ranked.



Suppose that this preference profile is 1-Euclidean with respect to an embedding E on the real line. The two possible extremities of E are either (i) 1 and 4 or (ii) 2 and 3. In the first case (i), we deduce from the preferences of Agents 1 and 4 that $E(1) < E(2) < E(3) < E(4)$ (or the reverse order). However, following such an embedding, Agent 2 must prefer Agent 3 to Agent 4, a contradiction. In the second case (ii), we deduce from the preferences of Agents 2 and 3 that $E(2) < E(1) < E(4) < E(3)$ (or the reverse order). However, following such an embedding, Agent 1 must prefer Agent 4 to Agent 3, a contradiction.

Note that this counterexample is trivially minimal for a roommate market, since a smaller counterexample would involve only two agents.

While 1-Euclidean preferences form a subdomain of both single-peaked preferences and globally-ranked preferences, there is no relationship between single-peakedness and globally-ranked preferences. Indeed, as shown in Example 2, a globally-ranked preference profile may not be 1-Euclidean and thus may not be single-peaked. Conversely, we show below with a minimal counterexample that a single-peaked preference profile may not be globally-ranked.

Observation 3. *There is no relationship between the globally-ranked and the single-peaked preference domains.*

Example 3. *Consider a roommate market with four agents. The preferences of the agents are given below.*

$$
\begin{array}{lccccc}
1: & 2 & \succ & 3 & \succ & 4 \\
2: & 3 & \succ & 4 & \succ & 1 \\
3: & 4 & \succ & 2 & \succ & 1 \\
4: & 1 & \succ & 2 & \succ & 3
\end{array}
$$

Observe that this preference profile is single-peaked with respect to, e.g., the linear order $1 > 2 > 3 > 4$ (or its reverse order). Suppose that this preference profile is globally-ranked with respect to a global order $\rhd$ over all possible matched pairs. It follows from the preferences of Agent 2 that $\{2,3\} \rhd \{1,2\}$, from the preferences of Agent 1 that $\{1,2\} \rhd \{1,4\}$, and from the preferences of Agent 4 that $\{1,4\} \rhd \{3,4\}$. By transitivity, we thus get that $\{2,3\} \rhd \{3,4\}$, a contradiction with the preferences of Agent 3 where $4 \succ_3 2$.

Note that this counterexample is trivially minimal for a roommate market since a smaller counterexample would involve only two agents.

We know that 1-Euclidean preferences are both globally-ranked and single-peaked. However, the reverse is not true: we show below with a minimal counterexample that a globally-ranked and single-peaked preference profile may not be 1-Euclidean, even in markets matching agents with each other.

Observation 4. *The 1-Euclidean preference domain is strictly contained in the intersection of the globally-ranked and the single-peaked preference domains.*

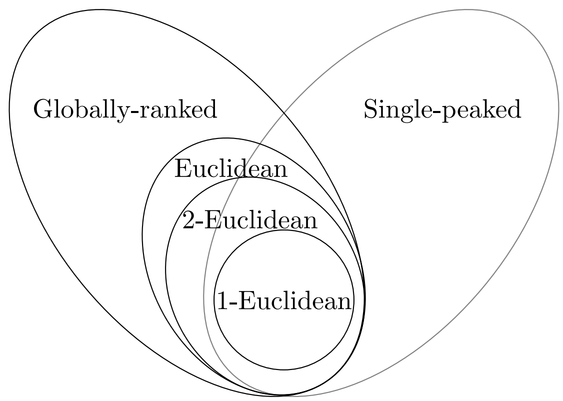


Figure 1: Relationships between globally-ranked, single-peaked, and Euclidean preference domains.

Example 4. *Consider a roommate market with four agents. The preferences of the agents are given below.*

$$\begin{array}{lccccc} 1: & 2 & \succ & 3 & \succ & 4 \\ 2: & 3 & \succ & 4 & \succ & 1 \\ 3: & 2 & \succ & 1 & \succ & 4 \\ 4: & 2 & \succ & 1 & \succ & 3 \end{array}$$

Observe that this preference profile is single-peaked with respect to, e.g., the linear order $1 > 2 > 3 > 4$ *(or its reverse order). Moreover, this preference profile is globally-ranked with respect to the global order* $\{2,3\} \rhd \{2,4\} \rhd \{1,2\} \rhd \{1,3\} \rhd \{1,4\} \rhd \{3,4\}$ *over all possible matched pairs in the market. Suppose that this preference profile is 1-Euclidean with respect to an embedding* E *on the real line. The two possible extremities of* E *can only be* 3 *and* 4*, say* $E(3) < E(4)$*. If* $E(3) < E(1) < E(2) < E(4)$*, then it contradicts the preferences of Agent* 3 *who prefers Agent* 2 *to Agent* 1*. Otherwise, i.e., if* $E(3) < E(2) < E(1) < E(4)$*, then it contradicts the preferences of Agent* 4 *who prefers Agent* 2 *to Agent* 1*.*

Note that this counterexample is trivially minimal for a roommate market since a smaller counterexample would involve only two agents.

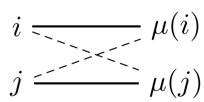
Based on Observations 1-4, the inclusion relationships among the restricted domains considered here can be depicted as shown in Figure 1.

While assuming that all agents have 1-Euclidean preferences certainly represents a strong restriction, there are nevertheless some applications where this assumption is not unreasonable, in both marriage and roommate markets. For example, in job markets, preferences could be 1-Euclidean because employees prefer one workplace to another if it is closer to their home, or when pairing workers in offices with a joint thermostat, workers could prefer co-workers whose most preferred room temperature is closer to their own. Moreover, when forming pairs of students for the realization of a project, a student could prefer to be matched with a student who is most productive during the same time of the day.

In this article, we focus on globally-ranked, single-peaked, and 1-Euclidean preference restrictions. Note that all considered preference restrictions are recognizable in polynomial time: a polynomial-time algorithm for checking single-peakedness is provided by Bartholdi, III and Trick (1986) and has been improved by Escoffier, Lang, and Öztürk (2008), while polynomial-time algorithms for checking the 1-Euclidean property have been proposed by Doignon and Falmagne (1994), Knoblauch (2010), and Elkind and Faliszewski (2014). Moreover, checking whether a preference profile is globally-ranked boils down to checking the acyclicity of the directed graph defined over all possible matched pairs in the market and where there is an arc from a pair $\{i, j\}$ to a pair $\{i, k\}$ if and only if $k \succ_i j$ (Abraham et al., 2008); this can thus be done in polynomial time. However, in general, recognizing d-Euclidean preferences for $d \geq 2$ is intractable (Peters, 2017).

3.3 Rational Swaps

We study sequences of matchings in which two pairs of the current matching are permuted. More formally, we assume that a swap w.r.t. a pair of agents $\{i, j\}$ transforms a matching μ into a matching μ' where agents i and j have exchanged their matches, i.e., $\mu'(i) = \mu(j)$ and $\mu'(j) = \mu(i)$, while the rest of the matching remains unchanged, i.e., $\mu'(k) = \mu(k)$ for every $k \notin \{i, j, \mu(i), \mu(j)\}$.



— matching μ

--- matching μ' after the swap w.r.t. pair $\{i, j\}$

We furthermore require these swaps to be *rational* in the sense that they result from an agreement among agents, and thus make the agents involved in the agreement better off.

The most natural notion of rationality in our definition of a swap is exchange-rationality, which requires that the two agents who exchange their matches are better off (Alcalde, 1994). A swap w.r.t. a pair of agents $\{i, j\}$ from matching μ is *exchange rational (ER)* if the agents who exchange their matches are better off, i.e.,

$$\mu(j) \succ_i \mu(i) \text{ and } \mu(i) \succ_j \mu(j). \tag{ER-swap}$$

Exchange-rationality is the only meaningful notion of swap rationality in housing markets because only one side of the market has preferences. However, several notions of rationality emerge in marriage and roommate markets, where agents are matched with each other. One could demand that only two of the agents who agree to form a new pair need to be better off. This notion of rational swaps is based on the classic idea of *blocking pairs*, which forms the basis of the standard notion of stability (Gale & Shapley, 1962). A swap w.r.t. a pair of agents $\{i, j\}$ from matching μ between agents is *blocking pair (BP) rational* if one of the new pairs in μ' forms a blocking pair, where both agents are better off, i.e.,

$$\left[\mu(j) \succ_i \mu(i) \text{ and } i \succ_{\mu(j)} j\right] \quad \text{or} \quad \left[\mu(i) \succ_j \mu(j) \text{ and } j \succ_{\mu(i)} i\right]. \tag{BP-swap}$$

We usually refer to a *BP*-swap by mentioning the associated blocking pair ($\{i, \mu(j)\}$ or $\{j, \mu(i)\}$). The old partners of the blocking pair are also assumed to be matched with each other.[2]

Finally, in marriage and roommate markets, a stronger notion of rationality is that of a *fully rational swap*, which makes all four involved agents better off. A swap w.r.t. a pair of agents $\{i, j\}$ from matching μ is *fully rational (FR)* if all four agents involved in the swap are better off, i.e.,

$$\mu(j) \succ_i \mu(i), \quad \mu(i) \succ_j \mu(j), \quad j \succ_{\mu(i)} i, \text{ and } \quad i \succ_{\mu(j)} j. \tag{\textit{FR}-swap}$$

Note that for marriage and roommate markets, an *FR*-swap w.r.t. a pair of agents $\{i, j\}$ from a matching μ is an *ER*-swap w.r.t. pair $\{i, j\}$ or $\{\mu(i), \mu(j)\}$ and also a *BP*-swap w.r.t blocking pair $\{i, \mu(j)\}$ or $\{j, \mu(i)\}$. We thus obtain the following implications:

$$\textit{BP}\text{-swap} \quad \Leftarrow \quad \textit{FR}\text{-swap} \quad \Rightarrow \quad \textit{ER}\text{-swap}$$

The different types of swap rationality are illustrated in the following example.

Example 5. *Consider a roommate market with six agents. The preferences of the agents are given below, where the initial assignment is marked with boxes.*

$$\begin{array}{lccccccccc}
1: & 4 & \succ & \boxed{3} & \succ & 6 & \succ & 5 & \succ & 2 \\
2: & 3 & \succ & 1 & \succ & \boxed{4} & \succ & 6 & \succ & 5 \\
3: & 6 & \succ & 2 & \succ & \boxed{1} & \succ & 5 & \succ & 4 \\
4: & 5 & \succ & 1 & \succ & 3 & \succ & \boxed{2} & \succ & 6 \\
5: & 2 & \succ & \boxed{6} & \succ & 4 & \succ & 1 & \succ & 3 \\
6: & 4 & \succ & 3 & \succ & 1 & \succ & 2 & \succ & \boxed{5}
\end{array}$$

The swap w.r.t. pair of agents $\{1, 2\}$, represented above by circles, which matches Agent 1 with Agent 4 and Agent 2 with Agent 3, is an FR-swap because every involved agent is better off. Hence, this is also an ER-swap for pair $\{1, 2\}$ or $\{3, 4\}$ because in both pairs the agents prefer to exchange their partners. It is also a BP-swap for blocking pair $\{2, 3\}$ or $\{1, 4\}$ because in both pairs the agents prefer to be together than with their current partner.

The swap w.r.t. pair of agents $\{1, 6\}$, represented above by stars, is a BP-swap for blocking pair $\{3, 6\}$ because Agent 3, the old partner of Agent 1, prefers to be with Agent 6, as well as Agent 6 who prefers Agent 3 to her old partner Agent 5. This is not an ER-swap (and hence not an FR-swap) because neither the agents in pair $\{1, 6\}$ nor in pair $\{3, 5\}$ want to exchange their partners.

The swap w.r.t. pair of agents $\{4, 6\}$, represented above by diamonds, is an ER-swap for pair $\{4, 6\}$ because Agent 4 prefers the current partner of Agent 6, i.e., Agent 5, to her current partner and Agent 6 prefers the current partner of Agent 4, i.e., Agent 2, to her current partner. This is not a BP-swap (and hence not an FR-swap) because it matches Agent 4 with Agent 5, who prefers to stay with her current partner, and Agent 6 with Agent 2, who prefers to stay with her current partner.

2. Once the old partners are alone, they have an incentive to form a new pair. Roth and Vande Vate (1990) therefore decompose *BP*-swaps into two steps. We do not explicitly consider these steps in order to always maintain a perfect matching (cf. Knuth (1976)).

Stability can now be defined according to the different notions of rational swaps. A matching μ is *σ-stable*, for $\sigma \in \{FR, ER, BP\}$, if no σ-swap can be performed from matching μ. A sequence of σ-swaps, for $\sigma \in \{FR, ER, BP\}$, corresponds to a sequence of matchings $(\mu^0, \mu^1, \ldots, \mu^r)$ such that a σ-swap transforms each matching μ^t into matching μ^{t+1} for every $0 \le t < r$. Then, matching μ is *σ-reachable* from initial matching μ^0 if there exists a sequence of σ-swaps $(\mu^0, \mu^1, \ldots, \mu^r)$ such that $\mu^r = \mu$. When the context is clear, we omit σ and the initial matching μ^0.

A *σ-dynamics* is defined according to initial matching μ^0 and a type σ of rational swaps. We say that the σ-dynamics converges if, starting from any initial matching μ^0, every sequence of σ-swaps terminates in a σ-stable matching.

In this article, we consider the following two decision problems related to the possibility and necessity, respectively, for the swap dynamics to eventually reach a Pareto-optimal matching.

$\exists$-σ-ParetoSequence	
Input:	Matching market $(N, (A_i)_{i\in N}, \succ, \mu^0)$, type σ of rational swaps
Question:	Does there exist a sequence of σ-swaps starting from initial matching μ^0 which terminates in a Pareto-optimal σ-stable matching?

