The final publication is available at 10.1016/j.mathsocsci.2015.12.001

A new solution for roommate problems: The $\mathcal{Q}$-stable matchings*

Péter Biró†, Elena Iñarra‡, and Elena Molis§

September 3, 2014

Abstract

In this paper we propose a new family of matchingsas solution for the roommate problem with strict preferences, when stable matchings may not exist. To define these matchings we proceed as follows: We introduce the solution of maximum irreversibility, a strong notion of stability, and consider two other existing solutions that deal with unsolvable roommate problems: the almost stable matchings (Abraham et al. [2]) and the maximum stable matchings (Tan [31] [33]). Although each of these core consistent solutions is a good candidate for solving roommate problems, we find that it is not possible to reconcile almost stability with any of the other two. Hence we select the family of matchings, the $\mathcal{Q}$-stable family, that lie in the intersection of the maximum irreversible matchings and maximum stable matchings. Then we offer an efficient algorithm to compute a member of this family: a $\mathcal{Q}^*$-stable matching.

*This research is supported by the Spanish Ministry of Science and Innovation (ECO2010-17049 and ECO2012-31346), co-funded by ERDF, by Basque Government IT-568-13 and by the Government of Andalusia Project for Excellence in Research (P07.SEJ.02547). Péter Biró also acknowledges the support from the Hungarian Academy of Sciences under its Momentum Programme (LD-004/2010), and the Hungarian Scientific Research Fund, OTKA, grant no. K108673.

†Institute of Economics, Research Centre for Economic and Regional Studies, Hungarian Academy of Sciences, H-1112, Budaörsi út 45, Budapest, Hungary. Email: peter.biro@krtk.mta.hu. He a visiting professor at Department of Economics, Stanford University in 2014.

‡*BRiDGE,* University of the Basque Country (UPV/EHU). and Instituto de Economía Pública. Email: elena.inarra@ehu.es

§GLOBE, Departamento de Teoría e Historia Económica, Universidad de Granada.. Email: emolis@ugr.es

1 Introduction

Gale and Shapley [13] introduce one-to-one matching problems. They first define, the marriage problem, a two-sided matching problem in which agents are divided in two disjoint groups and an agent can only be matched to an agent in the other group. Then, they proceed with the roommate problem, a one-sided matching problem in which all agents belong to a unique group and every agent can be matched to any other. The authors propose stable matchings as solution for those problems. A matching is stable if no two agents prefer one another to their current partners. They show that for the marriage problem a stable matching always exists, while this could not be true for the roommate problem. The following example, slightly modified from the original one in Gale and Shapley [13], illustrates this case.

Example 1 *Consider the following 4-agent problem:*

a_1	a_2	a_3	a_4
a_2	a_3	a_1	a_1
a_3	a_1	a_2	a_2
a_4	a_4	a_4	a_3
a_1	a_2	a_3	a_4

To see that no stable matching exists, suppose that one of the first three agents, say agent a_3, is either unmatched or matched to agent a_4. Then, agents a_2 and a_3 form a blocking pair. Similar arguments can be applied to agents a_1 and a_2.

In the last decades an extensive literature on one-to-one matching problems has emerged both in Economics and in Computer Science. However, most works focus on the marriage problem while the roommate problem has been much less studied. This can be explained by two reasons: First, a greater number of economic issues can be modeled as a two-sided problem than as a one-sided one. Second, the impossibility of finding a stable matching and the more complex structure of the roommate problem may have discouraged researchers from analyzing it.

Pairing policemen to patrol, pilots to flight (see Cechlárová and Ferková [8]), students to share double rooms in colleges or marriages between agents of the same-sex are significant problems to be analyzed. In sport competitions the manner in which agents are paired in teams of two players, such as double tennis or paddle, may affect the final result. The kidney exchange problem has been modeled as a roommate problem (see Irving [28]). Furthermore there are centrally coordinated programs such as the odd shoe exchanges,[1] holidays homes exchanges,[2] and centralized pairing methods used in chess competitions (see Kujansuu *et al.* [22]), which also suggest potential applications of the roommate problem. Moreover, as Klaus *et al.* [21] point out, the roommate problem has interest in itself since it boils down into hedonic coalition formation (see Bogomolnaia and Jackson [6]) and network formation problems (see Jackson and Watts [18]).

In the scarce literature about one-sided matching problems a common practice is to restrict the analysis to those problems in which a stable matching exists (solvable problems) (see for instance, Gusfield and Irving [15], Chung [9], Diamantoudi *et al.* [11], Klaus and Klijn [20] and Gudmundsson [14]). However, restricting the attention to solvable roommate problems ignores a significant subclass of problems without stable matchings (unsolvable problems). This fact has been corroborated by Pittel and Irving [26] who observe that as the number of agents increases the probability of a roommate problem being unsolvable also increases fairly steeply.

The aim of the current paper is to propose a new solution for the roommate problems with strict preferences.[3] Indeed it is essential to require a solution which provides a stable matching when dealing with solvable problems and some matching otherwise. Hence we focus on core consistent solutions[4]. At the interface between Economics and Computer Sciences several solutions have been proposed explicitly for unsolvable problems, but still there is a pending discussion in depth regarding comparison between solutions as well as scope for a new one.

Two interesting solutions have been analyzed in the literature for unsolvable problems: *The almost stable matchings*, proposed by Abraham *et al.* [2], form a subclass of Pareto-optimal matchings with the minimum number of blocking pairs.[5] The notion of *maximum internal stability* introduced by Tan [31], single

[1] http://www.oddshoe.org/

[2] http://www.exchangeholidayhomes.com/

[3] Except for small number of references, we have deliberately avoided the analysis of a variety of roommate problem reappraisals.

[4] Other solutions have been proposed in the literature, for instance popular matching. However, this solution is not core-consistent and for that reason we have ruled it out from our analysis. For more details in such a solution see Biro et al. [5].

[5] See Biermann [3] for a critical evaluation of this criterion in the marriage problem.

out matchings with the largest set of pairs that are stable among them. However, the following basic proposal has been overlooked. Consider the case in which two agents are top choice for each other. Once this pair is formed it never splits. A less extreme case is the existence of a set of agents forming a pairing so strongly stable that they are stably paired within them and none of them prefers an agent outside to her current partner no matter how the outside agents are matched. Hence, once these pairs are formed they never break. Thus, we believe that a *maximum irreversible* set of pairs should form part of the matching selected to solve any roommate problem.

Certainly each of the three mentioned solutions show sufficient grounds to be considered as a good candidate for solving roommate problems. Then it makes sense to consider to make a proposal that could conciliate most if not all of the mentioned solutions.

From the study of the relationship between these solutions we find that the almost stable solution is incompatible with the other two. Moreover it happens that the problem of finding a matching with the minimum number of blocking pairs is NP-hard. Hence, our next step (move) is to search a solution that it could conciliate the notions of maximum internally stability and maximum irreversibility. Thus, we select the set of matchings that lie in the intersection of the two solutions and refer to as the Q-stable matchings. Since our late motive is the selection of a single matching to solve the roommate problem, an essential criterion to take into account is the possibility of determining one of those matchings. We offer an efficient algorithm to compute such a matching.

Finally, let us extend what we have learnt from two-sided matching problems to one-sided matching problems. In two-sided matching problems if agents interact freely, and after a match they decide systematically what to do next then they eventually reach a stable matching. It is also known that market frictions may prevent to obtain a stable matching.[6] This justifies the presence of clearinghouses[7] to which agents submit preference lists to a policy maker who, following a procedure, implements the desired matching. Similarly in one-sided matchings it is important to identify some matching resulting from a decentralized process. For the roommate problem it is known that the blocking

[6]A well-known documented episode of unraveling in matching markets for medical interns shows that contracts for interns were signed two years earlier than students' graduation (see, for instance, Echenique and Pereyra [12] and the references therein).

[7]For instance, the National Resident Matching Program (NRMP) matches physicians and residency programs in the United States.

dynamic between agents leads to an *absorbing set* of matchings (see Inarra et al. [17] and Klaus et al. [21]). In fact, once one of these matchings has been reached, the blocking dynamic of the agents does not allow to abandon that set. Hence, we proceed studying whether our proposal is one of the elements of an absorbing set. We find that, although not all Q-stable matchings belong to an absorbing set the matching determined by the algorithm does. Therefore we are providing to policy makers with a procedure that implements a Q-stable matching for solving roommate problems.

The paper is organized as follows: Section 2 contains the preliminaries of the paper. In Section 3 we present and discuss the notion of maximum irreversibility and the other two solutions existent in the literature for the unsolvable roommate problem, and we proceed with the comparison of these three core consistent solutions. In Section 4 we introduce the Q-stable matchings and an algorithm to compute one of them. We also show that such a matching belongs to an absorbing set. Section 5 concludes.

