THE ROOMMATE PROBLEM WITH EXTERNALITIES

JOSÉ LUIS CONTRERAS AND JUAN PABLO TORRES-MARTÍNEZ

Abstract. This paper extends the roommate problem to include externalities, allowing preferences for a partner to depend on the situation of the others. Assuming that everyone has prudent expectations about other agents' reactions to deviations, stability concepts for matchings and partitions of the set of agents are proposed and characterized. We prove that any roommate problem with externalities has a stable partition and that a stable matching exists if there is a stable partition without odd rings. These results allow us to find restrictions on the space of preferences ensuring the existence of a stable matching. We also show that some classical properties are lost in the presence of externalities: the existence of paths to stability from any unstable matching, the coincidence of the core with the set of stable matchings, and the invariance of the set of agents that are alone in a stable matching.

Keywords. Roommate problems - Externalities - Stable matching - Stable partition
JEL Classification. D62, C78.

1. Introduction

The roommate problem, introduced by Gale and Shapley (1962), is a one-to-one matching problem where agents may form pairs among them. It extends the classical marriage market framework allowing individuals to form couples without dividing the population into two disjoint sets. Hence, several situations with economic and social interest are captured: marriage markets with same-sex unions, vertical/horizontal integration of firms, the formation of pairs of students/workers to carry out an assignment, or the bilateral trading of commodities, to mention some examples.

As shown by Gale and Shapley (1962), the absence of a two-sided demographic structure compromises the existence of a *stable matching* in the roommate problem, i.e., an individually rational distribution of agents in couples that is stable with respect to bilateral deviations. This negative result motivates the search for alternative solution concepts. In this direction, Tan (1991) and Tan and Hsueh (1995) show that any roommate problem has a *stable partition*. That is, a partition of the set of agents into ordered groups that associates to each individual an acceptable potential partner within her group in such form that the following stability property holds: there are no blocking pairs of a matching among the agents paired with individuals as preferred as their potential partners.[1] Tan (1991) and Tan and Hsueh (1995) also find a necessary and sufficient condition for the existence of a stable matching:

Date: May, 2018.

We are grateful to Milton Braitt, María Haydée Fonseca, Daniel Jaar, Adriana Piazza, Francisco J. Pino, and Matteo Triossi for helpful comments and discussions.

José Luis Contreras, National Productivity Commission, Chile. e-mail: jlcontreras@cnp.gob.cl

Juan Pablo Torres-Martínez, Department of Economics, University of Chile. e-mail: juan.torres@fen.uchile.cl.

[1]In a stable partition, a group of agents $\{a_1, \ldots, a_k\}$ is *ordered* in the sense that a_{i-1} is the potential partner of a_i for each $i \in \{1, \ldots, k\}$ (subscripts modulo k). Also, when $k \geq 3$, each a_i prefers to be matched with a_{i+1} than with a_{i-1}.

the absence of stable partitions with odd rings, i.e., non-trivial groups of agents with an odd number of members. Departing from this characterization, it is possible to find restrictions on the space of preferences ensuring the existence of a stable matching. For instance, the *no-odd-rings condition* of Chung (2000) or the *symmetric utilities hypothesis* of Rodrigues-Neto (2007).

Although not all roommate problems have a stable matching, it is worth noting that some interesting properties are satisfied when this happens. First, the set of stable matchings coincides with the *core*, which is the set of individually rational matchings that are stable with respect to group deviations (cf., Alcalde (1994)). Second, there always exist paths to stability, as for any unstable matching we can reach a stable outcome through a finite sequence of blocking pairs (cf., Chung (2000), Diamantoudi, Miyagawa, and Xue (2004)). Third, the Lone Wolf Theorem holds, because an agent that is alone in a stable matching will remain alone in any other stable matching (cf., Gusfield and Irving (1989), Klaus and Klijn (2010)).