$\forall$-σ-ParetoSequence	
Input:	Matching market $(N, (A_i)_{i\in N}, \succ, \mu^0)$, type σ of rational swaps
Question:	Do all sequences of σ-swaps starting from initial matching μ^0 terminate in a Pareto-optimal σ-stable matching?

In order to tackle these questions, we also study the stability and convergence properties of the considered dynamics in the three types of matching markets.

4. Exchange Rational Swaps

In housing markets, every *ER*-swap represents a Pareto improvement. Hence, since the number of agents and objects is finite, *ER*-dynamics always converges, and the existence of *ER*-stable matchings is guaranteed (simply because every Pareto-optimal matching happens to be *ER*-stable). However, it may be impossible to reach a Pareto-optimal matching from a given matching by only applying *ER*-swaps, as illustrated below (an observation already made by Damamme et al. (2015)).

Observation 5. *ER-dynamics may not converge to a Pareto-optimal matching in housing markets.*

Example 6. *Consider a housing market with $n \geq 3$ agents. The preferences of the agents are given below, where the initial assignment is marked with boxes, and $[\ldots]$ denotes an arbitrary order over the rest of the objects.*

$$
\begin{array}{lccccc}
1: & o_1 & \succ & \boxed{o_2} & \succ & [\ldots] \\
2: & o_2 & \succ & \boxed{o_3} & \succ & [\ldots] \\
3: & o_3 & \succ & \boxed{o_4} & \succ & [\ldots] \\
\vdots & & & \vdots & & \\
n: & o_n & \succ & \boxed{o_1} & \succ & [\ldots]
\end{array}
$$

Observe that no ER-swap is possible in this instance; therefore, the initial matching (boxed objects) is the unique ER-reachable matching. However, there exists a unique Pareto-optimal matching (circled objects), and this matching is different from the initial one. Note that, in such an instance, even if exchanges involving up to $n-1$ agents are allowed, the Pareto-optimal matching will not be reached: the only ER-exchange would involve all n agents.

Nevertheless, Damamme et al. (2015) have shown that *ER*-dynamics always converges to a Pareto-optimal matching in housing markets when the agents' preferences are single-peaked.

In marriage and roommate markets, an *ER*-stable matching may not exist, even for single-peaked preferences. This notably shows that a Pareto-optimal matching may not be stable (recall that our goal is the convergence to a Pareto-optimal matching and not only a Pareto-optimal matching, which may not end the swap dynamics). For the sake of clarity, we reproduce below the counterexamples for the existence of an *ER*-stable matching provided in the conclusion of Cechlárová (2002) for the marriage market and in Example 3.5 of Alcalde (1994) for the roommate market.

Example 7 (Alcalde (1994), Cechlárová (2002))**.** *Consider a marriage market with two women and two men. The preferences are given below and the initial assignment is marked with boxes.*

$$
\begin{array}{lccc}
w_1: & m_1 & \succ & \boxed{m_2} \\
w_2: & m_2 & \succ & \boxed{m_1}
\end{array}
\qquad\qquad
\begin{array}{lccc}
m_1: & \boxed{w_2} & \succ & w_1 \\
m_2: & \boxed{w_1} & \succ & w_2
\end{array}
$$

There are only two possible matchings: the encircled matching and the boxed matching. None of them is ER-stable because from the encircled matching, the two men can swap and from the boxed matching, the two women can swap. However, these two matchings are Pareto-optimal. Note that the preferences are trivially single-peaked.

Consider now a roommate market with four agents. The preferences of the agents are given below.

$$
\begin{array}{lccccc}
1: & 2 & \succ & \boxed{3} & \succ & 4 \\
2: & \boxed{4} & \succ & 1 & \succ & 3 \\
3: & \boxed{1} & \succ & 4 & \succ & 2 \\
4: & 3 & \succ & \boxed{2} & \succ & 1
\end{array}
$$

There are only three possible matchings: the encircled matching, the boxed matching and the diamond matching. None of them is ER-stable: from the diamond matching, all pairs of agents can swap their partners; from the encircled matching, agents 2 and 3 can swap their partners; and from the boxed matching, agents 1 and 4 can swap their partners. The encircled and the boxed matchings are both Pareto-optimal. The preferences are single-peaked with respect to the order: $3 > 2 > 1 > 4$.

Moreover, determining whether there exists an *ER*-stable matching has been shown to be NP-hard in both marriage and roommate markets (Cechlárová, 2002; Cechlárová & Manlove, 2005).

However, we prove that, for globally-ranked preferences, an *ER*-stable matching always exists, and, in addition, the convergence to such a matching is guaranteed.

Proposition 1. *ER-dynamics always converges in marriage and roommate markets for globally-ranked preferences.*

Proof. Denote by $\rhd$ the global order over all possible matched pairs in the market such that the preferences of the agents are globally-ranked with respect to this global order. Define as a potential function $f : \mu \rightarrow \mathbb{R}$ the function which assigns to each matching the sum of ranks in order $\rhd$ of all the assigned pairs in the matching, i.e., $f(\mu) = \sum_{\{i,j\} s.t. \mu(i)=j} rank_{\rhd}(\{i,j\}))$ with $rank_{\rhd}(\cdot)$ the function which gives the rank of a possible matched pair in order $\rhd$. Now consider a sequence of *ER*-swaps given by the sequence of matchings $(\mu^0, \mu^1, \ldots, \mu^r)$. Between each matchings μ^t and μ^{t+1}, with $0 \leq t < r$, an *ER*-swap is performed, say w.r.t. a pair of agents $\{i, j\}$. That means, by definition of an *ER*-swap, that agents i and j prefer to exchange their partners in μ^t, and thus, $\mu^t(j) \succ_i \mu^t(i)$ and $\mu^t(i) \succ_j \mu^t(j)$. This implies, by correlation of the preferences, that $\{i, \mu^t(j)\} \rhd \{i, \mu^t(i))\}$ and $\{j, \mu^t(i)\} \rhd \{j, \mu^t(j)\}$. But agents i and $\mu^t(j)$ are matched in μ^{t+1}, as well as agents j and $\mu^t(i)$. Since the rest of the pairs remains unchanged between μ^t and μ^{t+1}, we get that $f(\mu^{t+1}) < f(\mu^t)$. Because the number of different matchings is finite, we can conclude that *ER*-dynamics always converges. $\square$

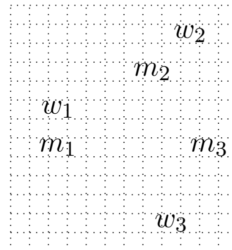
In general, an *ER*-stable matching may not be Pareto-optimal, and thus the convergence to a Pareto-optimal matching is not guaranteed, as shown in the next proposition, whose proof exhibits a counterexample which is similar to Example 6. This holds even when an *ER*-stable matching exists and under 2-Euclidean preferences, which are a preference restriction even stronger than globally-ranked preferences (see Figure 1).

Proposition 2. *ER-dynamics may not converge to a Pareto-optimal matching, in marriage and roommate markets, even when an ER-stable matching exists and for 2-Euclidean preferences.*

Proof. Consider a marriage market with three women and three men. The preferences are given below, and the initial assignment is marked with boxes.

$$
\begin{array}{lccccc}
w_1 : & m_1 & \succ & \boxed{m_2} & \succ & m_3 \\
w_2 : & m_2 & \succ & \boxed{m_3} & \succ & m_1 \\
w_3 : & m_3 & \succ & \boxed{m_1} & \succ & m_2
\end{array}
\qquad\qquad
\begin{array}{lccccc}
m_1 : & w_1 & \succ & \boxed{w_3} & \succ & w_2 \\
m_2 : & w_2 & \succ & \boxed{w_1} & \succ & w_3 \\
m_3 : & w_3 & \succ & \boxed{w_2} & \succ & w_1
\end{array}
$$

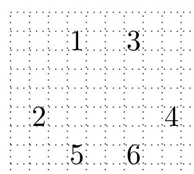
No *ER*-swap is possible from the initial matching (boxed agents), therefore the initial matching is the unique *ER*-reachable matching. However, there is another matching (circled agents) which is the unique Pareto-optimal matching. Note that this preference profile is globally-ranked with respect to, e.g., the global order $\{w_1, m_1\} \rhd \{w_2, m_2\} \rhd \{w_3, m_3\} \rhd \{w_1, m_2\} \rhd \{w_2, m_3\} \rhd \{w_3, m_1\} \rhd \{w_1, m_3\} \rhd \{w_2, m_1\} \rhd \{w_3, m_2\}$ over all possible matched pairs in the market and, even more, it is 2-Euclidean with respect to, e.g., the following planar embedding.



Now, consider a roommate market with six agents. Preferences of the agents are given below, where the initial partner of each agent is marked with boxes and $[\dots]$ denotes an arbitrary order over the rest of the agents.

$$\begin{array}{lccccc}
1: & 3 & \succ & \boxed{2} & \succ & [\dots] \\
2: & 5 & \succ & \boxed{1} & \succ & [\dots] \\
3: & 1 & \succ & \boxed{4} & \succ & [\dots] \\
4: & 6 & \succ & \boxed{3} & \succ & [\dots] \\
5: & 2 & \succ & \boxed{6} & \succ & [\dots] \\
6: & 4 & \succ & \boxed{5} & \succ & [\dots]
\end{array}$$

No *ER*-swap is possible from the initial matching (boxed agents), thus the initial matching is the unique *ER*-reachable matching. However, there is another matching (circled agents) which is the unique Pareto-optimal matching. Note that this preference profile is globally-ranked with respect to, e.g., the global order $\{4,6\} \rhd \{1,3\} \rhd \{3,4\} \rhd \{2,5\} \rhd \{1,2\} \rhd \{5,6\} \rhd [\dots]$ over all possible matched pairs and, even more, it is 2-Euclidean with respect to, e.g., the following planar embedding.



Both counterexamples are minimal because, in a smaller instance, a matching would be composed of exactly two pairs of agents, and thus, any swap would involve all the agents. Therefore, a matching which Pareto-dominates a stable matching would make every agent better off in comparison to the current stable one, and thus, there would exist an *ER*-swap, a contradiction. $\square$

Note that the above preference profiles are not 1-Euclidean. In fact, they are not even single-peaked. Again, more positive results can be obtained by restricting the domain of admissible preferences.

Proposition 3. *Every ER-stable matching is Pareto-optimal when preferences are single-peaked in marriage and roommate markets.*

Proof. Let μ be an *ER*-stable matching. For any two agents i and j (in N for roommate markets, or both in either W or M for marriage markets) it holds that $\mu(i) \succ_i \mu(j)$ or $\mu(j) \succ_j \mu(i)$. Suppose that μ is not Pareto-optimal, i.e., there is another matching μ' such that $\mu'(i) \succeq_i \mu(i)$ for every $i \in N$ and there exists $j \in N$ such that $\mu'(j) \succ_j \mu(j)$. Then, there exists a Pareto improving cycle from μ to μ' along agents $(n_1, \dots, n_k)$ such that each agent n_i, $1 \leq i \leq k$, is matched in μ' with agent $\mu(n_{(i \bmod k)+1})$. For marriage markets, the agents in $(n_1, \dots, n_k)$ are restricted by definition to only one side of the market, but it impacts both sides since the agents exchange agents of the other side. But there is no problem with the preferences of the matched agents because no agent is worse off in μ' compared to μ. The same holds for roommate markets. Since μ is *ER*-stable, it holds that $k > 2$. However, for single-peaked preferences, one can prove, by following the same proof by induction as Damamme et al. (2015)'s Proposition 1, that a Pareto improving cycle of any length cannot occur, contradicting the fact that μ is Pareto dominated. We reproduce below the main arguments of the proof, for the sake of self-containment. We will mainly use the property that every single-peaked preference profile is worst-restricted: for any triple of alternatives $X := \{x, y, z\}$, there always exists an alternative within X which is never ranked last by an agent when restricting the preference profile to X.

Let x_i denote the agent $\mu(n_i)$ matched with agent n_i in matching μ, for $1 \leq i \leq k$. By definition of the improving cycle, each agent n_i prefers $x_{(i \bmod k)+1}$ to x_i, for $1 \leq i \leq k$. Moreover, each agent $n_{(i \bmod k)+1}$ must prefer $x_{(i \bmod k)+1}$ to x_i, for $1 \leq i \leq k$, otherwise a possible swap would exist between agents n_i and $n_{(i \bmod k)+1}$. It follows that, for the base case with $k = 3$, we have the following preferences, which violate worst-restrictedness, a contradiction.

$$\begin{array}{ccccc} x_2 & \succ_{n_1} & x_1 & \succ_{n_1} & x_3 \\ x_3 & \succ_{n_2} & x_2 & \succ_{n_2} & x_1 \\ x_1 & \succ_{n_3} & x_3 & \succ_{n_3} & x_2 \end{array}$$

Assume now, by induction, that no improving cycle of length k' is possible for every $k' < k$ for a given k such that $3 < k \leq n$. Let us now suppose that there exists an improving cycle of length k. Every agent n_i must prefer x_i to any other agent x_j, for $1 \leq i \leq k$ with $j \in \{1, \dots, k\} \setminus \{i, (i \bmod k) + 1\}$, otherwise the agents $\{n_i, n_j, n_{j+1}, \dots, n_k, n_1, \dots, n_{i-1}\}$ would form an improving cycle of length strictly smaller than k, contradicting the induction assumption. Consider an arbitrary agent n_i within the improving cycle, with $1 \leq i \leq k$, and the agent, denoted by x_w, that n_i ranks last within $\{x_1, x_2, \dots, x_k\}$. By the previous observations, the agents n_{w-1} and n_w must have the following preferences: $x_w \succ_{n_{w-1}} x_{w-1} \succ_{n_{w-1}} x_{w+1}$ and $x_{w+1} \succ_{n_w} x_w \succ_{n_w} x_{w-1}$ (where $w - 1 = k$ if $w = 1$, and $w + 1 = 1$ if $w = k$). It follows that the three agents n_i, n_{w-1} and n_w violate worst-restrictedness for the three alternatives x_w, x_{w-1} and x_{w+1}, a contradiction. $\square$

Propositions 1 and 3 allow us to conclude that sequences of *ER*-swaps will always terminate in Pareto-optimal matchings when preferences are both single-peaked and globally-ranked, like in 1-Euclidean preferences.