2 Preliminaries

In a roommate problem , a finite set of agents $N = \{a_1, \ldots, a_n\}$ has to be partitioned into pairs and singletons. Each agent has strict preference over potential roommates with the possibility of having a room for herself/himself. Formally, a *roommate problem*, or a *problem* for short is a pair $(N, (\succeq_{a_i})_{a_i \in N})$ (or $(N, \succeq)$ for short) where N is a finite set of agents and for each agent $a_i \in N$, $\succeq_{a_i}$ is a complete, transitive preference relation defined over N. Preferences are strict, *i.e.*, $a_k \succeq_{a_i} a_j$ and $a_j \succeq_{a_i} a_k$ if and only if $a_j = a_k$. The strict preference relation associated with $\succeq_{a_i}$ is denoted by $\succ_{a_i}$. Agent a_j is *acceptable* for agent a_i if $a_j \succ_{a_i} a_i$. Otherwise it is said to be *unacceptable*. A solution to a problem, a *matching*, is a function $\mu : N \to N$ such that if $\mu(a_i) = a_j$ then $\mu(a_j) = a_i$. Thus, a matching is a set of disjoint pairs and singletons formed by the agents in N. Let $\mu(a_i)$ denotes agent's a_i partner at matching μ. If $\mu(a_i) = a_i$, then agent a_i is unmatched in μ. A matching μ with all its agents paired is called *complete*. Given $S \subseteq N$, $S \neq \emptyset$, let $\mu(S) = \{\mu(a_i) : a_i \in S\}$. That is, $\mu(S)$ is the set of partners of the agents in S under matching μ. Let $\mu \mid_S$ denotes the restriction of μ to agents in S. If $\mu(S) = S$, then $\mu \mid_S$ is a matching in $(S, (\succ_{a_i})_{a_i \in S})$.

A matching μ is *blocked by a pair* $\{a_i, a_j\} \subseteq N$ if $a_j \succ_{a_i} \mu(a_i)$ and $a_i \succ_{a_j} \mu(a_j)$, that is a_i and a_j prefer each other to her current partner (if any) in μ. If pair $\{a_i, a_j\}$ blocks matching μ then $\{a_i, a_j\}$ is called a *blocking pair* of μ. Let $\{a_i, a_j\}$ blocks a matching μ. A matching μ' is obtained from μ by *satisfying* $\{a_i, a_j\}$ if $\mu'(a_i) = a_j$, their partners (if any) under μ are alone in μ', and the remaining agents are matched as in μ. A matching without blocking pairs is called a *stable* matching. A problem is called *solvable* if the set of stable matchings is non-empty and *unsolvable* otherwise.

We extend each agent's preferences over her potential partners to the set of matchings in the following way: We say that agent a_i prefers μ' to μ, and denote it by $\mu' \succ_i \mu$ if and only if agent a_i prefers her partner at μ' to her partner at μ, $\mu'(a_i) \succ_{a_i} \mu(a_i)$. (We say that agent a_i is indifferent between matchings μ' and μ, denoted by $\mu' \sim_{a_i} \mu$ if she is matched to the same partner in both matchings).

Stable partitions

Tan [32] establishes the necessary and sufficient condition for the solvability of a problem with strict preferences using the notion of stable partition. Formally the notion of stable partition is formally defined as follows:

Let $A = \{a_1, ..., a_k\} \subseteq N$ be an ordered set of agents. The set A is a *ring* if $k \geq 3$ and for all $i \in \{1, ..., k\}$, $a_{i+1} \succ_{a_i} a_{i-1} \succ_{a_i} a_i$ (subscript modulo k). The set A is a *pair of mutually acceptable agents* if $k = 2$ and for all $i \in \{1, 2\}$, $a_{i-1} \succ_{a_i} a_i$ (subscript modulo 2). The set A is a *singleton* if $k = 1$.

A *stable partition* is a partition $\mathcal{P}$ of N such that:
(i) For all $A \in \mathcal{P}$, the set A is a ring, a pair of mutually acceptable agents or a singleton, and
(ii) For all $A, B \in \mathcal{P}$ where $A = \{a_1, ..., a_k\}$ and $B = \{b_1, ..., b_l\}$ (possibly $A = B$), the following condition holds:

$$\text{if } b_j \succ_{a_i} a_{i-1} \text{ then } b_{j-1} \succ_{b_j} a_i,$$

for all $i \in \{1, ..., k\}$ and $j \in \{1, ..., l\}$ such that $b_j \neq a_{i+1}$.

Thus, a stable partition is a partition of the set of agents such that each set in a stable partition is either a ring, a pair of mutually acceptable individuals, or a singleton, and the partition satisfies the (usual) stability condition between any two sets and also within each set.[8] The following assertions are proven by Tan [32].

Remark 1 *(i) A problem $(N, \succ)$ has no stable matchings if and only if there exists a stable partition with an odd ring. (ii) All stable partitions have exactly the same odd rings and singletons. (iii) All even rings in a stable partition can be broken into pairs of mutually acceptable agents while preserving stability.*

Without loss of generality hereafter we suppose that the even parties are always pairs in any stable partition we are working with.

Since stable partition plays a significant role in the present work and his interpretation is not that easy, in the appendix we introduce informally the algorithm introduced by Tan and Hsueh [34], and illustrate it with a numerical example which we believe it clarifies its meaning.

3 Core consistent solutions

In this section we first introduce a notion of strong stability that we believe is appropriate to be considered in the search of a matching as stable as possible. Then we consider two existing solutions proposed for unsolvable problems.

3.1 Maximum irreversibility

Consider the case in which two agents are top choice for each other. Once this pair is formed it never splits.[9] A less extreme case is the existence of a set of agents forming a pairing so strongly stable that they are stably paired within them and none of them prefers an agent outside to her current partner no matter how the outside agents are matched. Therefore, once these pairs are formed they never break. We call this set of pairs ' irreversible".

A problem may have matchings with different irreversible sets but it seems natural to require that some of the largest ones is contained into the proposed matching. Formally,

Definition 1 *(i) A set of agents $S \subseteq N$ form an* irreversible *set of pairs μ_S if there is no pair $\{a_i, a_j\}$ (possibly $a_i = a_j$) such that $\{a_i, a_j\} \cap S \neq \emptyset$ such that*

[8]Stable partitions are also called *stable half-matchings* in some recent papers, such as Biró *et al.* [4]. A half-matching is a well- known notion in graph theory that also helps to understand the meaning of this notion.

[9]This property, called 'mutually best" property was introduced by Toda [35] for the marriage problem and by Can and Klaus [7] for the roommate problem.

$\{a_i, a_j\}$ *blocks* μ_S. *(ii) Matching* μ *is* maximum irreversible *if it contains the largest irreversible set of pairs.*

The set of maximum irreversible matchings is core-consistent and the larger the set of irreversible pairs is, the more selective this criterion will be. To see the robustness of this solution consider the following example:

Example 2 *Consider the following* 10-*agent problem:*

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
a_2	a_3	a_1	a_8	a_9	a_4	a_6	a_5	a_7	a_1
a_3	a_1	a_2	a_9	a_1	a_5	a_8	a_7	a_4	a_2
a_4	a_{10}	a_4	a_3	a_6	a_7	a_9	a_4	a_5	a_3
a_6	a_5	a_5	a_6	a_8	a_8	a_5	a_6	a_1	a_4
a_5	a_6	a_6	a_7	a_7	a_9	a_1	a_1	a_2	a_5
a_7	a_7	a_7	a_{10}	a_4	a_2	a_2	a_2	a_3	a_6
a_8	a_8	a_8	a_1	a_2	a_3	a_3	a_9	a_6	a_7
a_9	a_4	a_9	a_2	a_3	a_1	a_4	a_1	a_8	a_8
a_{10}	a_9	a_{10}	a_5	a_{10}	a_{10}	a_{10}	a_3	a_{10}	a_9
a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}

Matching $\mu_1 = \{\{a_1, a_2\}, \{a_3\}, \{a_4, a_8\}, \{a_5, a_9\}\{a_7, a_6\}, \{a_{10}\}\}$ *is maximum irreversible, with an irreversible set of three pairs. The pairing* $\{\{a_4, a_8\}, \{a_5, a_9\}\{a_7, a_6\}\}$ *is stable and no agent in it prefers any other agent outside to her current partner. Hence, once this pairing is formed these pairs keep together.*

However, for unsolvable problems the set of irreversible pairs of a problem might be empty. Hence, this criterion by itself might be indefinite. In what follows, we present two core consistent solutions that have been proposed in the literature.

3.2 Two known solutions

In this subsection we present two core consistent solutions that deal explicitly with unsolvable problems: almost stable matchings and the maximum internally stable matchings.