However, classical roommate problems ignore situations where individuals recognize that their wellbeing is affected by the other couples formed. For instance, in oligopolistic markets, firms decide between vertical or horizontal integration considering the organizational structure of the economy. In labor markets, to choose a partner to work with, the relative importance of factors as personality or technical skills depends on the characteristics of the teams that are competing. In marriage markets, class/caste systems make preferences for a couple dependent on other couples' attributes.

Motivated by these possibilities, we extend the classical roommate problem to include externalities, addressing the validity of the results described above when preferences for a partner may depend on the situation of others. Externalities are arbitrary and are incorporated assuming that everyone has preferences over the set of matchings, instead of over the set of agents. Inspired by Sasaki and Toda (1996), we concentrate our analysis on environments where agents do not have any information about other agents' preferences and, therefore, they are *prudent* about other agents' reactions to deviations.[2] That is, two agents block a matching only if their situation improves in any scenario where they form a pair (i.e., independently of later deviations from others). Using this blocking concept, we introduce and characterize stability concepts for both matchings and partitions of the set of agents.

Although it may seem difficult that prudent agents form a blocking pair, roommate problems with externalities may not have a stable matching. However, a stable partition always exists (see Theorem 1). Even more, any stable partition without odd rings induces a stable matching and, therefore, we can find sufficient conditions on the space of preferences to guarantee that each roommate problem has an efficient and stable outcome (see Theorems 2 and 3).

As we show in several examples, unlike what happens in classical roommate problems, the inclusion of externalities compromises the existence of paths to stability, the core may be an empty-set even when there are stable matchings, and the Lone Wolf Theorem does not necessarily hold. Furthermore, the necessary and sufficient condition for the existence of a stable matching is lost in the presence of externalities: stable partitions with odd rings and stable matchings may coexist. Hence, even in an

[2]In our framework, the prudence of agents is a necessary and sufficient condition to guarantee solvability in the entire subclass of marriage problems with externalities (see Sasaki and Toda (1996, Proposition 3.1 and Theorem 4.1)).

environment where individuals are conservative when evaluating a deviation, the presence of externalities can have a profound effect on the characteristics of stable matchings.

In our model, a stable partition exists even without prudence. That is, we may allow agents to have information about other agents' preferences and to form heterogeneous and sophisticated beliefs about their reactions to deviations (see the remark after Theorem 1). However, as stable matchings are required to be compatible with individual beliefs, the relationship between stable partitions without odd rings and stable matchings—a key property to determine sufficient conditions for solvability of the roommate problem—crucially depends on the prudence of agents (see the remark after Theorem 2).

This paper is organized as follows. In the next section we describe the roommate problem with externalities and characterize the existence of stable partitions and stable matchings. In Section 3 we analyze the effects of externalities on the structure of the roommate problem. In Section 4 we recover, as byproducts of our main results, well-known characterizations of stability for marriage markets and for roommate problems without externalities. Section 5 is devoted to some concluding remarks.

2. Roommate Problems with Externalities

A roommate problem with externalities $(N,(\succ_a)_{a\in N})$ is characterized by a finite set of agents N and a preference profile $(\succ_a)_{a\in N}$ over the set of matchings $\mathcal{M}=\{\mu:N\to N:\mu(\mu(a))=a,\,\forall a\in N\}$, where each preference $\succ_a$ is strict, complete, and transitive. In this context, an agent can form a pair with any other individual or may remain alone. However, in contrast to the classical roommate problem, her preferences for a partner may depend on the situation of the others.

We assume that everyone has *prudent expectations* about other agents' reactions to deviations and, therefore, a pair of agents will block a matching only if their situation improves in any scenario where they are coupled, i.e., they improve independent of later deviations by other people.

Definition 1. A matching $\mu\in\mathcal{M}$ is *blocked by a pair* $(a,b)\in N\times N$ when $\rho\succ_a\mu$ and $\rho\succ_b\mu$ for all $\rho\in\mathcal{M}$ such that $\rho(a)=b$.[3] A matching μ is *stable* if it does not have any blocking pair.