Corollary 1. *ER-dynamics always converges to a Pareto-optimal matching in marriage and roommate markets for 1-Euclidean preferences.*

For more general preferences, an interesting computational question is whether, given a preference profile and an initial assignment, a Pareto-optimal matching can be reached via *ER*-swaps. In the context of housing markets, the complexity of this question was mentioned as an open problem by Damamme et al. (2015). It turns out that this problem is computationally intractable for all kinds of matching markets considered in this paper, even for globally-ranked preferences.

Theorem 1. $\exists$-ER-ParetoSequence *is* NP*-hard in housing, marriage, and roommate markets, even for globally-ranked preferences.*

Proof. Let us first consider the case of housing markets. We perform a reduction from 2P1N-SAT, a variant of SAT known to be NP-complete (Yoshinaka, 2005), where the goal is to decide the satisfiability of a CNF propositional formula where each variable appears exactly twice as a positive literal and once as a negative literal. The idea of the proof is close to the one given by Gourvès et al. (2017) for proving NP-hardness of determining whether a given object is reachable by a given agent. From an instance of 2P1N-SAT with formula φ on m clauses $C_1, \ldots, C_m$ and p variables $x_1, \ldots, x_p$, we build a housing market $(N, O, \succ, \mu^0)$ as follows.

For each clause C_j, with $1 \leq j \leq m$, we construct two clause-agents in N denoted by K_j and K'_j and two clause-objects in O denoted by k_j and k'_j such that $\mu^0(K_j) = k_j$ and $\mu^0(K'_j) = k'_j$. For each variable x_i, with $1 \leq i \leq p$, we construct six literal-agents in N corresponding to two copies of each literal, namely agents Y_i^ℓ and Z_i^ℓ who correspond to the ℓ^{th} ($\ell \in \{1, 2\}$) positive occurrence of variable x_i in formula φ, denoted by x_i^ℓ, and $\overline{Y_i}$ and $\overline{Z_i}$ who correspond to the negative occurrence of variable x_i in formula φ, denoted by $\overline{x_i}$; we also create their associated literal-objects y_i^ℓ, z_i^ℓ, $\overline{y_i}$ and $\overline{z_i}$ such that $\mu^0(Y_i^\ell) = y_i^\ell$, $\mu^0(Z_i^\ell) = z_i^\ell$, $\mu^0(\overline{Y_i}) = \overline{y_i}$ and $\mu^0(\overline{Z_i}) = \overline{z_i}$. The literal-agents are divided into two sets, denoted by Y and Z, which correspond to the original agents and their copy, respectively, i.e., $Y := \bigcup_{1 \leq i \leq p} \{Y_i^1, Y_i^2, \overline{Y}_i\}$ and $Z := \bigcup_{1 \leq i \leq p} \{Z_i^1, Z_i^2, \overline{Z}_i\}$. Three additional agents B, T and T' are created in N, with their initial assigned objects denoted by b, t, and t', respectively.

The preferences of the agents are given below for each $1 \leq i \leq p$ and $1 \leq j < m$ (notation $[\ldots]$ denotes an arbitrary order over the rest of the objects, $\{\mathbf{y}_j\}$ is an arbitrary order over the literal-objects in $\mathbf{y} := \bigcup_{1 \leq i \leq p} \{y_i^1, y_i^2, \overline{y_i}\}$ which are associated with literals of clause C_j, and $cl(\ell_i)$ is the index of the clause in which literal ℓ_i appears).

$$
\begin{array}{rl}
T: & t' \succ \{\mathbf{y}_1\} \succ \boxed{t} \succ [\ldots] \\
K_j: & k'_j \succ \{\mathbf{y}_{j+1}\} \succ t \succ \{\mathbf{y}_j\} \succ \boxed{k_j} \succ [\ldots] \\
K_m: & b \succ t \succ \{\mathbf{y}_m\} \succ \boxed{k_m} \succ [\ldots] \\
Y_i^1: & z_i^1 \succ k_{cl(\overline{x_i})} \succ k_{cl(x_i^1)} \succ \overline{y_i} \succ \boxed{y_i^1} \succ [\ldots] \\
Y_i^2: & z_i^2 \succ y_i^1 \succ k_{cl(x_i^2)} \succ \overline{y_i} \succ \boxed{y_i^2} \succ [\ldots] \\
\overline{Y}_i: & \overline{z_i} \succ y_i^2 \succ \boxed{\overline{y_i}} \succ [\ldots] \\
B: & t \succ \boxed{b} \succ [\ldots]
\end{array}
$$

$$
\begin{array}{rl}
T': & k'_m \succ \{\mathbf{y}_1\} \succ \boxed{t'} \succ [\ldots] \\
K'_j: & k_j \succ \{\mathbf{y}_j\} \succ k'_m \succ \{\mathbf{y}_{j+1}\} \succ \boxed{k'_j} \succ [\ldots] \\
K'_m: & k_m \succ \{\mathbf{y}_m\} \succ \boxed{k'_m} \succ [\ldots] \\
Z_i^1: & y_i^1 \succ \overline{y_i} \succ k_{cl(x_i^1)} \succ k_{cl(\overline{x_i})} \succ \boxed{z_i^1} \succ [\ldots] \\
Z_i^2: & y_i^2 \succ \overline{y_i} \succ y_i^1 \succ k_{cl(x_i^2)} \succ \boxed{z_i^2} \succ [\ldots] \\
\overline{Z}_i: & \overline{y_i} \succ y_i^2 \succ \boxed{\overline{z_i}} \succ [\ldots]
\end{array}
$$

We claim that the formula φ is satisfiable if and only if the matching assigning to each agent her best object is reachable (this is the only Pareto-optimal matching). The global idea of the reduction is that the only way to reach this Pareto-optimal matching is to make object t reach agent K_m by first giving to each clause-agent K_j, via *ER*-swaps, a literal-object in $\{\mathbf{y}_j\}$, objects associated with the literals of clause C_j. Once object t reaches clause-agent K_m, each agent except K'_m exchanges with her prime version agent (agents

Z_i^1, Z_i^2, and $\overline{Z_i}$ are the prime versions of agents Y_i^1, Y_i^2, and $\overline{Y_i}$, respectively, and B is the prime version of K_m), and then the prime agents make among them the reverse sequence of swaps of the initial one where the goal was to make object t reach K_m, leading to the Pareto-optimal matching. By construction of the preferences among the literal-agents, once a literal-object associated with a positive (resp., negative) literal of a variable has been chosen to go with a clause-agent K_j, no literal-object associated with a negative (resp., positive) literal of this variable can reach a clause-agent. An illustration of the sequence of swaps that needs to occur is given in Figure 2.

Let us first assume that the formula φ is satisfiable with truth assignment ϕ. For each clause C_j, let us choose an arbitrary literal of clause C_j which is true in ϕ (such a literal must exist by satisfiability assumption) and denote by L_j and ℓ_j its associated literal-agent and literal-object, respectively. By construction, ℓ_j belongs to $\{\mathbf{y}_j\}$ and L_j initially owns ℓ_j. For each clause C_j, if L_j corresponds to a positive literal, then L_j directly exchanges object ℓ_j with agent K_j, otherwise, i.e., if $L_j = \overline{Y_i}$ for a given i such that $1 \leq i \leq p$, agent L_j first exchanges her object with agent Y_i^2 who then exchanges it with agent Y_i^1 who then exchanges it with agent K_j. Observe that all these swaps are exchange rational and eventually give to each agent K_j an object associated with one of the literals of C_j which are true in ϕ. At this point, we perform the sequence of swaps between the agents $\{T, K_1\}, \{K_1, K_2\}, \ldots, \{K_{m-1}, K_m\}$, which are exchange rational and lead to give object t to agent K_m. Now, let agents T, K_1, K_2, $\ldots$, and K_{m-1} exchange with their respective prime version agent, i.e., agent T exchange with T', K_1 with K_1' and so on, as well as literal-agents in Y with their associated agent in Z, i.e., for every $1 \leq i \leq p$, agent Y_i^ℓ exchange with agent Z_i^ℓ, for $\ell \in \{1, 2\}$, and agent $\overline{Y_i}$ exchange with agent $\overline{Z_i}$. Moreover, let agent K_m exchange with agent B. Observe that all these swaps are exchange rational and lead to give to each agent in $\{T, K_1, \ldots, K_m, B\} \cup Y$ her best object. This is also the case for agents in Z for which the associated agent in Y is not a chosen L_j (or does not correspond to the positive literal of a chosen negative L_j). We then reproduce the first sequence of swaps leading to give object t to agent K_m but in reverse order and over prime-version agents. Therefore, we get that agent T' obtains object k_m', and each agent K_j' first gets object ℓ_j and then, since each agent K_j' exchanges with the literal-agent in Z who corresponds to the literal-agent in Y who has participated in the first sequence of swaps to move object ℓ_j, we get that each agent K_j' gets her best object, as well as agents in Z, because their best object is the initial object of their associated agent in Y. Finally, we have reached a matching where every agent is assigned her best object. For a description of the different steps of the sequence of swaps, see Figure 2.

Let us now assume that the matching assigning to each agent her best object is reachable. That means that agent B obtains object t, but the only way for her to get it in *ER*-swaps is via agent K_m, because no other agent prefers object b to her initial object. Therefore, object t, which is initially owned by agent T must reach agent K_m. By construction of the preferences of the agents, this can only be done via all the agents K_1, K_2, $\ldots$, K_{m-1}, who must have previously gotten an object within $\{\mathbf{y}_j\}$ for each K_j. This can only be done by exchanging with agents in Y. Observe that a literal-object in $\mathbf{y}$ corresponding to a negative literal, say $\overline{y_i}$, can reach an agent K_j only via agent Y_i^2 and then agent Y_i^1. Therefore, by exchange rationality of the swaps and construction of the preferences, if a literal-object in $\mathbf{y}$ corresponding to a positive literal of a variable is chosen for reaching a clause-agent,

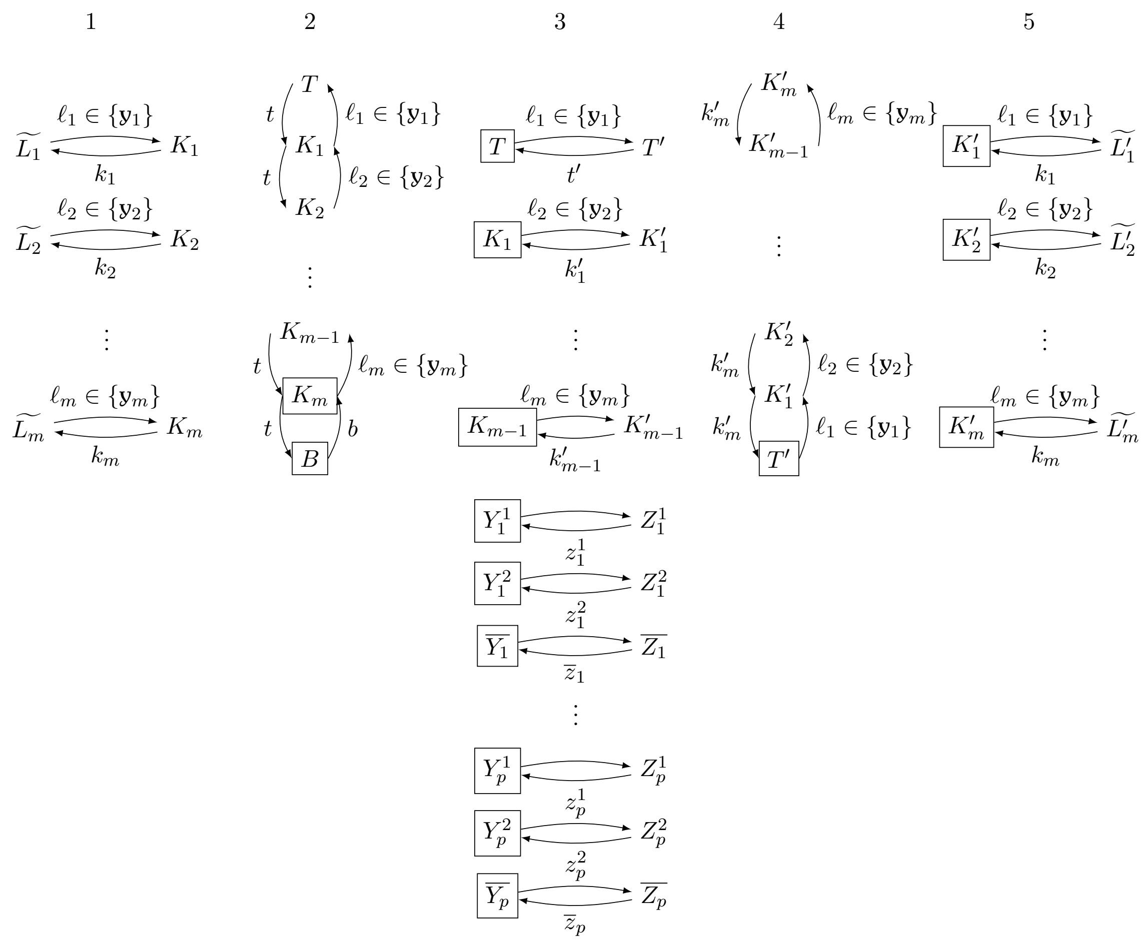


Figure 2: Description of the sequence of swaps in the complexity proof of Theorem 1 for reaching the only Pareto-optimal matching. Nodes are agents and arrows represent swaps where the exchanged objects are denoted along the arrows. Agents are boxed when they obtain their most preferred object. Agent $\widetilde{L_j}$ (resp., $\widetilde{L'_j}$) refers directly to the literal-agent in Y (resp., in Z) who is associated with a chosen literal of clause C_j if this literal is positive, or to literal-agent Y_i^1 (resp., Z_i^1) if this literal is negative of index i. In the latter case, before swaps described at global step 1, agent $\overline{Y_i}$ exchanges literal-object ℓ_j with agent Y_i^2 who then exchanges it with agent Y_i^1. Conversely, in the same case, after global step 5, agent Z_i^1 exchanges literal-object ℓ_1 with agent Z_i^2 who then exchanges it with agent $\overline{Z_i}$.

then the negative version of this object can never reach a clause-agent, and vice versa if the negative version is chosen. It follows that we must have chosen one literal-object per clause-agent and that, among all the chosen literal-objects, there cannot be two of them which correspond to opposite literals. Hence, by setting to true all the literals corresponding to chosen literal-objects, we get a truth assignment of the variables which satisfies formula φ.