Almost stability

Pareto optimality, one of the most relevant criterion in Economics, has also been applied to the roommate problem and to a variety of its extensions.[10] In our

[10] Sotomayor [30] studies Pareto optimality in the roommate problem with indifferences. Özkal-Sanver [25] establishes some impossibility results for roommate problems which includes

setup it can be defined as follows:

Definition 2 *A matching μ is* Pareto optimal *if there does not exist a matching μ' such that $\mu'(a_i) \succeq_{a_i} \mu(a_i)$ for all $a_i \in N$ and $\mu'(a_i) \succ_{a_i} \mu(a_i)$ for some $a_i \in N$.*

There is no doubt that Pareto optimality is an appealing criterion related to stability. A stable matching is Pareto-optimal (see *Proposition 5* in Abraham and Manlowe [1]) and a non-Pareto optimal matching is always unstable. This is because it is blocked by a set of agents who are better off in another matching.[11] However, Pareto optimality by itself is not a convincing criterion for selecting matchings in the roommate problem for two reasons. On the one hand, it requires that when any two agents block a matching by forming a new pair their partners, if any, must not get worse in the new matching. In our setting, however, it is enough that two agents improve forming a blocking pair, without considering the well being of their abandoned partners in the new matching formed.[12] On the other hand, it can select too many matchings: For solvable problems, the Pareto-optimal solution is core inclusive, that is, it selects all stable matchings and some unstable ones. For unsolvable problems, it suffers from a similar drawback; it can select matchings differing on the number of blocking pairs.

The idea of refining the set of Pareto optimal matchings was undertaken by Abraham and Manlove [1] who prove the following assertions:

Remark 2 *Let $bp(\mu)$ denote the set of blocking pairs of matching μ, that is $bp(\mu) = \{\{a_i, a_j\} \subseteq N : \{a_i, a_j\}$ blocks $\mu\}$ (i) If matching μ Pareto dominates matching μ' then $bp(\mu) \subset bp(\mu')$ (ii) If μ is a matching with the fewest number of blocking pairs of a problem then μ is Pareto optimal.*

Following the idea behind the previous results, Abraham et al. [2] study matchings with the fewest number of blocking pairs and call them almost stable matchings. Formally,

Definition 3 *A matching μ is* almost stable *if $|bp(\mu)| \leq |bp(\mu')|$ for all $\mu' \neq \mu$, where $|bp(\mu)|$ denotes the number of blocking pairs of matching μ.*

Pareto optimality as an axiom.

[11] If a matching is not Pareto optimal then it admits an improving coalition (see Proposition 6.24 in Manlove [23]) See Chapter 6 in this book for a survey on Pareto optimal matchings.

[12] This is not the case in problems in which bilateral approval is required to dissolve partnership see Morrill [24].

Maximum internal stability

A matching μ is *maximum stable* if it excludes the minimum number of agents such that the non excluded ones form a complete stable matching see Tan [31] [33]. For computing a maximum stable matching, given a stable partition Tan [31] proposes to delete one agent from each odd ring of the partition as well as all singletons. Then he defines the problem restricted to the set of non-deleted agents, keeping their original preferences over them. This new problem is solvable and the computation of a stable matching gives a maximum stable matching.

Example 2 (cont.) *In this problem,* $\mathcal{P} = \{\{a_1, a_2, a_3\}, \{a_4, a_6\}, \{a_5, a_8\}, \{a_7, a_9\}, \{a_{10}\}\}$ *is a stable partition and all maximum stable matchings have four stable pairs. Matching* $\mu_1 \supseteq \{\{a_2, a_3\}, \{a_4, a_6\}, \{a_5, a_8\}, \{a_7, a_9\}\}$ *is maximum stable and it can be derived from Tan's proposal by isolating one agent from the odd ring, agent* a_2*, and the singleton, agent* a_{10}*. Apart from the maximum stable matchings obtained in this manner there may be others. For instance,* $\mu_2 \supseteq \{\{a_2, a_{10}\}, \{a_4, a_6\}, \{a_5, a_8\}, \{a_7, a_9\}\}$ *which is also maximum stable.*

Tan's solution is applied to a setting in which a matching is defined as a set of pairs while isolated agents never form part of that matching. To adapt Tan's definition of a maximum stable matching to our setup, where a matching is a set of disjoint pairs and singletons formed by all the agents of a given set N, we must add to the maximum stable matching the deleted agents, so that all agents in the problem form part of that matching. One possibility is to consider the matching in which agents that are not paired in a stable manner are singletons. Another possibility is to matched some (or all) the singletons forming pairs among them. All these matchings are equally close to stability in the sense that they contain the same number of stable pairs. Hence the idea of maximum stability in our setting can be define as follows:

Definition 4 *(i) A set of agents* $T \subseteq N$ *form an* internally stable *set of pairs* μ_T *if there is no pair* $\{a_i, a_j\} \subseteq T$ *such that* $\{a_i, a_j\}$ *blocks* μ_T*. (ii) Matching* μ *is* maximum internally stable *if it contains the largest number of pairs which are internally stable.*

Example 2 (cont.) *Matching* $\mu = \{\{a_1, a_3\}, \{a_2\}, \{a_4, a_6\}, \{a_5, a_8\}, \{a_7, a_9\}, \{a_{10}\}\}$ *is maximum stable and internally stable, while matching* $\mu' = \{\{a_1, a_3\}, \{a_2, a_{10}\}, \{a_4, a_6\}, \{a_5, a_8\}, \{a_7, a_9\}\}$ *is maximum internally stable but not maximum stable.*

Considering the notion of stable partition, say $\mathcal{P}$, introduced above, Inarra *et al.* [16] define matchings associated to that partition, called $\mathcal{P}$*-stable matchings*. These matchings are formally defined as follows:

Let $\mathcal{P}$ be a stable partition. A $\mathcal{P}$*-stable matching* is a matching such that for each set $A = \{a_1, ..., a_k\} \in \mathcal{P}$, agent a_i is paired with either a_{i+1} or a_{i-1} for all $i \in \{1, ..., k\}$ except for a unique agent j who remains unmatched if A is odd or a singleton.[13] Hence, for each odd ring one agent is left out, and the rest of the agents of the ring are matched following its order, that is, they are matched to his subsequent or preceding agent. The reader may have noticed some similarities between $\mathcal{P}$-stable matchings and maximum internally stable matchings defined above. It turns out that the set of $\mathcal{P}$-stable matchings coincides with the set of maximum internally stable matchings that can be computed by Tan's algorithm.

Remark 3 *A $\mathcal{P}$-stable matching is maximum internally stable.*

The following result states some features of the $\mathcal{P}$-stable matchings obtained from the same stable partition $\mathcal{P}$.

Given two matchings μ and μ', we denote $\mu' RM$ if and only if μ' is obtained from μ by satisfying a blocking pair of μ (direct domination). We denote $\mu' R^T \mu$ if and only if there is a sequence of matchings $\mu = \mu_1, \ldots, \mu_k = \mu'$ such that for all $l \in \{1, \ldots, k-1\}$ μ_{l+1} is obtained from μ_l by satisfying a blocking pair of μ_l (indirect domination).

Remark 4 *Let $\mathcal{M}$ be the set of all matchings and let $\mathcal{P}$ be a stable partition. Consider $\mathcal{M}|_{\mathcal{P}} = \{\mu \in \mathcal{M} : \mu \text{ is a } \mathcal{P}\text{-stable matching of } \mathcal{P}\}$. Then (i) For any $\mu, \mu' \in \mathcal{M}|_{\mathcal{P}}$, $\mu R^T \mu'$. (ii) For all $a_i \in N$ belonging to an odd ring of $\mathcal{P}$, there exists a matching $\mu \in \mathcal{M}|_{\mathcal{P}}$ such that $\mu(a_i) = a_i$.*

As we shall see the $\mathcal{P}$-stable matchings play a significant role in Section 4.

3.3 Incompatibilities between solutions

In this subsection our purpose is to analyze if there exists a solution that could conciliate all the solutions presented above. Unfortunately we find that this is not possible. In what follows we show that the solution of almost stable matchings is incompatible with each of the other two solutions.

[13] Inarra *et al.* [16] show that from any matching there exists a sequence of blocking pairs reaching a $\mathcal{P}$-stable matching. See Roth and Vande Vate [27] and Diamantoudi *et al.* [11] for similar approaches to convergence to stability.

To prove the incompatibility between almost stable matchings and maximum internally stable matchings, we start showing the incompatibility between the latter and the family of Pareto optimal matchings in the following example.