To adapt the proof to the case of marriage and roommate markets, we now consider the objects in O as agents. More precisely, we build a marriage market $(N, \succ, \mu^0)$ where $N = M \cup W$ with $W = \{T, T', B, \{K_j, K'_j\}_{1 \le j \le m}, \{Y_i^1, Y_i^2, \overline{Y_i}, Z_i^1, Z_i^2, \overline{Z_i}\}_{1 \le i \le p}\}$ and $M = \{t, t', b, \{k_j, k'_j\}_{1 \le j \le m}, \{y_i^1, y_i^2, \overline{y_i}, z_i^1, z_i^2, \overline{z_i}\}_{1 \le i \le p}\}$. The preferences of women over men are the same as the preferences of agents over objects previously described, and the preferences of men over women are described below for $1 \le j < m$ and $1 \le i \le p$. The notation $\{\mathcal{Y}_j\}$ (resp., $\{\mathcal{Z}_j\}$) denotes an arbitrary order over the subset of literal-agents in Y (resp., Z) that are associated with a literal of clause C_j where each "negative" literal-agent $\overline{Y}_i$ (resp., $\overline{Z}_i$) is replaced by agent Y_i^1 (resp., Z_i^1). Moreover, $K_0 = T$, $K'_0 = T'$, and $[\dots]$ denotes an arbitrary order over the rest of the women.

$$
\begin{array}{rl}
t: & B \succ K_m \succ \cdots \succ K_1 \succ \boxed{T} \succ [\dots] \\
k_j: & K'_j \succ \{\mathcal{Z}_j\} \succ \{\mathcal{Y}_j\} \succ \boxed{K_j} \succ [\dots] \\
k_m: & K'_m \succ \{\mathcal{Z}_m\} \succ \{\mathcal{Y}_m\} \succ \boxed{K_m} \succ [\dots] \\
y_i^1: & Z_i^1 \succ Z_i^2 \succ Y_i^2 \succ K'_{cl(x_i^1)} \succ K'_{cl(x_i^1)-1} \succ K_{cl(x_i^1)-1} \succ K_{cl(x_i^1)} \succ \boxed{Y_i^1} \succ [\dots] \\
y_i^2: & Z_i^2 \succ \overline{Z_i} \succ \overline{Y_i} \succ K'_{cl(x_i^2)} \succ K'_{cl(x_i^2)-1} \succ K_{cl(x_i^2)-1} \succ K_{cl(x_i^2)} \succ \boxed{Y_i^2} \succ [\dots] \\
\overline{y_i}: & \overline{Z_i} \succ Z_i^2 \succ Z_i^1 \succ K'_{cl(\overline{x_i})} \succ K'_{cl(\overline{x_i})-1} \succ K_{cl(\overline{x_i})-1} \succ K_{cl(\overline{x_i})} \succ Y_i^1 \succ Y_i^2 \succ \boxed{\overline{Y_i}} \succ [\dots] \\
\hline
t': & T \succ \boxed{T'} \succ [\dots] \\
k'_j: & K_j \succ \boxed{K'_j} \succ [\dots] \\
k'_m: & T' \succ K'_1 \succ \cdots \succ \boxed{K'_m} \succ [\dots] \\
z_i^1: & Y_i^1 \succ \boxed{Z_i^1} \succ [\dots] \\
z_i^2: & Y_i^2 \succ \boxed{Z_i^2} \succ [\dots] \\
\overline{z_i}: & \overline{Y_i} \succ \boxed{\overline{Z_i}} \succ [\dots] \\
\hline
b: & K_m \succ \boxed{B} \succ [\dots]
\end{array}
$$

For roommate markets, we consider the same market but without distinguishing between men and women. The only difference in the preferences is that $[\dots]$ is an arbitrary order over all the rest of the agents. In such a way, there is no incentive to partner with an agent who belongs to the other side of the marriage market.

Note that the preferences are globally ranked with respect to, e.g., the following global order over all possible matched pairs in the market: $\{t, B\} \rhd \{b, K_m\} \rhd \{t, K_m\} \rhd \{b, B\} \rhd \{t', T\} \rhd \{k'_m, T'\} \rhd \{k_j, K'_j\}_{\forall j} \rhd \{k'_j, K_j\}_{\forall j < m} \rhd \{\overline{z_i}, \overline{Y_i}\}_{\forall i} \rhd \{\overline{y_i}, \overline{Z_i}\}_{\forall i} \rhd \{z_i^\ell, Y_i^\ell\}_{\ell \in \{1,2\}, \forall i} \rhd \{y_i^\ell, Z_i^\ell\}_{\ell \in \{1,2\}, \forall i} \rhd \{y_i^2, \overline{Z_i}\}_{\forall i} \rhd \{y_i^2, \overline{Y_i}\}_{\forall i} \rhd \{\overline{y_i}, Z_i^2\}_{\forall i} \rhd \{\overline{y_i}, Z_i^1\}_{\forall i} \rhd \{\overline{z_i}, \overline{Z_i}\}_{\forall i} \rhd \{y_i^1, Z_i^2\}_{\forall i} \rhd \{y_i^1, Y_i^2\}_{\forall i} \rhd \{\{\mathsf{y}_j\}, K'_j\}_{\forall j} \rhd \{k'_m, K'_1\} \rhd \cdots \rhd \{k'_m, K'_m\} \rhd \{\{\mathsf{y}_j\}, K'_{j-1}\}_{\forall j} \rhd \{t', T'\} \rhd \{k'_j, K'_j\}_{\forall j < m} \rhd \{k_{cl(x_i^1)}, Z_i^1\}_{\forall i} \rhd \{k_{cl(\overline{x_i})}, Z_i^1\}_{\forall i} \rhd \{k_{cl(x_i^2)}, Z_i^2\}_{\forall i} \rhd \{z_i^\ell, Z_i^\ell\}_{\ell \in \{1,2\}, \forall i} \rhd \{\{\mathsf{y}_j\}, K_{j-1}\}_{\forall j} \rhd \{t, K_{m-1}\} \rhd \cdots \rhd \{t, K_1\} \rhd \{t, T\} \rhd \{\{\mathsf{y}_j\}, K_j\}_{\forall j} \rhd$

$\{k_{cl(\overline{x_i})}, Y_i^1\}_{\forall i} \rhd \{k_{cl(x_i^1)}, Y_i^1\}_{\forall i} \rhd \{k_{cl(x_i^2)}, Y_i^2\}_{\forall i} \rhd \{k_j, K_j\}_{\forall j} \rhd \{\overline{y_i}, Y_i^1\}_{\forall i} \rhd \{\overline{y_i}, Y_i^2\}_{\forall i} \rhd \{y_i^\ell, Y_i^\ell\}_{\ell \in \{1,2\}, \forall i} \rhd \{\overline{y_i}, \overline{Y_i}\}_{\forall i} \rhd$ [arbitrary order over the rest of pairs], where an arbitrary order can be chosen when a set of pairs is mentioned.

Given these preferences for both marriage and roommate markets, a swap is rational for one side of the market if and only if it is also rational for the other side (in the constructed roommate market, no *ER*-swap can occur between two agents who were from two different sides in the marriage market). In other words, the set of *ER*-swaps is identical to the set of *FR*-swaps. Indeed, after any (possibly empty) sequence of swaps, one can observe that for any agent i, all the agents that are preferred to the current partner of agent i also prefer agent i to their current partner. Hence, the sequences of swaps that may occur are exactly the same as in the proof for housing markets, which completes the proof. □

For housing markets, the size of a sequence of *ER*-swaps is bounded by $\mathcal{O}(n^2)$ because every agent involved in the swap is strictly better off. Therefore, since checking the Pareto-optimality of a matching can be done in polynomial time, we also get membership in NP.

Corollary 2. ∃-ER-ParetoSequence *is* NP*-complete in housing markets, even for globally-ranked preferences.*

Note that the model of swap dynamics in housing markets used by Gourvès et al. (2017) takes into account an underlying graph structure where two agents can exchange their objects only if they are connected in the graph. They notably investigate the complexity of the problems of deciding whether a given agent can get a given object via a sequence of swaps and whether a given assignment is reachable by a sequence of swaps. They left open the complexity of these questions on a complete graph, but the proof of Theorem 1 shows that both problems are actually intractable.

Corollary 3. *In housing markets, determining whether a given agent can get a given object via a sequence of ER-swaps or whether a given matching can be reached via a sequence of ER-swaps is* NP*-hard.*

By slightly modifying the proof of Theorem 1, we can also show that it is NP-hard to detect a cycle in the *ER*-dynamics for marriage and roommate markets.[3] It follows that determining convergence is hard for markets that match agents with other agents.

Theorem 2. *Determining whether ER-dynamics is guaranteed to converge is* co-NP*-hard in marriage and roommate markets.*

Proof. We show that determining whether *ER*-dynamics can cycle in marriage and roommate markets is NP-hard. Consider the instance of a marriage market used in the proof of Theorem 1 and change the preferences of agents t and b by swapping the positions of agents B and K_m in their preference rankings. The formula of the 2P1N-SAT instance is satisfiable if and only if agent t is reachable for agent K_m through a configuration where agents K_m and B can exchange their partners within an *ER*-swap whereas their partners can then perform an *ER*-swap to come back to the previous matching. We thus get a cycle

3. Recall that there is no possibility of cycles in the *ER*-dynamics for housing markets. Indeed, the dynamics is proven to converge since an *ER*-swap makes every involved agent better off in housing markets.

in the *ER*-dynamics. This is the only possible cycle since all the other *ER*-swaps are *FR*-swaps; that is, they make all the agents involved in the swaps better off. The same proof adaptation works for the roommate market built in the proof of Theorem 1. $\square$

Nevertheless, the problem of convergence to a Pareto-optimal matching is not only due to convergence issues. For preferences more general than those restricted to the 1-Euclidean domain, recognizing the instances where *ER*-dynamics is guaranteed to converge to a Pareto-optimal matching is intractable, even when *ER*-dynamics always converges to a stable matching.

Theorem 3. ∀-ER-ParetoSequence *is* co-NP-*hard in housing, marriage, and roommate markets, even for globally-ranked preferences.*

Proof. The proof is very similar to the one presented in Theorem 1. We first consider the case of housing markets. We provide a reduction to the problem of deciding the existence of a sequence of *ER*-swaps terminating in a matching which is not Pareto-optimal. From an instance of 2P1N-SAT (Yoshinaka, 2005) with formula φ on m clauses and p variables, we build a housing market $(N, O, \succ, \mu^0)$ where the set of agents N and the set of objects O are the same as in the proof of Theorem 1, except that agent B and object b are not present. The preferences are only slightly different but, for the sake of clarity, we explicitly present them below, for $1 \leq j < m$ and $1 \leq i \leq p$ (the notations are the same as in the proof of Theorem 1).

$$
\begin{array}{rl|rl}
T: & t' \succ \{\mathsf{y}_1\} \succ \boxed{t} \succ [\dots] & T': & t \succ \{\mathsf{y}_1\} \succ \boxed{t'} \succ [\dots] \\
K_j: & k'_j \succ \{\mathsf{y}_{j+1}\} \succ t \succ \{\mathsf{y}_j\} \succ \boxed{k_j} \succ [\dots] & K'_j: & k_j \succ \{\mathsf{y}_j\} \succ t \succ \{\mathsf{y}_{j+1}\} \succ \boxed{k'_j} \succ [\dots] \\
K_m: & k'_m \succ t \succ \{\mathsf{y}_m\} \succ \boxed{k_m} \succ [\dots] & K'_m: & k_m \succ \{\mathsf{y}_m\} \succ \boxed{k'_m} \succ [\dots] \\
Y_i^1: & z_i^1 \succ k_{cl(\overline{x_i})} \succ k_{cl(x_i^1)} \succ \overline{y_i} \succ \boxed{y_i^1} \succ [\dots] & Z_i^1: & y_i^1 \succ \overline{y_i} \succ k_{cl(x_i^1)} \succ k_{cl(\overline{x_i})} \succ \boxed{z_i^1} \succ [\dots] \\
Y_i^2: & z_i^2 \succ y_i^1 \succ k_{cl(x_i^2)} \succ \overline{y_i} \succ \boxed{y_i^2} \succ [\dots] & Z_i^2: & y_i^2 \succ \overline{y_i} \succ k_{cl(x_i^2)} \succ y_i^1 \succ \boxed{z_i^2} \succ [\dots] \\
\overline{Y_i}: & \overline{z_i} \succ y_i^2 \succ \boxed{\overline{y_i}} \succ [\dots] & \overline{Z_i}: & \overline{y_i} \succ y_i^2 \succ \boxed{\overline{z_i}} \succ [\dots]
\end{array}
$$

We claim that there exists a sequence of *ER*-swaps which does not terminate in a Pareto-optimal matching if and only if formula φ is satisfiable. Observe that there exists a unique Pareto-optimal matching which is the matching assigning to every agent her most preferred object.

Suppose first that formula φ is satisfiable. Then, let us show that there exists a sequence of *ER*-swaps which does not terminate in the matching where every agent gets her best object. Following the arguments of the proof of Theorem 1, this is possible to make object t reach agent K_m, where object t is the second most preferred object of agent K_m. Afterwards, agent K_m cannot make any further *ER*-swap because no agent prefers object t to her current assigned object, except possibly agents K'_j for $1 \leq j < m$ and agent T but they cannot get object k'_m (which is less preferred than their initial object) which is the only one that agent K_m prefers to t. Note that all the clause-agents K_j get an object that they prefer to object t since they were all involved in the sequence of swaps leading to passing object t from agent T to agent K_m. Therefore, the associated sequence of swaps terminates in a matching where agent K_m can only get her second-most preferred object, and thus it is not Pareto-optimal.

Suppose now that formula φ is not satisfiable. According to the arguments of the proof of Theorem 1, then this is not possible to make agent K_m get object t. By construction of the preferences, the sequence of *ER*-swaps between only agents and their prime versions (recall that the prime version of an agent in Y is an agent in Z with the same indices) trivially leads to give to every agent her best object. Note that a "prime" agent cannot make swaps with another prime agent as long as she did not exchange with an "original" agent. Moreover, this original agent must be her associated prime agent because, among the prime objects, the original agents only prefer the object of their associated prime agent over their initial object. Note that, after this swap, the original agents get their most preferred object and thus are not involved in further swaps. Whatever the object owned by an original agent, a swap can occur with her associated prime agent because the prime agent has not made previous swaps (as noted before) and the prime agent has reversed preferences compared to the ones of the original agent, except for agents K_m and K'_m who have preferences which differ only with object t (K'_m does not prefer it to her initial object) but this is not a problem since object t cannot be reached by agent K_m. Therefore, at the end of any sequence of swaps, each original agent gets her most preferred object. For prime agents, if after the swap with their associated original agent, they do not get their most preferred object, this is because their associated original agent made previous swaps with other original agents. By reproducing the reverse sequence of swaps among the prime agents, they get the initial object of their associated agent which is their most preferred object. Note that at this point, there is no other sequence of swaps which can be applied without leading to give to each prime agent her most preferred object. Indeed, if it were the case, then, by construction of the preferences, there would still be a possible swap. Hence, all sequences of *ER*-swaps terminate in the Pareto-optimal matching.

To adapt the proof to the case of marriage and roommate markets, we now consider the objects in O as agents. More precisely, we build a marriage market $(N, \succ, \mu^0)$ where $N = M \cup W$ with $W = \{T, T', \{K_j, K'_j\}_{1\leq j\leq m}, \{Y_i^1, Y_i^2, \overline{Y_i}, Z_i^1, Z_i^2, \overline{Z_i}\}_{1\leq i\leq n}\}$ and $M = \{t, t', \{k_j, k'_j\}_{1\leq j\leq m}, \{y_i^1, y_i^2, \overline{y_i}, z_i^1, z_i^2, \overline{z_i}\}_{1\leq i\leq n}\}$. The preferences of women over men are the same as the preferences of agents over objects previously described, and the preferences of men over women are described below for $1 \leq j \leq m$ and $1 \leq i \leq p$ (the notations are the same as in the proof of Theorem 1).

$$
\begin{array}{rl|rl}
t: & T' \succ K'_1 \succ \cdots \succ K'_{m-1} \succ K_m \succ \cdots \succ K_1 \succ \boxed{T} \succ [\ldots] & t': & T \succ \boxed{T'} \succ [\ldots] \\
k_j: & K'_j \succ \{\mathcal{Z}_j\} \succ \{\mathcal{Y}_j\} \succ \boxed{K_j} \succ [\ldots] & k'_j: & K_j \succ \boxed{K'_j} \succ [\ldots] \\
y_i^1: & Z_i^1 \succ Z_i^2 \succ Y_i^2 \succ K'_{cl(x_i^1)} \succ K'_{cl(x_i^1)-1} \succ K_{cl(x_i^1)-1} \succ K_{cl(x_i^1)} \succ \boxed{Y_i^1} \succ [\ldots] & z_i^1: & Y_i^1 \succ \boxed{Z_i^1} \succ [\ldots] \\
y_i^2: & Z_i^2 \succ \overline{Z_i} \succ \overline{Y_i} \succ K'_{cl(x_i^2)} \succ K'_{cl(x_i^2)-1} \succ K_{cl(x_i^2)-1} \succ K_{cl(x_i^2)} \succ \boxed{Y_i^2} \succ [\ldots] & z_i^2: & Y_i^2 \succ \boxed{Z_i^2} \succ [\ldots] \\
\overline{y_i}: & \overline{Z_i} \succ Z_i^2 \succ Z_i^1 \succ K'_{cl(\overline{x_i})} \succ K'_{cl(\overline{x_i})-1} \succ K_{cl(\overline{x_i})-1} \succ K_{cl(\overline{x_i})} \succ Y_i^1 \succ Y_i^2 \succ \boxed{\overline{Y_i}} \succ [\ldots] & \overline{z_i}: & \overline{Y_i} \succ \boxed{\overline{Z_i}} \succ [\ldots]
\end{array}
$$

For roommate markets, we consider the same market but without distinguishing between men and women. The only difference in the preferences is that $[\ldots]$ is an arbitrary order over all the rest of the agents. In such a way, there is no incentive to partner with an agent who belongs to the other side of the marriage market.

Note that the preferences are globally ranked with respect to, e.g., the following global order over all possible matched pairs in the market: $\{t, T'\} \rhd \{k_j, K'_j\}_{\forall j} \rhd \{y_i^\ell, Z_i^\ell\}_{\ell\in\{1,2\},\forall i} \rhd \{\overline{y_i}, \overline{Z_i}\}_{\forall i} \rhd \{t', T\} \rhd \{k'_j, K_j\}_{\forall j} \rhd \{z_i^\ell, Y_i^\ell\}_{\ell\in\{1,2\},\forall i} \rhd \{\overline{z_i}, \overline{Y_i}\}_{\forall i} \rhd \{\overline{y_i}, Z_i^2\}_{\forall i} \rhd \{\overline{y_i}, Z_i^1\}_{\forall i} \rhd$

$\{y_i^2, \overline{Z_i}\}_{\forall i} \rhd \{y_i^2, \overline{Y_i}\}_{\forall i} \rhd \{k_{cl(x_i^\ell)}, Z_i^\ell\}_{\ell\in\{1,2\},\forall i} \rhd \{k_{cl(\overline{x_i})}, Z_i^1\}_{\forall i} \rhd \{y_i^1, Z_i^2\}_{\forall i} \rhd \{y_i^1, Y_i^2\}_{\forall i} \rhd \{k_{cl(\overline{x_i})}, Y_i^1\}_{\forall i} \rhd \{k_{cl(x_i^\ell)}, Y_i^\ell\}_{\forall i} \rhd \{\{\mathrm{y_j}\}, K_j'\}_{\forall j} \rhd \{t, K_1'\} \rhd \cdots \rhd \{t, K_{m-1}'\} \rhd \{\{\mathrm{y_j}\}, K_{j-1}'\}_{\forall j} \rhd \{\{\mathrm{y_j}\}, K_{j-1}\}_{\forall j} \rhd \{t, K_m\} \rhd \cdots \rhd \{t, K_1\} \rhd \{\{\mathrm{y_j}\}, K_j\}_{\forall j} \rhd \{\overline{y_i}, Y_i^1\}_{\forall i} \rhd \{\overline{y_i}, Y_i^2\}_{\forall i} \rhd \{t, T\} \rhd \{k_j, K_j\}_{\forall j} \rhd \{y_i^\ell, Y_i^\ell\}_{\ell\in\{1,2\},\forall i} \rhd \{\overline{y_i}, \overline{Y_i}\}_{\forall i} \rhd \{t', T'\} \rhd \{k_j', K_j'\}_{\forall j} \rhd \{z_i^\ell, Z_i^\ell\}_{\ell\in\{1,2\},\forall i} \rhd \{\overline{z_i}, \overline{Z_i}\}_{\forall i} \rhd$ [arbitrary order over the rest of pairs], where an arbitrary order can be chosen when a set of pairs is mentioned.

Given these preferences for both marriage and roommate markets, a swap is rational for one side of the market if and only if it is also rational for the other side (in the constructed roommate market, no *ER*-swap can occur between two agents who were from two different sides in the marriage market). In other words, the set of *ER*-swaps is identical to the set of *FR*-swaps. Indeed, after any sequence of swaps (even possibly empty), one can observe that for any agent i, all the agents that are preferred to the current partner of agent i also prefer agent i to their current partner. Hence, the sequences of swaps that may occur are exactly the same as in the proof for housing markets, which completes the proof. □

Since the size of a sequence of *ER*-swaps is bounded for housing markets and checking Pareto-optimality can be done in polynomial time, we get membership in co-NP.

Corollary 4. ∀-ER-ParetoSequence *is* co-NP*-complete in housing markets, even for globally-ranked preferences.*

5. Blocking Pair Swaps

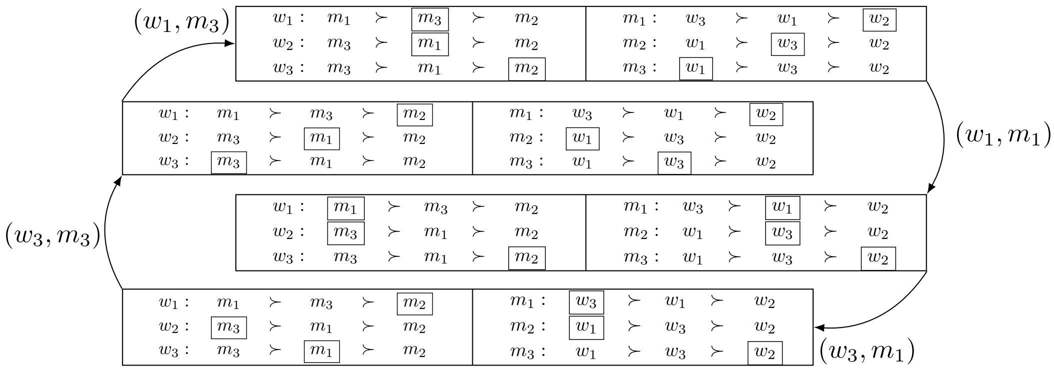
BP-swaps cannot occur in housing markets because objects can never be better off. We therefore focus in this section on matching markets that match agents with other agents.

First, by definition of a blocking pair, any *BP*-stable matching is Pareto-optimal. Indeed, consider a stable matching μ which is Pareto-dominated by matching μ'. Then, there exists an agent i such that $\mu'(i) \succ_i \mu(i)$. Since no agent is worse off in μ' compared to μ, it follows that for agent j such that $j = \mu'(i)$, it holds that $\mu'(j) \succeq_j \mu(j)$. But $\mu'(j) \neq \mu(j)$ because $i = \mu'(j)$ is strictly better off in μ'. So, (i, j) forms a blocking pair for matching μ, which contradicts its stability.

Observation 6. *Every BP-stable matching is Pareto-optimal.*

Moreover, a *BP*-stable matching always exists in marriage markets by the Deferred Acceptance algorithm (Gale & Shapley, 1962). However, the convergence to such a state is not guaranteed, as illustrated by Knuth (1976). The counterexample provided by Knuth (1976) can be completed in such a way that it holds even for single-peaked preferences. For the sake of clarity, we provide below a similar counterexample with a cycle in *BP*-dynamics for single-peaked preferences.

Example 8. *Consider a marriage market with three women and three men. The preferences of the agents are presented below, with the illustration of a cycle in BP-dynamics. The current matching is marked with boxed agents. The deviations are represented by arcs labeled with the blocking pair.*



Note that the preferences of the agents are single-peaked with respect to, e.g., the linear order $w_3 > w_1 > w_2 > m_3 > m_1 > m_2$. *Nevertheless, in this instance, there is a path of blocking pair swaps, which leads to a stable state. More precisely, by executing the swaps defined by the following sequence of blocking pairs,* $(\{w_1, m_1\}, \{w_3, m_1\}, \{w_1, m_3\})$, *we reach the matching assigning man* m_3 *to woman* w_1, *man* m_2 *to woman* w_2, *and man* m_1 *to woman* w_3, *which is BP-stable.*

Nevertheless, when old partners are not matched with each other, there always exists a sequence of *BP*-swaps leading to a stable matching (Roth & Vande Vate, 1990). However, when old partners are matched with each other, like in our setting, there exist marriage markets where no *BP*-stable matching can be reached from some initial matching (Tamura, 1993; Tan & Su, 1995), as reproduced below.[4]

Example 9. *Consider a marriage market with four women and four men. The preferences are given below and the initial assignment is marked with boxes.*

$w_1:$	$\boxed{m_1}$	$\succ$	m_3	$\succ$	m_2	$\succ$	m_4		$m_1:$	w_2	$\succ$	w_4	$\succ$	$\boxed{w_1}$	$\succ$	w_3
$w_2:$	$\boxed{m_2}$	$\succ$	m_4	$\succ$	m_3	$\succ$	m_1		$m_2:$	w_3	$\succ$	w_1	$\succ$	$\boxed{w_2}$	$\succ$	w_4
$w_3:$	m_3	$\succ$	m_1	$\succ$	$\boxed{m_4}$	$\succ$	m_2		$m_3:$	$\boxed{w_4}$	$\succ$	w_2	$\succ$	w_3	$\succ$	w_1
$w_4:$	m_4	$\succ$	m_2	$\succ$	m_1	$\succ$	$\boxed{m_3}$		$m_4:$	w_1	$\succ$	$\boxed{w_3}$	$\succ$	w_4	$\succ$	w_2

One can show that no BP-stable matching can be reached in this instance when starting from the initial matching given by boxes. At each step of the BP-dynamics, exactly one BP-swap is possible, implying that there is a unique possible execution of the dynamics. This unique execution of the BP-dynamics eventually assigns every woman with every possible man and comes back to the initial given assignment, forming a cycle.