Example 3 *Consider the following 8-agent problem:*

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
a_2	a_3	a_1	a_1	a_1	a_7	a_8	a_6
a_3	a_1	a_2	a_2	a_2	a_8	a_6	a_7
a_4	a_4	a_4	a_3	a_3	a_4	a_4	a_4
a_5	a_5	a_5	a_6	a_6	a_5	a_5	a_5
a_6	a_6	a_6	a_7	a_7	a_1	a_1	a_1
a_7	a_7	a_7	a_8	a_8	a_2	a_2	a_2
a_8	a_8	a_8	a_5	a_4	a_3	a_3	a_3
a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8

This example has a unique stable partition $\mathcal{P} = \{\{a_1, a_2, a_3\}, \{a_4, a_5\}, \{a_6, a_7, a_8\}\}$ with two odd rings $\{a_1, a_2, a_3\}$ and $\{a_6, a_7, a_8\}$. Every maximum internally stable matching has three stable pairs: one pair from each odd ring and the pair $\{a_4, a_5\}$. First, note that those matchings with singletons are not Pareto optimal since any other matching that joins them will Pareto dominate the original one. Hence we restrict our attention to those maximum internally stable matchings without singletons. There are nine of those matchings with three stable pairs each: $\mu_1 = \{\{a_1, a_2\}, \{a_3, a_6\}, \{a_4, a_5\}, \{a_7, a_8\}\}$; $\mu_2 = \{\{a_1, a_2\}, \{a_3, a_7\}, \{a_4, a_5\}, \{a_6, a_8\}\}$; $\mu_3 = \{\{a_1, a_2\}, \{a_3, a_8\}, \{a_4, a_5\}, \{a_6, a_7\}\}$; $\mu_4 = \{\{a_1, a_3\}, \{a_2, a_6\}, \{a_4, a_5\}, \{a_7, a_8\}\}$; $\mu_5 = \{\{a_1, a_3\}, \{a_2, a_7\}, \{a_4, a_5\}, \{a_6, a_8\}\}$; $\mu_6 = \{\{a_1, a_3\}, \{a_2, a_8\}, \{a_4, a_5\}, \{a_6, a_7\}\}$; $\mu_7 = \{\{a_1, a_6\}, \{a_2, a_3\}, \{a_4, a_5\}, \{a_7, a_8\}\}$; $\mu_8 = \{\{a_1, a_7\}, \{a_2, a_3\}, \{a_4, a_5\}, \{a_6, a_8\}\}$; $\mu_9 = \{\{a_1, a_8\}, \{a_2, a_3\}, \{a_4, a_5\}, \{a_6, a_8\}\}$. None of these matchings are Pareto-optimal since in each of them agents in the pair $\{a_4, a_5\}$ and agents in the pair coming from different odd rings can improve by rearranging their partners. Hence the following proposition can be established.

Proposition 1 *The intersection of Pareto-optimal matchings and maximum internally stable matchings may be empty.*

Since almost stable matchings are Pareto optimal the following corollary can be established.

Corollary 2 *The set of maximum internally matchings and the set of almost-stable matchings may have an empty intersection.*

This implies that the idea of finding a matching with the fewest number of blocking pairs is conflicting with the idea of finding a matching with the maximum number of stable pairs. In fact the example shows the well known trade off between Pareto optimality and stability.

Next, the incompatibility between almost stability and maximum irreversibility is shown in the following example.

Example 4 *Consider the following 8-agent problem:*

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
a_2	a_3	a_4	a_5	a_4	a_5	a_8	a_6
a_3	a_1	a_1	a_3	a_6	a_7	a_6	a_7
a_8	a_4	a_2	a_1	a_3	a_8	a_4	a_1
a_5	a_5	a_5	a_6	a_2	a_4	a_5	a_5
a_4	a_6	a_6	a_7	a_7	a_1	a_1	a_4
a_7	a_7	a_7	a_8	a_8	a_2	a_2	a_2
a_6	a_8	a_8	a_2	a_1	a_3	a_3	a_3
a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8

Matching $\mu = \{\{a_1, a_2\}, \{a_3, a_4\}, \{a_5, a_6\}, \{a_7, a_8\}\}$ *is the only one that is almost stable, being* $\{a_4, a_5\}$ *its unique blocking pair. However, pair* $\{a_4, a_5\}$ *is the maximum irreversible set and it is not contained in matching* μ*. Hence the following proposition can be established.*

Proposition 3 *The intersection of almost stable matchings and maximum irreversible matchings may be empty.*

Therefore almost stable matchings are incompatible with the other two solutions. Moreover Abraham *et al.* [2] show that the problem of finding a matching with the minimum number of blocking pairs is NP-hard and even hard to approximate. These results suggest to make a proposal for the roommate problem by conciliating maximum internally stable and maximum irreversible matchings.

4 $\mathcal{Q}$-stable matchings

Given the incompatibility between almost stable matchings and the other two solutions demostrated in the previous section, a natural question to ask is whether the intersection of the other two solutions is non-empty and the answer is in the affirmative. Matchings lying in the intersection are called *$\mathcal{Q}$-stable matchings* and are analyzed in this section.

Definition 5 *A matching is* Q*-stable if it is maximum internally stable, and maximum irreversible.*

As we have mentioned in the introduction our aim in this paper is to provide to the policy maker with a procedure to compute a Q-stable matching. In what follows we introduce an algorithm that does this job efficiently. The algorithm starts with a stable partition of a roommate problem and, by means of an iterative process, it removes from the preference lists those agents unable to form irreversible pairs. Then a stable partition with a maximum set of irreversible pairs is derived from which a Q-stable matching is finally obtained.

To proceed some additional notation is needed.

Given a stable partition $\mathcal{P}_t$. Let D_t be the set formed by the agents in the odd sets of $\mathcal{P}_t$, i.e. odd rings or singletons. Let S_t be the set of agents in pairs so that $N = D_t \cup S_t$. Let $\mathcal{P}_t|_{S_t}$ the set of pairs of partition $\mathcal{P}_t$. Notice that $\mathcal{P}_t|_{S_t}$ is a stable pairing of the agents in S_t.

The algorithm

Stage 1: Finding a maximum irreversible set of pairs

Step 1. Let $(N_1, (\succeq_{a_i}^{R_1})_{i \in N_1})$ be a problem where $N_1 = N$ and $(\succeq_{a_i}^{R_1}) = (\succeq_{a_i})$. Compute a stable partition $\mathcal{P}_1$ for $(N_1, (\succeq_{a_i}^{R_1})_{a_i \in N_1})$.[14]

Let $N_1 = D_1 \cup S_1$. If $S_1 = \emptyset$ then STOP and set $\mu_I = \mathcal{P}_1|_{S_1} = \emptyset$. If $S_1 = N_1$ then STOP and set $\mu_I = \mathcal{P}_1|_{S_1}$. Otherwise, for every agent $a_i \in S_1$ remove from $(\succeq_{a_i}^{R_1})$ every agent $a_k \in D_1$ and every agent $a_j \in S_1$, $a_j \neq a_i$, such that $a_k \succ_{a_i} a_j$ (a_i prefers a_k to a_j) for some $a_k \in D_1$. Go to next step.

Step t. Define a reduced problem $(N_t, (\succeq_{a_i}^{R_t})_{a_i \in N_t})$ where $N_t = S_{t-1}$ and $(\succeq_{a_i}^{R_t})$ is agent a_i's preference list after the clearing process over $(\succeq_{a_i}^{R_{t-1}})$. If no agent is removed from agent a_i's preference list, set $(\succeq_{a_i}^{R_t}) = (\succeq_{a_i}^{R_{t-1}})$. Compute a stable partition $\mathcal{P}_t$ for $(N_t, (\succeq_{a_i}^{R_t})_{a_i \in N_t})$ where $N_t = S_{t-1}$.

Let $N_t = D_t \cup S_t$. If $S_t = \emptyset$ then STOP and set $\mu_I = \mathcal{P}_t|_{S_t} = \emptyset$. If $S_t = N_t$ then STOP and set $\mu_I = \mathcal{P}_t|_{S_t}$. Otherwise, for every agent $a_i \in S_t$ remove from $(\succeq_{a_i}^{R_t})$ every agent $a_k \in D_t$ and every agent $a_j \in S_t$ $a_j \neq a_i$, such

[14] An algorithm which computes a stable partition in linear time can be found in Tan [32].

that $a_k \succ_{a_i} a_j$ (a_i prefers a_k to a_j) for some $a_k \in D_t$. Increase t and repeat this step.

Stage 2: Build a stable partition for $(N, \succeq)$.
Let I denote the set of matched agents in μ_I and let D denote the remaining agents. Join $\mathcal{P}_1 \mid_D$ with μ_I to determine a stable partition $\mathcal{P}^*$ on N. That is, $\mathcal{P}^* = \mathcal{P}_1 \mid_D \cup \mu_I$.

Stage 3: Build a matching from stable partition $\mathcal{P}^*$
From stable partition $\mathcal{P}^*$ derive a $\mathcal{P}^*$-stable matching. This matching is called Q^*-stable.

If the problem is solvable then the output of the algorithm is a stable matching and therefore it is immediate that it is maximum internally stable and maximum irreversible. That is,

Remark 5 *A stable matching is a* $\mathcal{Q}^*$*-stable.*

The previous remark allows us to focus on unsolvable problems. Hence from now on, we present some claims to prove that the algoritm provides a $\mathcal{Q}^*$-stable matching.

First, we present some claims, which are needed to show that the algorithm provides a matching with a set of irreversible pairs of maximum size.

Claim 1 *Suppose that* μ_I *is an irreversible set of pairs formed by a subset of agents* $I \subseteq N$ *and* $\mathcal{P} \mid_D$ *is a stable partition for the problem restricted to* $D = N \backslash I$. *Then* $\mathcal{P} = \mathcal{P} \mid_D \cup \mu_I$ *is a stable partition on* N.