In roommate markets, even the existence of a *BP*-stable matching is not guaranteed (Gale & Shapley, 1962), and this remains true even for single-peaked preferences. Nevertheless, checking the existence of a stable matching in a roommate market can be done in polynomial time (Irving, 1985) and, like in marriage markets, there always exists a sequence of *BP*-swaps leading to a stable matching when there exists one (Diamantoudi, Miyagawa, & Xue, 2004), and old partners are not matched with each other. However, a

4. The conference version of our paper, published in the proceedings of WINE-2019, contained an error by supposing that the result of Roth and Vande Vate (1990) holds even when matching the old partners of the blocking pair with each other.

counterexample for the existence of a path to stability can be found in our setting, where old partners are matched with each other, by adapting those of Tamura (1993) or Tan and Su (1995) for marriage markets (it suffices in Example 9 to rank agents of the same type at the end of the given preferences: by construction and the given initial assignment, the same necessary cycle will occur).

Chen (2020) has recently proved that the problem of the existence of a path to stability with respect to *BP*-swaps is hard in marriage markets with incomplete preferences (i.e., when allowing unacceptabilities in the preferences). It remains open whether this result still holds for complete preferences, as studied in this paper.

Nevertheless, in general, determining whether all sequences of *BP*-swaps terminate in a Pareto-optimal matching, i.e., checking the convergence of *BP*-dynamics to a Pareto-optimal matching, is hard. This is due to the hardness of checking the existence of a cycle in *BP*-dynamics.

Theorem 4. *Determining whether BP-dynamics can cycle in marriage and roommate markets is* NP*-hard.*

Proof. We perform a reduction from (3,B2)-SAT, a variant of 3-SAT known to be NP-complete (Berman, Karpinski, & Scott, 2003), where the goal is to decide the satisfiability of a CNF propositional formula with exactly three literals per clause and where each variable appears exactly twice as a positive literal and twice as a negative literal. From an instance of (3,B2)-SAT with formula φ on m clauses $C_1, \dots, C_m$ and p variables $x_1, \dots, x_p$, we build a marriage market $(N = W \cup M, \succ, \mu^0)$ as follows.

For each clause C_j, with $1 \leq j \leq m$, we create four clause-agents A_j, B_j, Q_j, and K_j, where $A_j, Q_j \in W$ and $B_j, K_j \in M$. For each occurrence of variable x_i, with $1 \leq i \leq p$, we create four literal-agents, i.e., agents $Z_i^\ell, D_i^\ell \in W$, and $Y_i^\ell, E_i^\ell \in M$ for the ℓ^{th} positive literal x_i^ℓ of x_i, with $\ell \in \{1,2\}$, and $\overline{Z}_i^\ell, \overline{D}_i^\ell \in W$, and $\overline{Y}_i^\ell, \overline{E}_i^\ell \in M$ for the ℓ^{th} negative literal $\overline{x}_i^\ell$ of x_i, with $\ell \in \{1,2\}$. Denote by A, B, Q, K, D, E, Y, and Z the sets of agents associated with the same letter.

The preferences of the agents are given below, for $1 \leq i \leq p$, $1 \leq j \leq m$, and $\ell \in \{1,2\}$, with the initial assignment marked with boxes. Notation $\{\mathcal{Y}_j\}$ (resp., $\{\mathcal{D}_j\}$, $\{\mathcal{E}_j\}$, and $\{\mathcal{Z}_j\}$) refers to an arbitrary order over the literal-agents in Y (resp., D, E, and Z) which correspond to the literals of clause C_j, and $[\dots]$ is an arbitrary order over the rest of the agents of the other type. In general, when a set is given in the preferences, it refers to an arbitrary order over its elements minus the elements of the set already explicitly given in the rest of the preference ranking. The notation $cl(\ell_i)$ refers to the index of the clause in which literal ℓ_i appears. Note that A_0 (resp., B_0) stands for A_m (resp., B_m) and $\{\mathcal{Y}_{m+1}\}$ stands for $\{\mathcal{Y}_1\}$.

$$
\begin{array}{ll|ll}
A_j: & \{\mathcal{Y}_{j+1}\} \succ \{\mathcal{Y}_j\} \succ \boxed{B_j} \succ B \succ [\dots] & B_j: & \boxed{A_j} \succ A \succ \{\mathcal{Z}_{j+1}\} \succ \{\mathcal{Z}_j\} \succ [\dots] \\
 & & & \\
Z_i^1: & \overline{Y}_i^1 \succ \overline{Y}_i^2 \succ Y_i^1 \succ \boxed{E_i^1} \succ & Y_i^1: & \boxed{D_i^1} \succ \{\mathcal{D}_{cl(x_i^1)}\} \succ Q \succ \overline{Z}_i^1 \succ \overline{Z}_i^2 \succ \\
 & \quad Y \succ B_{cl(x_i^1)} \succ B_{cl(x_i^1)-1} \succ [\dots] & & \quad A_{cl(x_i^1)} \succ A_{cl(x_i^1)-1} \succ Z_i^1 \succ Q_{cl(x_i^1)} \succ [\dots] \\
Z_i^2: & \overline{Y}_i^2 \succ \overline{Y}_i^1 \succ Y_i^2 \succ \boxed{E_i^2} \succ & Y_i^2: & \boxed{D_i^2} \succ \{\mathcal{D}_{cl(x_i^2)}\} \succ Q \succ \overline{Z}_i^2 \succ \overline{Z}_i^1 \succ \\
 & \quad Y \succ B_{cl(x_i^2)} \succ B_{cl(x_i^2)-1} \succ [\dots] & & \quad A_{cl(x_i^2)} \succ A_{cl(x_i^2)-1} \succ Z_i^2 \succ Q_{cl(x_i^2)} \succ [\dots] \\
\overline{Z}_i^1: & Y_i^1 \succ Y_i^2 \succ \overline{Y}_i^1 \succ \boxed{\overline{E}_i^1} \succ & \overline{Y}_i^1: & \boxed{\overline{D}_i^1} \succ \{\mathcal{D}_{cl(\overline{x}_i^1)}\} \succ Q \succ Z_i^1 \succ Z_i^2 \succ \\
 & \quad Y \succ B_{cl(\overline{x}_i^1)} \succ B_{cl(\overline{x}_i^1)-1} \succ [\dots] & & \quad A_{cl(\overline{x}_i^1)} \succ A_{cl(\overline{x}_i^1)-1} \succ \overline{Z}_i^1 \succ Q_{cl(\overline{x}_i^1)} \succ [\dots] \\
\overline{Z}_i^2: & Y_i^2 \succ Y_i^1 \succ \overline{Y}_i^2 \succ \boxed{\overline{E}_i^2} \succ & \overline{Y}_i^2: & \boxed{\overline{D}_i^2} \succ \{\mathcal{D}_{cl(\overline{x}_i^2)}\} \succ Q \succ Z_i^2 \succ Z_i^1 \succ \\
 & \quad Y \succ B_{cl(\overline{x}_i^2)} \succ B_{cl(\overline{x}_i^2)-1} \succ [\dots] & & \quad A_{cl(\overline{x}_i^2)} \succ A_{cl(\overline{x}_i^2)-1} \succ \overline{Z}_i^2 \succ Q_{cl(\overline{x}_i^2)} \succ [\dots] \\
 & & & \\
D_i^\ell: & K_{cl(x_i^\ell)} \succ \boxed{Y_i^\ell} \succ Y \succ [\dots] & E_i^\ell: & Q_{cl(x_i^\ell)} \succ \boxed{Z_i^\ell} \succ Y \succ [\dots] \\
\overline{D}_i^\ell: & K_{cl(\overline{x}_i^\ell)} \succ \boxed{\overline{Y}_i^\ell} \succ Y \succ [\dots] & \overline{E}_i^\ell: & Q_{cl(\overline{x}_i^\ell)} \succ \boxed{\overline{Z}_i^\ell} \succ Y \succ [\dots] \\
 & & & \\
Q_j: & \{\mathcal{Y}_j\} \succ \boxed{K_j} \succ \{\mathcal{E}_j\} & K_j: & \{\mathcal{D}_j\} \succ \boxed{Q_j}
\end{array}
$$

We claim that BP-dynamics can cycle if and only if formula φ is satisfiable. The global idea of the reduction is the following. At the initial matching, the only possible BP-swaps involve blocking pairs with literal-agents in D and clause-agents in K associated with the same clause. By their swap, a literal-agent in D associated with clause C_j and clause-agent K_j can "unlock" exactly one literal-agent in Y associated with clause C_j who will not be matched with her most preferred agent anymore, and thus could have an incentive to form a blocking pair. By construction of the preferences, the only possibility to get a cycle in BP-dynamics is that, for each clause C_j, exactly one literal-agent $\mathtt{Y}_j$ in Y associated with C_j is unlocked and the cycle involves a sequence of blocking pairs $\{A_j, \mathtt{Y}_j\}, \{A_j, \mathtt{Y}_{j+1}\}, \{A_{j+1}, \mathtt{Y}_{j+1}\}, \dots$ (with $j+1$ modulo m) all along the m clauses. For this cycle to occur, the unlocked literal-agents in Y must have been matched with their associated agent in Z. Therefore, two unlocked literal agents in Y participating in the cycle cannot correspond to opposite literals. Otherwise, one of them would be matched at a moment of the cycle with an agent in Z corresponding to her opposite literal and thus would not agree to form a blocking pair with a clause-agent. An illustration of the sequence of swaps that needs to occur is given in Figure 3.

Let us first assume that there exists a truth assignment ϕ of the variables such that formula φ is satisfiable. For each clause C_j, let us denote by $\mathtt{Y}_j$ an arbitrarily chosen literal-agent in Y such that the associated literal belongs to C_j and is true in ϕ. The agent in Z (resp., D and E) corresponding to the same literal is denoted by $\mathtt{Z}_j$ (resp., $\mathtt{D}_j$ and $\mathtt{E}_j$). For each $1 \leq j \leq m$, let us perform the BP-swap with respect to blocking pair $(\mathtt{D}_j, K_j)$, which matches $\mathtt{D}_j$ with K_j, and $\mathtt{Y}_j$ with Q_j. Then, for each $1 \leq j \leq m$, let us perform the BP-swap with respect to blocking pair $(\mathtt{Z}_j, \mathtt{Y}_j)$, which matches $\mathtt{Z}_j$ with $\mathtt{Y}_j$, and $\mathtt{E}_j$ with Q_j. At this point, we reach a matching μ where exactly one literal-agent in Y per clause is matched with her associated literal-agent in Z, while all the other literal-agents in Y are matched with their best partner.

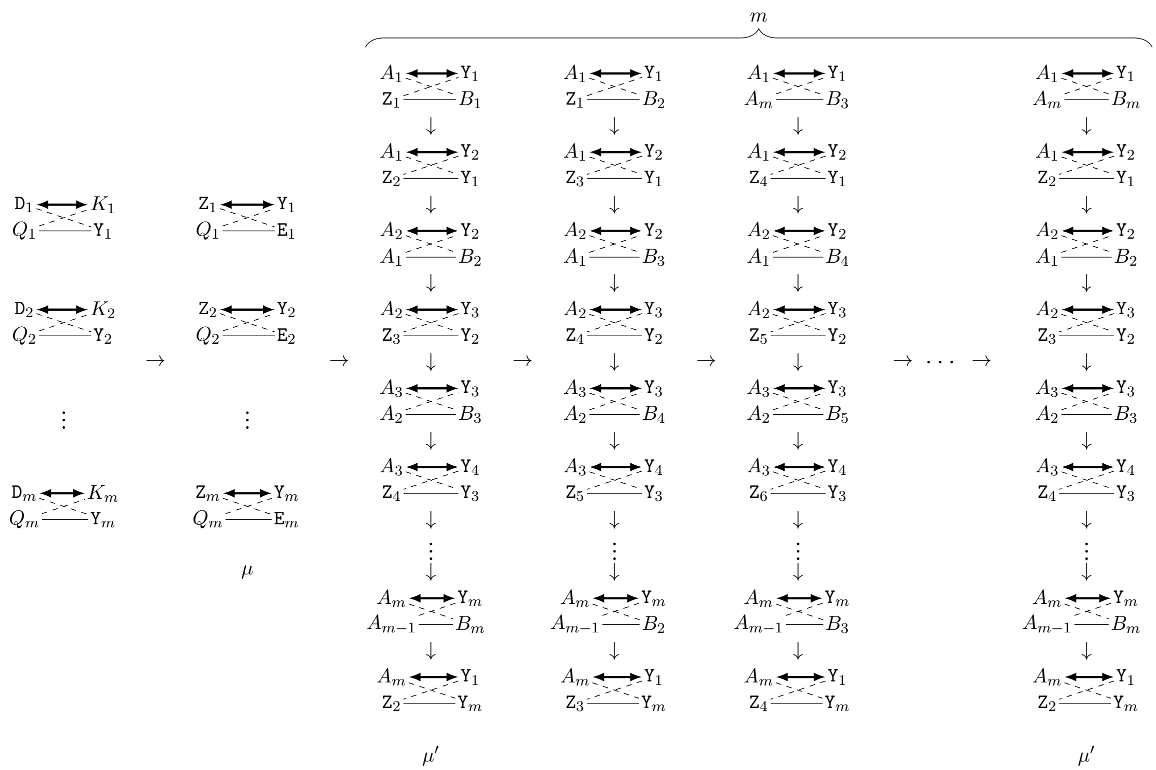


Figure 3: Description of the sequence of swaps in the proof of Theorem 4 for reaching a cycle in the dynamics. A swap is represented by a group of four agents between which dashed edges represent the old matching and plain edges the new matching resulting from the swap. The blocking pair is symbolized by a thick double arc. Arrows indicate the order between the swaps, and a matching name at the bottom of a sub-sequence of swaps indicates the name of the matching reached after all the swaps of the column. Notation $\mathtt{Y}_j$ (resp., $\mathtt{Z}_j$, $\mathtt{D}_j$ and $\mathtt{E}_j$) refers to a chosen literal-agent in Y (resp., in Z, in D and in E) such that the same associated literal belongs to C_j and is true in ϕ.

Let us consider the following sequence of blocking pairs: $\{A_1, \mathtt{Y}_1\}, \{A_1, \mathtt{Y}_2\}, \{A_2, \mathtt{Y}_2\}, \{A_2, \mathtt{Y}_3\}, \{A_3, \mathtt{Y}_3\}, \ldots, \{A_m, \mathtt{Y}_m\}, \{A_m, \mathtt{Y}_1\}$. Let us denote by μ' the matching reached after one iteration of this sequence of BP-swaps. One can observe that, by repeating the sequence of BP-swaps corresponding to this sequence of blocking pairs m times from matching μ, we come back to the matching reached after the first sequence of swaps, i.e., matching μ', as illustrated in Figure 3. It is easy to verify that all the swaps involved in the sequence are BP-swaps. Indeed, before the first iteration, agent A_1 is matched with agent B_1 but prefers to be with an agent in $\{\mathcal{Y}_1\}$, to which agent $\mathtt{Y}_1$ belongs. At the same time, agent $\mathtt{Y}_1$ prefers to be with the clause-agent in A corresponding to the clause to which her associated literal belongs, i.e., A_1, than being with her current partner $\mathtt{Z}_1$. Therefore, the swap with respect to blocking pair $\{A_1, \mathtt{Y}_1\}$ is a BP-swap which matches A_1 with $\mathtt{Y}_1$ and B_1 with $\mathtt{Z}_1$. The second swap involves agent A_1, who even prefers to be with a literal-agent associated with clause C_2, and $\mathtt{Y}_2$, who also prefers to be with the clause-agent indexed just before the clause to which her associated literal belongs, i.e., C_1, than with her current partner $\mathtt{Z}_2$. This swap matches A_1 with $\mathtt{Y}_2$ and $\mathtt{Y}_1$ with $\mathtt{Z}_2$. The process continues in the same way by assigning at each step one clause-agent in A with one literal-agent in Y, all the rest of clause-agents in A with clause-agents in B, one clause-agent in B with one literal-agent in Z (i.e., B_1 with $\mathtt{Z}_1$), and all the other literal-agents in Y with literal-agents in Z. The partners in Z of agents from Y correspond either to the same literal if the agent in Y has not been involved in a swap yet, or to one of the chosen true literals in ϕ. Since there are no two opposite literals which are both true in ϕ, an agent Y_i^ℓ (resp., $\overline{Y}_i^\ell$) corresponding to an $\mathtt{Y}_j$ for some clause C_j cannot be matched during this process with an opposite literal-agent $\overline{Z}_i^{\ell'}$ (resp., $Z_i^{\ell'}$). Therefore, any such agent Y_i^ℓ (resp., $\overline{Y}_i^\ell$), corresponding to an $\mathtt{Y}_j$, is matched, before being matched with A_{j-1}, with an agent in Z who is different from $\overline{Y}_i^{\ell'}$ (resp., $Y_i^{\ell'}$) that she prefers less than A_{j-1}, making this swap BP-rational. After the first "round" of swaps, where all BP-swaps in the sequence of blocking pairs have been performed once, we need $m-1$ new rounds with the same sequence of BP-swaps to come back to this matching, in order to let the chosen agents in Z reach the same partners in Y (there is only a shift of one clause per round for these agents). Finally, we have a cycle in BP-dynamics.

Let us now assume that there exists a cycle in BP-dynamics. First of all, observe that in the initial matching, only the agents in D and K are able to form a blocking pair. They cannot be involved in the cycle of the dynamics because they both agree that their best partner is within the set of agents with the same clause index in the other set. Observe that such a BP-swap w.r.t. blocking pair $\{D_i^\ell, K_{cl(x_i^\ell)}\}$ (resp., $\{\overline{D}_i^\ell, K_{cl(\overline{x}_i^\ell)}\}$) matches Y_i^ℓ (resp., $\overline{Y}_i^\ell$) with $Q_{x_i^\ell}$ (resp., $Q_{\overline{x}_i^\ell}$) and thus "unlocks" agent Y_i^ℓ (resp., $\overline{Y}_i^\ell$), in the sense that she will be able to make further BP-swaps since she is not matched with her best partner anymore. Note that at most one literal-agent in Y per clause C_j can be unlocked. Indeed, at this point, such a literal-agent in Y can only be matched with a partner in $\{\mathcal{D}_j\}$ or with Q_j. If she is matched with an agent in $\{\mathcal{D}_j\}$, then no swap in a blocking pair from her or from this agent will assign her a partner outside $\{\mathcal{D}_j\}$, and $\{\mathcal{D}_j\}$ are her best possible partners, so she has no incentive to form a blocking pair with other agents.

We recall that so far, the cycle has not yet begun. Now, the unlocked literal-agents in Y are able to exchange with agents in Z or A, which is the only way to get a cycle, since a cycle cannot occur within D, E, Q, or K, regarding the current assignment and the construction of the preferences of these agents. If the literal-agents in Y directly exchange with the clause-agents in A, a cycle cannot occur because the literal-agents in Y would exchange their partners in Q and, by construction of the preferences, once a literal-agent in Y is matched with a clause-agent in Q with a different clause-index from her own associated clause, she will not be involved in a further swap. Therefore, the unlocked literal-agents in Y must be paired first via a BP-swap with the literal-agent in Z associated with the same literal (they could also make a BP-swap with a literal-agent in Z associated with an opposite literal but it would prevent them to perform further swaps, by construction of the preferences). Now, the only possible BP-swaps which would not block a cycle involve unlocked agents in Y and clause-agents in A. Let us denote by L_j an unlocked agent in Y associated with clause C_j. Such an L_j can be matched via a BP-swap to clause-agent A_j or A_{j-1}. In the first case, the only possibility for the two agents to not be together anymore is because of a BP-swap of A_j with unlocked agent L_{j+1}. After such a swap, A_j has no incentive to change her partner so the only possibility for this partnership to break is that the unlocked agent L_{j+1} makes a BP-swap with A_{j+1} and then we get back to the first case for L_{j+1} and A_{j+1}. By construction of the preferences, the only way to get a cycle is to continue like this until we get back to the case of A_j and L_j (note that L_{m+1} is L_1). If we are in the second case, as we previously stated, L_j has an incentive to leave A_{j-1} to go with A_j via a BP-swap and then we get back to the first case.

Therefore, to come back to a previous matching we need an alternation of BP-swaps along the unlocked agents in Y and clause-agents in A in such a way that each agent will participate every time in two consecutive blocking pairs, i.e., L_j forms a blocking pair with A_j who then forms a blocking pair with L_{j+1} who then forms a blocking pair with A_{j+1} and so on. In order to get a cycle, it follows that there must be exactly one unlocked agent in Y per clause.

Observe that, in order to come back to a previous matching, the matched agents in Z of the unlocked agents in Y must be successively matched with each of the unlocked agents in Y. Therefore, there are no two unlocked agents in Y associated with opposite literals; otherwise, at some point, an unlocked agent in Y would be matched with an agent she prefers to the clause-agents in A, and thus, a cycle could not occur.

To summarize, by setting to true the literals associated with the unlocked agents in Y, we get a valid truth assignment of the variables which satisfies all the clauses.

This proof can be adapted to roommate markets by assuming that, in the preferences, [...] is an arbitrary order over the remaining agents where the agents of the same "type" in the marriage market are ranked last. In such a way, there will be no more swaps than in the constructed marriage market since no agent has an incentive to be matched with an agent who was of the same type in the marriage market. $\square$

Corollary 5. $\forall$-BP-ParetoSequence *is* co-NP*-hard in marriage and roommate markets.*

Nevertheless, when the preferences are globally-ranked, we can always reach a stable matching thanks to BP-dynamics in both settings. Indeed, it has been proved that BP-dynamics always converges in marriage markets with globally-ranked preferences (Ackermann, Goldberg, Mirrokni, Röglin, & Vöcking, 2011). In roommate markets, there always exists a BP-stable matching under globally-ranked preferences (Abraham et al., 2008)[5], and this stable matching is unique under 1-Euclidean preferences (Arkin, Bae, Efrat, Okamoto, Mitchell, & Polishchuk, 2009). We prove that, in addition to the existence, globally-ranked preferences enable the convergence to a BP-stable matching, thanks to a potential function argument.

Proposition 4. *BP-dynamics always converges in roommate markets for globally-ranked preferences.*

Proof. Denote by $\rhd$ the global order over all possible matched pairs in the market such that the preferences of the agents are globally-ranked with respect to this global order. Let $d(\mu)$ be the $n/2$-vector of the ranks in $\rhd$ of all the different matched pairs of μ, i.e., $d(\mu) = (rank_{\rhd}(\{i,j\}))_{i,j \ s.t. \ \mu(i)=j}$ with $rank_{\rhd}$ the function which gives the rank of the pairs in order $\rhd$. Now consider a sequence of BP-swaps given by the following sequence of matchings $(\mu^0, \mu^1, \ldots, \mu^r)$. Then, between each pair of matchings μ^t and μ^{t+1} with $0 \leq t < r$, a BP-swap is performed, say w.r.t. blocking pair $\{i,j\}$ of agents. By definition of a BP-swap, agents i and j prefer to be together than being with their partner in μ^t, so $j = \mu^{t+1}(i) \succ_i \mu^t(i)$ and $i = \mu^{t+1}(j) \succ_j \mu^t(j)$, which implies, by correlation of the preferences, that $\{i,j\} \rhd \{i, \mu^t(i)\}$ and $\{i,j\} \rhd \{j, \mu^t(j)\}$. Therefore, $(rank_{\rhd}(\{i,j\}), rank_{\rhd}(\{\mu^t(i), \mu^t(j)\}))$ is lexicographically strictly smaller than $(rank_{\rhd}(\{i, \mu^t(i)\}), rank_{\rhd}(\{j, \mu^t(j)\}))$. Since the rest of the pairs remains unchanged between μ^t and μ^{t+1}, it follows that $d(\mu^{t+1})$ is lexicographically strictly smaller than $d(\mu^t)$. Because the number of different matchings is finite, we can conclude that BP-dynamics always converges. $\square$

Since every BP-stable matching is Pareto-optimal, we obtain the following corollary.

Corollary 6. *BP-dynamics always converges to a Pareto-optimal matching in marriage and roommate markets when the preferences are globally-ranked.*

6. Fully Rational Swaps

Just as in the case of ER-swaps and housing markets, FR-swaps always represent Pareto improvements because all involved agents are strictly better off after the swap. Hence, FR-stable matchings are guaranteed to exist because every Pareto-optimal matching is FR-stable, and FR-dynamics always converges because the number of agents is finite.

In Section 4, we have shown that ER-dynamics always converges to a Pareto-optimal matching when the preferences of the agents are 1-Euclidean. It turns out that this does not hold for FR-dynamics.

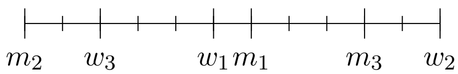
Proposition 5. *A sequence of FR-swaps may not converge to a Pareto-optimal matching in marriage and roommate markets, even for 1-Euclidean preferences.*

Proof. Consider a marriage market with three women and three men. The preferences are given below, where the initial assignment is marked with boxes.

5. More precisely, Abraham et al. (2008) talk about weakly stable matchings since they deal with a model where agents may be unmatched or indifferent between partners, two characteristics that we do not allow in our model.

$$
\begin{array}{lccccc}
w_1: & m_1 & \succ & \boxed{m_3} & \succ & m_2 \\
w_2: & m_3 & \succ & m_1 & \succ & \boxed{m_2} \\
w_3: & m_2 & \succ & \boxed{m_1} & \succ & m_3
\end{array}
\qquad
\begin{array}{lccccc}
m_1: & w_1 & \succ & \boxed{w_3} & \succ & w_2 \\
m_2: & w_3 & \succ & w_1 & \succ & \boxed{w_2} \\
m_3: & w_2 & \succ & \boxed{w_1} & \succ & w_3
\end{array}
$$

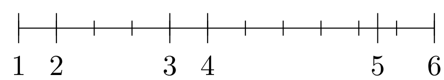
The initial matching is the only reachable matching, because no *FR*-swap is possible in this matching. However, there is another matching (circled agents) which is not reachable but which Pareto-dominates this only reachable matching. The preferences are 1-Euclidean w.r.t. the following embedding on the real line.



Now, consider a roommate market with six agents. The preferences of the agents are given below, where the initial assignment is marked with boxes.

$$
\begin{array}{lccccccccc}
1: & 2 & \succ & 3 & \succ & 4 & \succ & 5 & \succ & \boxed{6} \\
2: & 1 & \succ & \boxed{3} & \succ & 4 & \succ & 5 & \succ & 6 \\
3: & 4 & \succ & \boxed{2} & \succ & 1 & \succ & 5 & \succ & 6 \\
4: & 3 & \succ & 2 & \succ & \boxed{5} & \succ & 1 & \succ & 6 \\
5: & 6 & \succ & \boxed{4} & \succ & 3 & \succ & 2 & \succ & 1 \\
6: & 5 & \succ & 4 & \succ & 3 & \succ & 2 & \succ & \boxed{1}
\end{array}
$$

The initial matching is the only reachable matching, because there is no *FR*-swap from this matching. However, there is another matching (circled agents) which is not reachable but which Pareto-dominates this only reachable matching. The preferences are 1-Euclidean w.r.t. the following embedding on the real line.