Proof. No pair $\{a_i, a_j\} \subseteq I$ can block[15] $\mathcal{P}$ by the stability of μ_I, no pair $\{a_i, a_j\} \subseteq D$ can block by the stability of $\mathcal{P} \mid_D$, and no pair $\{a_i, a_j\}$ with $a_i \in I$ and $a_j \in D$ can block by the irreversibility of μ_I. ■

Claim 2 *If agent* a_i *either belongs to an odd ring or she is a singleton in a stable partition* $\mathcal{P}$ *then she can never be part of an irreversible set of pairs.*

Proof. Suppose for a contradiction that a_i is part of an irreversible matching μ_I. Then, by Claim 1, μ_I could be extended to a stable partition $\mathcal{P}' = \mathcal{P}_{N \backslash I} \cup \mu_I$. But then the set of odd rings and singletons would not be the same in $\mathcal{P}$ and $\mathcal{P}'$, contradicting Remark (ii) 1. ■

[15] Abusing of language we say that a pair of agents $\{a_i, b_j\}$ block partition $\mathcal{P}$ if $b_j \succ_{a_i} a_{i-1}$ then $b_{j-1} \succ_{b_j} a_i$.

Claim 3 *The set of pairs μ_I derived in Stage 1 of the algorithm, is maximum irreversible.*

Proof. μ_I is irreversible by construction since there is no agent $a_i \in I$ such that a_i prefers an agent a_k outside I to her current partner. It remains to prove the maximality of μ_I. That is, if there exists another irreversible set of pairs $\mu_{I'}$ with $I' \subseteq N$, then $I' \subseteq I$. By contradiction, suppose that there exists $\mu_{I'}$ such that $I' \setminus I \neq \emptyset$. Then there is an agent $a_i \in I'$ ($a_i \notin I$) who, by Claim 2, is part of some pair in any stable partition. This agent is removed at some step t of Stage 1 of the algorithm, and that implies the existence of an agent $a_j \in D_t$ such that $a_j \succ_{a_i} \mu_{I'}(a_j)$, contradicting that $\mu_{I'}$ is irreversible. ■

The previous claim and the argument in its proof imply the following corollary.[16]

Corollary 4 *For a roommate problem the set of agents matched in the largest irreversible set of pairs are the same.*

Matching $\mathcal{Q}^*$-stable matching is maximum irreversible by Claim 3 and maximum internally stable by Remark 3 then the following theorem can be established.

Theorem 5 *There always exists a $\mathcal{Q}^*$-stable matching for any roommate problem.*

Regarding the complexity of the algorithm the following result is established.

Proposition 6 *A $\mathcal{Q}^*$-stable matching can be computed in $O(mn)$ time, where n is the number of agents and m is the total length of the preference lists.*

Proof. Stage 1 can be invoked at most n times since the set of agents in pairs in the initial partition $\mathcal{P}_1$ can only shrink, and it is stopped when it does not shrink. The execution of each step takes linear time in m, which is the total length of the preference lists, since a stable partition can be found with Tan's algorithm [32] in $O(m)$ time, and the clearing process in Stage 1 can also be

[16] This corollary is closely related to Proposition 3 in Inarra *et al.* [17] although they are proven in different manners.

conducted in linear time. Therefore the algorithm terminates in $O(mn)$ time.[17]

■

Let us illustrate the algorithm and the previous results with an example:

Example 5 *Consider the following 10-agents problem:*

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}
a_2	a_3	a_1	a_3	a_4	a_7	a_5	a_9	a_{10}	a_{11}	a_8	a_1
a_3	a_1	a_2	a_5	a_6	a_5	a_6	a_4	a_8	a_9	a_{10}	a_2
a_4	a_6	a_4	a_6	a_7	a_1	a_1	a_5	a_1	a_4	a_1	a_3
a_5	a_4	a_5	a_2	a_1	a_9	a_2	a_{11}	a_6	a_5	a_2	a_4
a_6	a_5	a_6	a_1	a_2	a_2	a_3	a_1	a_3	a_1	a_3	a_5
a_7	a_7	a_7	a_8	a_3	a_3	a_4	a_2	a_4	a_2	a_4	a_6
a_8	a_8	a_8	a_7	a_8	a_8	a_8	a_3	a_5	a_3	a_5	a_7
a_9	a_9	a_9	a_9	a_9	a_4	a_9	a_6	a_7	a_6	a_6	a_8
a_{10}	a_{10}	a_{10}	a_{10}	a_{10}	a_{10}	a_{10}	a_7	a_2	a_7	a_7	a_9
a_{11}	a_{11}	a_{11}	a_{11}	a_{11}	a_{11}	a_{11}	a_{10}	a_{11}	a_8	a_9	a_{10}
a_{12}	a_{12}	a_{12}	a_{12}	a_{12}	a_{12}	a_{12}	a_{12}	a_{12}	a_{12}	a_{12}	a_{11}
a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}

Stage 1: *Finding a maximum irreversible set of pairs*

Step 1 *Computing a stable partition for the problem* $(N_1, (\succeq^{R_1})_{a_i \in N_1})$*,* $\mathcal{P}_1 = \{\{a_1, a_2, a_3\}, \{a_4, a_5\}, \{a_6, a_7\}, \{a_8, a_{11}\}, \{a_9, a_{10}\}, \{a_{12}\}\}$ *is obtained, where* $D_1 = \{a_1, a_2, a_3, a_{12}\}$ *and* $S_1 = \{a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, a_{11}\}$*. Remove from the list of preferences of each agent in* S_1 *all agents in* D_1 *and those that are less preferred, except for herself.*

Step 2 *A reduced problem* $(N_2, (\succeq^{R_2})_{i \in N_2})$*, is defined where* $S_1 = N_2$ *:*

a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}
a_4	a_4	a_7	a_5	a_9	a_{10}	a_{11}	a_8
	a_6	a_5	a_6	a_4	a_8	a_9	a_{10}
	a_7	a_6	a_7	a_5	a_9	a_4	a_{11}
	a_5			a_{11}		a_5	
				a_8		a_{10}	

[17] Tan and Hsueh [34] propose another algorithm, which constructs a stable partition incrementaly and whose complexity is $O(n^3)$ where n is the number of agents. This algorithm can be seen as a generalization of Roth and Vande Vate [27] procedure of convergence to a stable marriage.

Computing a stable partition for this reduced problem, $\mathcal{P}_2 = \{\{a_4\}, \{a_5, a_6, a_7\}, \{a_8, a_11\}, \{a_9, a_{10}\}\}$ *is obtained, where* $D_2 = \{a_4, a_5, a_6, a_7\}$ *and* $S_2 = \{a_8, a_9, a_{10}, a_{11}\}$*. Remove from the list of preferences of each agent in* S_2 *all agents in* D_2 *and those that are less preferred, except for herself.*

Step 3 *A reduced problem* $(N_3, (\succeq^{R_3})_{i \in N_3})$ *is defined where* $S_2 = N_3$:

a_8	a_9	a_{10}	a_{11}
a_9	a_{10}	a_{11}	a_8
a_8	a_8	a_9	a_{10}
	a_9	a_{10}	a_{11}

Computing a stable partition for the previous reduced problem $\mathcal{P}_3 = \{\{a_8, a_9\}, \{a_{10}, a_{11}\}\}$ *is obtained, where* $D_3 = \emptyset$ *and* $S_3 = \{a_8, a_9, a_{10}, a_{11}\}$*. Since* $S_3 = N_3$ *STOP. Set* $\mu_I = \{\{a_8, a_9\}, \{a_{10}, a_{11}\}\}$.

Stage 2: Let $I = \{a_8, a_9, a_{10}, a_{11}\}$ *and* $D = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_{12}\}$*. The following stable partition* $\mathcal{P}^* = \mathcal{P}_1 \mid_D \cup \mu_I = \{\{a_1, a_2, a_3\}, \{a_4, a_5\}, \{a_6, a_7\}, \{a_8, a_9\}, \{a_{10}, a_{11}\}, \{a_{12}\}\}$ *is determined.*

Stage 3: From $\mathcal{P}^*$*-stable partition let agent* a_1 *to be left out from the ring while agents* a_2 *and* a_3 *are matched preserving the ring ordering and the remaining agents are matched as in* $\mathcal{P}^*$*. The resulting matching* $\{\{a_1\}, \{a_2, a_3\}, \{a_4, a_5\}, \{a_6, a_7\}, \{a_8, a_9\}, \{a_{10}, a_{11}\}, \{a_{12}\}\}$ *is a* $\mathcal{Q}^*$*-stable matching.*

Note that we have started with partition $\mathcal{P}_1 = \{\{a_1, a_2, a_3\}, \{a_4, a_5\}, \{a_6, a_7\}, \{a_8, a_{11}\}, \{a_9, a_{10}\}, \{a_{12}\}\}$ *and have found that the* $\mathcal{P}_1$*-stable matching derived from it is not maximum irreversible. With the algorithm partition* $\mathcal{P}^* = \{\{a_1, a_2, a_3\}, \{a_4, a_5\}, \{a_6, a_7\}, \{a_8, a_9\}, \{a_{10}, a_{11}\}, \{a_{12}\}\}$ *is reached and a* $\mathcal{Q}^*$*-stable matching* $\{\{a_1\}, \{a_2, a_3\}, \{a_4, a_5\}, \{a_6, a_7\}, \{a_8, a_9\}, \{a_{10}, a_{11}\}, \{a_{12}\}\}$ *is obtained given that it is maximum internal stable and it contains the maximum irreversible set of pairs* $\{\{a_8, a_9\}, \{a_{10}, a_{11}\}\}$.

Until now we have overlooked the importance of maximizing the number of pairs matched in a matching. However, in many applications an essential objective is to match as many agents as possible. Think for example on the problem of dividing agents in a fixed number of two-person rooms or on the kidney exchange problem.[18] In those cases we can join the single agents of the

[18] For more details see the survey on market design for kidney exchange by Sönmez and Ünver [29].

$\mathcal{Q}^*$-stable matching outcome of the algorithm. In the example above matching $\{\{a_1, a_{12}\}, \{a_2, a_3\}, \{a_4, a_5\}, \{a_6, a_7\}, \{a_8, a_9\}, \{a_{10}, a_{11}\}\}$ is obtained.

Remark 6 *Every matching formed by joining mutually acceptable unmatched agents from a $\mathcal{Q}^*$-stable matching is also a $\mathcal{Q}^*$-stable matching.*

We have given an algorithm to compute a matching which lies in the intersection of the set of maximum internal stable matchings and the set of maximum irreversible stable matchings. It can be checked that these two solutions do not contain each other. In Example 6, $\mathcal{P}_1$-stable matching is maximum internally stable and not maximum irreversible while, for instance, matching $\{\{a_1\}, \{a_2\}, \{a_3\}, \{a_4, a_5\}, \{a_6, a_7\}, \{a_8, a_9\}, \{a_{10}, a_{11}\}, \{a_{12}\}\}$ is maximum irreversible and not maximum internally stable.

4.1 $\mathcal{Q}$-stable matchings and absorbing sets

For the roommate problem Inarra *et al.* [17] study to which matchings a decentralized process may lead to. They consider a dynamic process in which a matching is adjusted when a blocking pair of agents mutually decide to become partners. Either this change gives a stable matching or a new blocking pair of agents will generate another matching and so on. If there are stable matchings the process eventually converges to one of them. Otherwise it will lead to a set of matchings (an absorbing set) such that any matching in the set can be obtained from any other and it is impossible to escape from the matchings in that set.[19] As we have mentioned in the introduction it is important to investigate whether our proposal, the $\mathcal{Q}^*$-stable matching, is achievable from a free interactions of agents. That is, whether it belongs to an absorbing set, task that we undertake in this subsection.

A non-empty set of matchings $\mathcal{A}$ is an *absorbing set* if the following conditions hold: *(i) For all* $\mu, \mu' \in \mathcal{A}$ $(\mu \neq \mu')$, $\mu' R^T \mu$. *(ii) For all* $\mu \in \mathcal{A}$ there is no $\mu' \notin \mathcal{A}$ such that $\mu' RM$.

Condition *(i)* says that every matching in an absorbing set is (directly or indirectly) dominated by any other matching in the same set. Condition *(ii)* no matching in an absorbing set is directly dominated by a matching outside that set.

[19]Klaus *et al.* [21] prove that only the matchings in absorbing sets are *stochastically stable*.

The following remark states some properties of absorbing sets and their matchings.

Remark 7 *(i) Absorbing sets satisfy the property of* outer stability*, which requires that every matching not belonging to an absorbing set is (indirectly) dominated by the matchings of an absorbing set (Kalai* et al. *[19]). (ii) Every absorbing set contains a $\mathcal{P}$-stable matchings but not all $\mathcal{P}$-stable matchings belong to an absorbing sets (Inarra et al. [17]).*

Next, consider the relationship between the $\mathcal{Q}$-stable matchings and absorbing sets. We find that not all $\mathcal{Q}$-stable matchings belong to an absorbing set. In Example 1 all matchings with at least one pair of agents are Q-stable, however, matching $\mu_1 = \{\{2\}, \{3\}, \{1, 4\}\}$ does not belong to unique absorbing set. To see this let $\mu_2 = \{\{2, 3\}, \{1, 4\}\}$. It is immediate that $\mu_2 R^T \mu_1$ but not $\mu_1 R^T \mu_2$. Hence condition 1 of the definition of absorbing set is not satisfied. However, the matching provided by the algorithm belongs to an absorbing set.

Let $\mathcal{P}$ be a stable partition and denote by S_1 the set of pairs of $\mathcal{P}$ and BY?? $D_1 = N \backslash S_1$.

Claim 4 *Let μ be a $\mathcal{P}$-stable matching and let $U = S_1 \setminus I$ Then $\mu' R^T \mu$ where μ' is a matching such that $\mu'|_U(a_i) = a_i$ for all $a_i \in U$ and $\mu'|_{N\setminus U}(a_j) = \mu|_{N\setminus U}(a_j)$ for all $a_j \in N \setminus U$.*

Proof. We show that there is a sequence of matchings from μ to μ', in which all pairs in $\mu|_U$ become singles while the rest of agents are paired as in μ. Notice that each pair in $\mu|_U$ can be broken by (i) a single agent of an odd ring or (ii) an agent in U who has previously become single. Consider the following iterative process:

For t=1. Let $U_1 = \{a_i \in U : b_j \succ_{a_i} \mu(a_i)$ for some $b_j \in D_1\}$ and let $\mathcal{M}|_\mathcal{P}$ be the set of $\mathcal{P}$-stable matchings of $\mathcal{P}$ so that $\mu \in \mathcal{M}|_\mathcal{P}$. Thus, U_1 is the set of agents who block some matching in $\mathcal{M}|_\mathcal{P}$ with a single agent of an odd ring of $\mathcal{P}$.[20] Note that $U_1 \neq \emptyset$ otherwise $S_1 = I$ and we are done. Set $\mu = \mu^1$ and consider $\{a_i, a_{i+1}\} \in \mu^1|_{U_1}$ such that $b_i \succ_{a_i} a_{i+1}$ and $a_i \succ_{b_i} b_i$ with $b_i \in D_1$. W.l.o.g. assume that $\mu^1(b_i) = b_i$, otherwise by Remark 4 there exists another matching $\hat{\mu} \in \mathcal{M}|_\mathcal{P}$ such that $\hat{\mu} R^T \mu^1$

[20] By definition of stable partition no agent in U prefers a singleton of the partition to her partner in the partition and therefore this type of pairs cannot block any $\mathcal{P}$-stable matching.

and $\hat{\mu}(b_i) = b_i$. Matching μ^1 is blocked by $\{a_i, b_i\}$ forming matching μ_1^1 in which $\mu_1^1(a_{i+1}) = a_{i+1}$. By the stability of partition $\mathcal{P}$, $b_{i-1} \succ_{b_i} a_i$ and $b_i \succ_{b_{i-1}} \mu^1(b_{i-1}) = b_{i-2}$. Thus, matching μ_1^1 is blocked by $\{b_{i-1}, b_i\}$ forming matching μ_2^1 in which $\mu_2^1(a_i) = a_i$ and $\mu_2^1(a_{i+1}) = a_{i+1}$. We repeat this step for all pairs in $\mu|_{U_1}$ until their agents become singles.[21] Hence, we achieve a matching μ_k^1 such that $\mu_k^1(a_i) = a_i$ for all $a_i \in U_1$ and $\mu_k^1(a_j) = \mu^1(a_j)$ for all $a_j \in N \setminus U_1$ and $\mu_k^1 R^T \mu^1$. Then go to next step.

For t¿1. Let μ_l^{t-1} be the matching obtained at the end of Step $t-1$ such that $\mu_l^{t-1} R^T \mu$. Set $\mu_l^{t-1} = \mu^t$. Let $U_t = \{a_i \in U : a_j \succ_{a_i} \mu(a_i) \text{ for some } a_j \in U_{t-1}\}$. If $U_t = \emptyset$ then $\mu^t = \mu'$ and we are done. Otherwise there is a pair $\{a_l, a_{l+1}\} \in \mu|_{U \setminus U^{t-1}}$ where $U^{t-1} = \bigcup_{i=1}^{t-1} U_i$ such that $a_t \succ_{a_l} a_{l+1}$ and $a_l \succ_{a_t} a_t$ for some $a_t \in U^{t-1}$. Then matching μ^t is blocked by $\{a_l, a_t\}$ forming a new matching μ_1^t in which $\mu^t(a_{l+1}) = a_{l+1}$. By the stability of partition $\mathcal{P}$, $\mu(a_t) = a_{t-1} \succ_{a_t} a_l$ and $a_t \succ_{a_{t-1}} a_{t-1}$ so that μ_1^t is blocked by pair $\{a_t, a_{t-1}\}$ forming matching μ_2^t in which $\mu_2^t(a_l) = a_l$, $\mu_2^t(a_{l+1}) = a_{l+1}$ and $\mu_2^t(a_t) = a_{t-1}$. Thus, pair $\{a_t, a_{t-1}\}$ has been formed again and we need to split it. Suppose that a_{t-1} and a_t become singles for first time at Step j, then repeat the reasoning followed backwards from Step j until we reach a matching in which all agents in $U^{t-1} \cup \{a_l, a_{l+1}\}$ are alone, and the remaining agents are paired as in μ. Then iterate this step for all pairs in $\mu|_{U_t}$ until all agents in $U^t = U_t \cup U^{t-1}$ become singles. Hence, we achieve a matching μ_m^t such that $\mu_m^t(a_i) = a_i$ for all $a_i \in U^t$ and $\mu_m^t(a_j) = \mu(a_j)$ for all $a_j \in N \setminus U^t$ and $\mu_m^t R^T \mu$. Then increase t and repeat this step.