Note that both counterexamples are minimal because, in a smaller instance, a matching would be composed of exactly two pairs of agents, and thus, any swap would involve all the agents. Therefore, a matching in which Pareto-dominates a stable matching would make every agent better off in comparison to the current stable one, and thus, there would exist an *FR*-swap, a contradiction. □

The proofs of Theorems 1 and 3 only dealt with instances in which *FR*-swaps are identical to *ER*-swaps. We thus immediately obtain hardness of ∃-FR-ParetoSequence and ∀-FR-ParetoSequence. An *FR*-swap makes four agents strictly better off and no agent worse off. Thus, the size of a sequence of *FR*-swaps is bounded by $\mathcal{O}(n^2)$. Moreover, the Pareto-optimality of a matching can be checked in polynomial time. Therefore, we get the membership of the problems in NP and co-NP, respectively.

Theorem 5. ∃-FR-ParetoSequence *is* NP*-complete in marriage and roommate markets, even for globally-ranked preferences.*

Theorem 6. ∀-FR-ParetoSequence *is* co-NP*-complete in marriage and roommate markets, even for globally-ranked preferences.*

Market	Preferences	Exchange Rational Swaps	Blocking Pair Swaps	Fully Rational Swaps
Housing	General / GR	■■□ (Obs. 5)	–	–
	SP	■■■ (Damamme et al., 2015)	–	–
	1-D	■■■	–	–
Marriage	General	□□□	■□□ (Gale & Shapley, 1962)	■■□
	GR	■■□ (Prop. 1 and Prop. 2)	■■■ (Cor. 6)	■■□
	SP	□□□ (Cechlárová, 2002)	■□□	■■□
	1-D	■■■ (Cor. 1)	■■■	■■□ (Prop. 5)
Roommate	General	□□□	□□□ (Gale & Shapley, 1962)	■■□
	GR	■■□ (Prop. 1 and Prop. 2)	■■■ (Cor. 6)	■■□
	SP	□□□ (Alcalde, 1994)	□□□	■■□
	1-D	■■■ (Cor. 1)	■■■	■■□ (Prop. 5)

Table 1: Summary of the results on the *existence of a stable matching* (■□□), the *guarantee of convergence to a stable matching* (■■□), and the *guarantee of convergence to a Pareto-optimal matching* (■■■) for the three different matching markets under study, according to different types of rational swaps and under different preference domains (General, globally-ranked (GR), single-peaked (SP), and 1-Euclidean (1-D)). All results are tight and we always list the strongest property that is satisfied. The only meaningful type of rational swaps in housing markets are exchange-rational swaps; hence, the empty spaces.

7. Conclusion

We have studied the properties of different dynamics of rational swaps in matching markets with initial assignments and, in particular, the question of convergence to a Pareto-optimal matching. For all considered settings, the dynamics may not terminate in a Pareto-optimal matching because *(i)* there is no stable matching, *(ii)* the dynamics does not converge, or *(iii)* the stable matching that is eventually reached is not Pareto-optimal. An overview of our results according to different preference restrictions is given in Table 1.

From a computational perspective, determining whether there *exists* a sequence of rational swaps terminating in a Pareto-optimal matching is NP-hard for fully rational swaps and exchange rational swaps in all matching markets, even for globally-ranked preferences (Theorems 1 and 5). However, the convergence to a Pareto-optimal matching, that is, whether *all* sequences of swaps terminate in a Pareto-optimal matching, is co-NP-hard to decide (Corollary 5). Unsurprisingly, the same hardness result holds for fully rational and exchange rational swaps, even for globally-ranked preferences (Theorems 3 and 6). Our computational results are summarized in Table 2. Even if the existence of a sequence of swaps terminating in a Pareto-optimal matching is not guaranteed for single-peaked preferences in marriage and roommate markets, it would be interesting to know whether this preference restriction is nevertheless sufficient for efficiently solving our computational problems in these markets.

The convergence to a Pareto-optimal matching in housing markets for exchange rational dynamics and single-peaked preferences (Damamme et al., 2015) does not hold for more general settings where the "objects" are agents who have preferences. However, this con-

Market	Prefs	Exchange Rational Swaps		Blocking Pair Swaps		Fully Rational Swaps	
		∃-ParetoSeq	∀-ParetoSeq	∃-ParetoSeq	∀-ParetoSeq	∃-ParetoSeq	∀-ParetoSeq
Housing	General / GR	NP-c. (Cor. 2)	co-NP-c. (Cor. 4)	–	–	–	–
	SP	in P (Damamme et al., 2015)	in P	–	–	–	–
Marriage / Roommate	General	NP-h. (Th. 1)	co-NP-h. (Th. 3)	?	co-NP-h. (Cor. 5)	NP-c. (Th. 5)	co-NP-c. (Th. 6)
	GR	NP-h. (Th. 1)	co-NP-h. (Th. 3)	in P (Cor. 6)	in P (Cor. 6)	NP-c. (Th. 5)	co-NP-c. (Th. 6)

Table 2: Summary of the computational results on the existence (∃-ParetoSeq) or the guarantee (∀-ParetoSeq) of sequences of rational swaps terminating in a Pareto-optimal matching for the three different matching markets under study, according to different types of rational swaps and under different preference domains (General, globally-ranked (GR), and single-peaked (SP)). The only meaningful type of rational swaps in housing markets are exchange-rational swaps; hence, the empty spaces.

vergence is guaranteed under 1-Euclidean preferences in marriage and roommate markets. Hence, the generalization of this convergence result to more general settings requires more structure in the preferences.

A natural extension of this work would be to study meaningful dynamics for hedonic games, where agents form groups consisting of more than two agents.

Acknowledgments

This material is based upon work supported by the Deutsche Forschungsgemeinschaft under grant BR 2312/12-1. Preliminary results of this paper were presented at the 15th Conference on Web and Internet Economics (New York, December 2019). We are grateful to an anonymous reviewer for introducing us to the notion of globally-ranked preferences and to Jiehua Chen for valuable discussions on blocking pair dynamics.

References

Abdulkadiroğlu, A., Che, Y.-K., Pathak, P. A., Roth, A. E., & Tercieux, O. (2017). Minimizing justified envy in school choice: The design of New Orleans' OneApp. Tech. rep. 23265, National Bureau of Economic Research Working Paper.

Abdulkadiroğlu, A., & Sönmez, T. (1999). House allocation with existing tenants. *Journal of Economic Theory*, *88*(2), 233–260.

Abeledo, H., & Rothblum, U. G. (1995). Paths to marriage stability. *Discrete Applied Mathematics*, *63*(1), 1–12.

Abraham, D. J., Cechlárová, K., Manlove, D. F., & Mehlhorn, K. (2005). Pareto optimality in house allocation problems. In *Proceedings of the 16th International Symposium* *on Algorithms and Computation (ISAAC)*, Vol. 3341 of *Lecture Notes in Computer Science (LNCS)*, pp. 1163–1175.

Abraham, D. J., Levavi, A., Manlove, D. F., & O'Malley, G. (2008). The stable roommates problem with globally ranked pairs. *Internet Mathematics*, *5*(4), 493–515.

Abraham, D. J., & Manlove, D. F. (2004). Pareto optimality in the roommates problem. Tech. rep. TR-2004-182, University of Glasgow, Department of Computing Science.

Ackermann, H., Goldberg, P. W., Mirrokni, V. S., Röglin, H., & Vöcking, B. (2011). Uncoordinated two-sided matching markets. *SIAM Journal on Computing*, *40*(1), 92–106.

Ackermann, H., Goldberg, P. W., Mirrokni, V. S., Röglin, H., & Vöcking, B. (2008). A unified approach to congestion games and two-sided markets. *Internet Mathematics*, *5*(4), 439–457.

Alcalde, J. (1994). Exchange-proofness or divorce-proofness? Stability in one-sided matching markets. *Review of Economic Design*, *1*(1), 275–287.

Alcalde-Unzu, J., & Molis, E. (2011). Exchange of indivisible goods and indifferences: The top trading absorbing sets mechanisms. *Games and Economic Behavior*, *73*(1), 1–16.

Arkin, E. M., Bae, S. W., Efrat, A., Okamoto, K., Mitchell, J. S. B., & Polishchuk, V. (2009). Geometric stable roommates. *Information Processing Letters*, *109*(4), 219–224.

Aziz, H. (2019). Algorithms for Pareto optimal exchange with bounded exchange cycles. *Operations Research Letters*, *47*(5), 344–347.

Aziz, H., Brandt, F., & Harrenstein, P. (2013). Pareto optimality in coalition formation. *Games and Economic Behavior*, *82*, 562–581.

Aziz, H., & Goldwaser, A. (2017). Coalitional exchange stable matchings in marriage and roommate markets (extended abstract). In *Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pp. 1475–1477.

Bartholdi, III, J., & Trick, M. (1986). Stable matching with preferences derived from a psychological model. *Operations Research Letters*, *5*(4), 165–169.

Berman, P., Karpinski, M., & Scott, A. (2003). Approximation hardness of short symmetric instances of MAX-3SAT. Tech. rep. TR03-049, Electronic Colloquium on Computational Complexity (ECCC).

Black, D. (1948). On the rationale of group decision-making. *Journal of Political Economy*, *56*(1), 23–34.

Cechlárová, K. (2002). On the complexity of exchange-stable roommates. *Discrete Applied Mathematics*, *116*(3), 279–287.

Cechlárová, K., & Manlove, D. F. (2005). The exchange-stable marriage problem. *Discrete Applied Mathematics*, *152*(1–3), 109–122.

Chen, J. (2020). Reaching stable marriage via divorces is hard. Tech. rep., http://arxiv.org/abs/1906.12274.

Coombs, C. H. (1950). Psychological scaling without a unit of measurement. *Psychological Review*, *57*(3), 145–158.

Coombs, C. H. (1964). *A Theory of Data.* Wiley.

Damamme, A., Beynier, A., Chevaleyre, Y., & Maudet, N. (2015). The power of swap deals in distributed resource allocation. In *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pp. 625–633.

Diamantoudi, E., Miyagawa, E., & Xue, L. (2004). Random paths to stability in the roommate problem. *Games and Economic Behavior*, *48*(1), 18–28.

Doignon, J.-P., & Falmagne, J.-C. (1994). A polynomial time algorithm for unidimensional unfolding representations. *Journal of Algorithms*, *16*(2), 218–233.

Elkind, E., & Faliszewski, P. (2014). Recognizing 1-Euclidean preferences: An alternative approach. In *Proceedings of the 7th International Symposium on Algorithmic Game Theory (SAGT)*, Lecture Notes in Computer Science (LNCS), pp. 146–157.

Escoffier, B., Lang, J., & Öztürk, M. (2008). Single-peaked consistency and its complexity. In *Proceedings of the 18th European Conference on Artificial Intelligence (ECAI)*, pp. 366–370.

Gabow, H. N., Kaplan, H., & Tarjan, R. E. (2001). Unique maximum matching algorithms. *Journal of Algorithms*, *40*(2), 159–183.

Gale, D., & Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, *69*(1), 9–15.

Gourvès, L., Lesca, J., & Wilczynski, A. (2017). Object allocation via swaps along a social network. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 213–219.

Gusfield, D., & Irving, R. W. (1989). *The stable marriage problem: Structure and algorithms.* MIT Press, Cambridge, MA, USA.

Hoefer, M., Vaz, D., & Wagner, L. (2018). Dynamics in matching and coalition formation games with structural constraints. *Artificial Intelligence*, *262*, 222–247.

Hoefer, M. (2013). Local matching dynamics in social networks. *Information and Computation*, *222*, 20–35.

Huang, S., & Xiao, M. (2020). Object reachability via swaps under strict and weak preferences. *Autonomous Agents and Multi-Agent Systems*, *34*(2), 1–33.

Irving, R. W. (1985). An efficient algorithm for the "stable roommates" problem. *Journal of Algorithms*, *6*(4), 577–595.

Irving, R. W., & Manlove, D. F. (2002). The stable roommates problem with ties. *Journal of Algorithms*, *43*(1), 85–105.

Klaus, B., Manlove, D. F., & Rossi, F. (2016). Matching under preferences. In Brandt, F., Conitzer, V., Endriss, U., Lang, J., & Procaccia, A. D. (Eds.), *Handbook of Computational Social Choice*, chap. 14, pp. 333–355. Cambridge University Press.

Knoblauch, V. (2010). Recognizing one-dimensional Euclidean preference profiles. *Journal of Mathematical Economics*, *46*(1), 1–5.

Knuth, D. E. (1976). *Mariages stables.* Les Presses de l'Université de Montréal.

Manlove, D. F. (2013). *Algorithmics of Matching Under Preferences.* World Scientific Publishing Company.

McVitie, D. G., & Wilson, L. B. (1971). The stable marriage problem. *Communications of the ACM*, *14*(7), 486–490.

Morrill, T. (2010). The roommates problem revisited. *Journal of Economic Theory*, *145*(5), 1739–1756.

Peters, D. (2017). Recognising multidimensional Euclidean preferences. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, pp. 642–648.

Roth, A. E., & Vande Vate, J. H. (1990). Random paths to stability in two-sided matching. *Econometrica*, *58*(6), 1475–1480.

Saffidine, A., & Wilczynski, A. (2018). Constrained swap dynamics over a social network in distributed resource reallocation. In *Proceedings of the 11th International Symposium on Algorithmic Game Theory (SAGT)*, Vol. 11059 of *Lecture Notes in Computer Science (LNCS)*, pp. 213–225.

Salonen, H., & Salonen, M. A. (2018). Mutually best matches. *Mathematical Social Sciences*, *91*, 42–50.

Shapley, L. S., & Scarf, H. (1974). On cores and indivisibility. *Journal of Mathematical Economics*, *1*(1), 23–37.

Tamura, A. (1993). Transformation from arbitrary matchings to stable matchings. *Journal of Combinatorial Theory, Series A*, *62*(2), 310–323.

Tan, J., & Su, W. (1995). On the divorce digraph of the stable marriage problem. *Proceedings of the National Science Council, Republic of China (A)*, *19*(2), 342–354.

Yoshinaka, R. (2005). Higher-order matching in the linear lambda calculus. In *Proceedings of the 16th International Conference on Term Rewriting and Applications (RTA)*, pp. 235–249.