Since the number of agents in U is finite, the process finishes in finite time.

■

Proposition 7 *The $\mathcal{Q}^*$-stable matching obtained as the output of the algorithm belongs to an absorbing set.*

Proof. Let μ^* be the resulting matching of the algorithm. Assume that μ^* does not belong to an absorbing set. By Remark 7 and from the definition of absorbing set there exists a $\mathcal{P}$-stable matching μ in an absorbing set $\mathcal{A}$ such that $\mu R^T \mu^*$ but not $\mu^* R^T \mu$.

[21] Remark 4 can be extended to any set of matchings such that the agents in the odd rings are paired as in the set of $\mathcal{P}$-stable matchings and the rest of agents are paired equally in all matchings in the set.

Since $\mu R^T \mu^*$ then $\mu_I^* = \mu_I$ since μ_I^* is a maximum irreversible set of pairs. By Remark 1 $\mu^*|_{N\setminus U} = \mu|_{N\setminus U}$, hence $\mu^*|_U \neq \mu|_U$ otherwise μ and μ^* coincide.

By Claim 4 $\mu' R^T \mu$ such that $\mu'(a_i) = a_i$ for all $a_i \in U$ and $\mu'(a_j) = \mu(a_j)$ for all $a_j \in N \setminus U$. But since $\mu^*(a_i) \succ a_i$ for all $a_i \in U$, then it is easy to see that $\mu^*(a_i) R^T \mu' R^T \mu$, contradicting the initial assumption. ■

Remark 8 *Every matching formed by joining mutually acceptable unmatched agents in a $\mathcal{Q}^*$-stable matching belongs to an absorbing set.*

5 Discussion: Q-stability for hedonic games

To conclude we discuss some problems that share with the roommate problem the lack of stable outcomes and where the extension of Q-stable matchings to these setups can generate a good candidate as solution. The first problem we consider is the most natural extension of the roommate problem, the so called hedonic games introduced by Dréze and Greenberg **??**.

Consider a *hedonic game* $(N, (\succeq_i)_{i\in N})$ where N is the finite set of players, and $\succeq_i$ is a complete and transitive preference relation on subsets of agents that include agent i. A set $S \subseteq N$ is called a coalition. Let $S_i(N) = \{S \subseteq N : i \in S\}$ the set of coalitions that contain agent i. A *coalition structure* $\mathcal{S}$ on N is a partition of N into disjoint coalitions. We denote the set of all coalition structures by $\mathfrak{P}$. Let $S(i, \mathcal{S})$ be a coalition in $\mathcal{S}$ that contains agent i. Given $T \subseteq N$ and $\mathcal{S} \in \mathfrak{P}$ such that $\bigcup_{i\in T} S(i, \mathcal{S}) = T$, we denote $\mathcal{S}|_T$ the coalition structure on T.

A coalition structure $\mathcal{S}$ is *stable* if for every $S' \subseteq N$ there exists an agent $i \in S'$ such that $S(i, \mathcal{S}) \succ_i S'$.

It is well-known that for hedonic games a stable coalition structure may not exist. To cope with this lack of stability we propose Q-stable coalition structures as a solution. Thus, in this setting we introduce the following concepts.

Definition 6 *(i) A set of agents T form an* irreversible set *of (disjoint) coalitions $\mathcal{S}|_T$ if for all S' with $S' \cap T \neq \emptyset$ there exists an agent $i \in T$ such that $S(i, \mathcal{S}|_T) \succ_i S'$. (ii) A coalitional structure is* maximum irreversible *if it containsan an irreversible set of the largest possible size.*

Definition 7 *(i) A set of agents T form an* internally stable set *of (disjoint) coalitions $\mathcal{S}|_T$ if for every coalition $S' \subseteq T$ there exists and agent $i \in S'$ such* *that* $S(i, \mathcal{S}|_T) \succ_i S'$. *(ii) A coalitional structure is* maximum internally stable *if it contains an internally stable set of the largest possible size.*

The intersection of maximum irreversible coalition structures and maximum internally stable coalition structures determine the Q-stable matchings of these problems. Formally,

Definition 8 *A coalition structure is* Q*-stable if it is maximum internally stable and maximum irreversible.*

Example 6 *Consider the following 7-agent problem, where we only take into account those coalitions in which all its members prefer to form such coalitions to be single:*

1	2	3	4	5	6	7
$\{1,2\}$	$\{2,3\}$	$\{1,3,4\}$	$\{1,3,4\}$	$\{5,6,7\}$	$\{5,6\}$	$\{5,7\}$
$\{1,3,4\}$	$\{2,3,7\}$	$\{2,3\}$	$\{1,4\}$	$\{5,6\}$	$\{5,6,7\}$	$\{6,7\}$
$\{1,4\}$	$\{2,4\}$	$\{2,3,7\}$	$\{2,4\}$	$\{5,7\}$	$\{1,6\}$	$\{5,6,7\}$
$\{1,7\}$	$\{1,2\}$				$\{6,7\}$	$\{2,3,7\}$

In this hedonic game there is no stable coalition structure, since agents 1*,* 2*,* 3 *and* 4 *cannot be partitioned in a stable manner. Any partition containing* $\{1,2\}$ *is blocked by coalitions* $\{2,3\}$ *and* $\{2,4\}$*. Any partition containing* $\{2,3\}$ *or* $\{2,4\}$ *is blocked by coalition* $\{1,3,4\}$ *and any partition containing* $\{1,3,4\}$ *is blocked by* $\{1,2\}$*. On the one hand, the maximum irreversible set of coalitions is* $\mathcal{S}|_T = \{\{5,6,7\}\}$*. On the other hand, the set of agents* $V = \{1,3,4,5,6,7\}$ *form a maximum internally stable set of coalitions* $\mathcal{S}|_V = \{\{1,3,4\},\{5,6,7\}\}$*. Therefore the unique* Q*-stable coalition structure is* $\mathcal{S}^* = \{\{2\},\{1,3,4\},\{5,6,7\}\}$*.*

References

[1] Abraham, D. J., and Manlove, D. F. (2004). Pareto optimality in the roommates problem. Technical Report No. TR-2004-182 of the Computing Science Department of Glasgow University.

[2] Abraham, D. J., Biró, P., and Manlove, D. F. (2006). “Almost stable” matchings in the Roommates problem. In Approximation and Online Algorithms. 3879, 1-14.

[3] Biermann F. M. (2011) A measure to compare matchings in marriage markets. The Hebrew University of Jerusalem. Discussion Paper # 575.

[4] Biró, P., Cechlárová, K., and Fleiner, T. (2008). The dynamics of stable matchings and half-matchings for the stable marriage and roommates problems. International Journal of Game Theory, 36, 333-352.

[5] Biró, P., Irving, R., and Manlove, D.F. (2010) Popular matchings in the marriage and roommates problems. Proceedings of the seventh International Conference on Algorithms and Complexity (CIAC), 97-108.

[6] Bogomolnaia, A., and Jackson, M. O. (2002). The stability of hedonic coalition structures. Games and Economic Behavior, 38(2), 201-230.

[7] Can, B. and Klaus, B. (2013) Consistency and population sensitivity properties in marriage and roommate markets. Social Choice and Welfare, 41: 835-862.

[8] Cechlárová, K., and Ferková, S. (2004). The stable crews problem. Discrete applied mathematics, 140(1), 1-17.

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

[10] Dréze, J., and Greenberg, J., (1980). Hedonic coalitions: optimality and stability. Econometrica 48, 987–1003.

[11] Diamantoudi, E., Miyagawa, E., and Xue, L. (2004). Random paths to stability in the roommate problem. Games and Economic Behavior, 48(1), 18-28.

[12] Echenique F., and J. Pereyra (2013) "Strategic uncertainty and unraveling in matching markets" Caltech SS Working Paper 1373.

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

[14] Gudmundsson, J. (2013). When do stable roommate matchings exist? A review. Review of Economic Design, 1-11.

[15] Gusfield, D., and Irving, R. W. (1989). The stable marriage problem: structure and algorithms (Vol. 54). Cambridge: MIT press.

[16] Inarra, E., Larrea, C., and Molis, E. (2008). Random paths to P-stability in the roommate problem. International Journal of Game Theory, 36(3-4), 461-471.

[17] Inarra, E., Larrea, C., and Molis, E. (2013). Absorbing sets in roommate problems. Games and Economic Behavior, 81, 165–178.

[18] Jackson, M. O., and Watts, A. (2002). On the formation of interaction networks in social coordination games. Games and Economic Behavior, 41(2), 265-291.

[19] Kalai, E., Pazner, E.A., Schmeidler, D. (1976). Admissible outcomes of social bargaining processes as collective choice correspondence. Econometrica 63, 299-325.

[20] Klaus, B., and Klijn, F. (2010). Smith and Rawls share a room: stability and medians. Social Choice and Welfare, 35(4), 647-667.

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

[22] Kujansuu, E., Lindberg, T. and Makinen, E. (1999). The stable roommates problem and chess tournament pairings. Divulgaciones Matemáticas 7(1), 19-28.

[23] Manlove, D.F. (2013). Algorithms of Matching Under Preferences. World Scientific Publishing.

[24] Morrill, T. (2010). The roommates problem revisited. Journal of Economic Theory, 145(5), 1739-1756.

[25] Özkal-Sanver, İ. (2010). Impossibilities for roommate problems. Mathematical Social Sciences, 59(3), 360-363.

[26] Pittel, B. G., and Irving, R. W. (1994). An upper bound for the solvability probability of a random stable roommates instance. Random Structures & Algorithms, 5(3), 465-486.

[27] Roth, A. E., and Vande Vate, J. H. V. (1990). Random paths to stability in two-sided matching. Econometrica 58: 1475-1480.

[28] Roth, A. E., Sönmezc, T. and Ünverd, M. U. (2005) Pairwise kidney exchange Journal of Economic Theory 125: 151 − 188

[29] Sönmez, T., and Ünver, M. U. (2010). Market design for kidney exchange. Oxford Handbook of Market Design, Oxford University Press.

[30] Sotomayor, M. (2011). The Pareto-stability Concept is a Natural Solution Concept for Discrete Matching Markets with Indifferences. International Journal of Game Theory, 40(3), 631-644.

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

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

[33] Tan, J. J. (1991b). Stable matchings and stable partitions. International Journal of Computer Mathematics, 39(1-2), 11-20.

[34] Tan, J. J., and Hsueh, H-C. (1995). A generalization of the stable matching problem. Discrete Applied Mathematics, 59(1), 87-102.

[35] Toda, M. (2006). Monotonicity and consistency in matching markets. International Journal of Game Theory, 34, 13-31.

6 Appendix[22]

Since the notion of stable partition plays a significant role in the current work and his interpretation is not that easy, we present informally this algorithm described as an acceptance-rejection procedure, which is illustrated with a numerical example, and then we give its formal definition.

Let $(N, \succeq)$ be a problem where $N = \{a_1, ..., a_n\}$ is the set of agents. Let $(N_k, (\succeq)^k)$ be the restricted problem where $N_k = \{a_1, ..., a_k\}$ and $(\succeq)^k$ is the preference list of agents in N_k in which $N \backslash N_k$ agents have been deleted. Assume that we have already found a stable partition, say $\mathcal{P}_k$, a for $(N_k, (\succeq)^k)$, $1 \leq k \leq n-1$ and that one additional agent a_{k+1} is added. The question is whether a stable partition $\mathcal{P}_{k+1}$ for the enlarged problem $(N_{k+1}, (\succeq)^{k+1})$ can be determined. The answer in in the affirmative. The following acceptance-rejection procedure determines it:

Let problem $(N_k, (\succeq)^k)$ and let a_{k+1} be the newcomer. By embedding agent a_{k+1} into the existent lists and adding her own list to problem $(N_k, (\succeq)^k)$, problem $(N_{k+1}, (\succeq)^{k+1})$ is constructed. (Note that $\{a_1\}$ is the unique stable partition for $(N_1, (\succeq)^1))$. Given a stable partition $\mathcal{P}_k$ for $(N_k, (\succeq)^k)$ let agent a_{k+1} propose to the set of agents in N_k according to her preference order:

1. If nobody accepts her proposal, then a_{k+1} is alone and stable partition $\mathcal{P}_{k+1} = \mathcal{P}_k \cup \{a_k\}$ is obtained.

2. If a_{k+1} is accepted by agent x there are three possible cases:

(i) If x is unmatched in $\mathcal{P}_k$ then $\mathcal{P}_{k+1} = (\mathcal{P}_k \backslash \{x\}) \cup \{x, a_{k+1}\}$ and stable partition $\mathcal{P}\backslash_{k+1}$ for $(N_{k+1}, (\succeq)^{k+1}$ is obtained.

(ii) If x is currently in an odd ring, say $(a_1, a_2, ..., a_{2m}, x)$, then the arrival of a_{k+1} decomposes the set into pairs and $\mathcal{P}_{k+1} = \{\{a_1, a_2\}, \{a_3, a_4\}....\{a_{2m-1}, a_{2m}\}\} \cup \{\{a_{k+1}, x\}\}$ becomes a stable partition for $(N_{k+1}, (\succeq)^{k+1}$.

(iii) If x is in a mutually aceptable pair say $\{x, y\}$ in $\mathcal{P}_k$ then y becomes single ans is the new proposer. In this phase a proposal-rejection sequence takes place which may stops in 1 in 2 (i) or 2 (ii). In both cases the desired stable partition comes out. Otherwise an agent who made a proposal once

[22] The content of this appendix follows Section 4.3 in Manlove [23]. (See this section for a detailed analysis)

receives a proposal later and repetition takes place. Then stop. All agents involved in the cycle form a set and stable partition P_{k+1} for problem $(N_{k+1}, (\succeq)^{k+1})$ is constructed.

Example 7 *Consider the following 7-agent problem:*

a_1	a_2	a_3	a_4	a_5	a_6	a_7
a_3	a_1	a_4	a_2	a_6	a_7	a_5
a_2	a_4	a_2	a_5	a_4	a_5	a_6
a_4	a_3	a_1	a_3	a_7	a_2	a_1
a_5	a_6	a_7	a_7	a_2	a_3	a_2
a_6	a_7	a_5	a_6	a_3	a_1	a_3
a_7	a_5	a_6	a_1	a_1	a_4	a_4
a_1	a_2	a_3	a_4	a_5	a_6	a_7

W.l.o.g assume that agents arrive to the process in the following arbitrary order: $a_1, a_2, a_3, ..., a_7$.

- *a_1 arrives and forms a stable partition $\{a_1\}$ for $(N_1, (\succeq)^1)$.*
- *a_2 arrives and proposes to a_1. Since a_1 rather matches a_2 than being alone, accepts the proposal and $\{a_1, a_2\}$ is stable partition for $(N_2, (\succeq)^2)$ (see Case 2 (i)).*
- *a_3 arrives and proposes to a_2 who rejects then proposes to a_1 who accepts. Pair $\{a_1, a_3\}$ forms and a_2, now alone, proposes to a_1 who rejects and then to a_3 who accepts. Pair $\{a_2, a_3\}$ and a_1 is abandoned. This agent proposes a_3 who rejects and then to a_2 who accepts and pair $\{a_1, a_2\}$ is formed and a_3 is a proposer again. The cycling agents get together in the set $\{a_1, a_2, a_3\}$ and stable partition $\{\{a_1, a_2, a_3\}\}$ for $(N_3, (\succeq)^3)$ is obtained (see Case 2 (iii)).*
- *a_4 arrives and proposes to a_2 who accepts. Stable partition $\{\{a_1, a_3\}, \{a_2, a_4\}\}$ for $(N_4, (\succeq)^4)$ is obtained (see Case 2 (ii)).*
- *a_5 arrives and is rejected for all agents in the process. Hence stable partition$\{\{a_1, a_3\}, \{a_2, a_4\}, \{a_5\}\}$ for $(N_5, (\succeq)^5)$ is obtained. (See Case 1)*
- *a_6 arrives and proposes to a_5 who accepts. Stable partition $\{\{a_1, a_3\}, \{a_2, a_4\}, \{a_5, a_6\}\}$ for $(N_6, (\succeq)^6)$ is obtained (see Case 2 (i)).*

- *a_7 arrives and proposes to a_5 who rejects, then to a_6 who acepts and form: $\{\{a_1, a_3\}, \{a_2, a_4\}, \{a_6, a_7\}\{a_5\}\}$. a_5 proposes to a_4 who rejects and then to a_7 who accepts forming $\{\{a_1, a_3\}, \{a_2, a_4\}, \{a_5, a_7\}\{a_6\}\}$. a_6 proposes to a_7 who rejects and to a_5 who accepts forming $\{\{a_1, a_3\}, \{a_2, a_4\}, \{a_6, a_7\}\{a_5\}\}$. Then a_5 proposes to a_6 who rejects, then to a_4 who also rejects and then to a_7 who accepts forming $\{\{a_1, a_3\}, \{a_2, a_4\}, \{a_5, a_6\}\{a_7\}\}$ and we reach a cycle. The cycling agents get together $\{a_5, a_6, a_7\}$ and stable partition $\{\{a_1, a_3\}, \{a_2, a_4\}, \{a_5, a_6, a_7\}\}$ for $(N_7, (\succeq)^7) = (N, (\succeq))$ is obtained for (see Case 2 (iii))*

  The outcome is stable by construction: there is not a pair of agents belonging to different sets or within a set who block the stable partition.