The following example shows that the set of stable matchings may be empty.

Example 1. Let $(N,(\succ_a)_{a\in N})$ be a roommate problem characterized by $N=\{1,2,3\}$ and

$$\mu_3\succ_1\mu_2\succ_1\mu_1\succ_1\mu_0,\qquad \mu_1\succ_2\mu_3\succ_2\mu_2\succ_2\mu_0,\qquad \mu_2\succ_3\mu_1\succ_3\mu_3\succ_3\mu_0,$$

where μ_0 is the matching where all agents are single and μ_i is the matching where i is the only agent alone, with $i\in\{1,2,3\}$. In this context, a stable matching does not exist. Indeed, μ_0 is blocked by $(2,3)$, μ_1 is blocked by $(1,3)$, μ_2 is blocked by $(1,2)$, and μ_3 is blocked by $(2,3)$. □

Since roommate problems with externalities may not have stable matchings, we will focus first on stability properties of ordered *partitions* of the set of agents, i.e., collections of pairwise disjoint, ordered, and nonempty subsets of N covering N.

[3] The case where a blocks μ to become alone is captured by taking $b=a$.

We need some definitions. Let $\mathcal{M}(a,b) = \{\mu \in \mathcal{M} : \mu(a) = b\}$ be the set of matchings where a and b form a couple. Agent a considers b *acceptable* when she cannot block a matching in $\mathcal{M}(a,b)$ becoming alone. Also, a considers that b *is more interesting than* c, with $b \neq c$, when she cannot block a matching in $\mathcal{M}(a,b)$ by forming a pair with c. Given $k \geq 3$, an ordered set $\{a_1,\dots,a_k\} \subseteq N$ is a *ring* when, for every $i \in \{1,\dots,k\}$, a_i considers that a_{i+1} is more interesting than a_{i-1}, although both are acceptable to her, where subscripts are modulo k.[4] An *odd ring* is a ring with an odd number of elements.

A matching μ *is induced by a partition* P, denoted by $\mu \in \mathcal{M}(P)$, when the following properties hold:

- If $\{a_1,\dots,a_{2k}\} \in P$, then $\mu(a_r) \in \{a_{r-1}, a_{r+1}\}$ for all $r \in \{1,\dots,2k\}$, where $k \geq 1$.
- If $\{a_1,\dots,a_{2k+1}\} \in P$, then there is an agent a_i which is alone under μ and $\mu(a_r) \in \{a_{r-1}, a_{r+1}\}$ for all $r \in \{1,\dots,2k+1\}$, $r \neq i$, where $k \geq 0$.

DEFINITION 2. Given a roommate problem with externalities $(N, (\succ_a)_{a\in N})$, a partition P of the set of agents is *stable* when the following properties hold:

(i) Each element of P is a ring, a pair of mutually acceptable agents, or a singleton.

(ii) If $\mu \in \mathcal{M}(P)$ is an unstable matching, then its blocking pairs always include an agent that is alone under μ and belongs to a ring of P.

To give a more intuitive description of the stability of partitions, suppose that when P is formed agents develop expectations about the partner they will have. Hence, if $\{a_1 \dots, a_k\} \in P$, a_i expects that a matching in $\mathcal{M}(a_i, a_{i+1})$ being realized, although she does not discard the realization of a matching in $\mathcal{M}(a_i, a_{i-1})$. With this interpretation, the stability of P can be understood as a requirement aligning expectations with preferences. In particular, any pair of agents that block a matching induced by P includes an individual a_i whose expectations were frustrated, as she was not matched with a_{i+1} or a_{i-1}. That is, a_i is alone and belongs to a ring of P.[5]

THEOREM 1. *Any roommate problem with externalities* $(N, (\succ_a)_{a\in N})$ *has a stable partition.*

Proof. Given $a, b \in N$, let $\mu_{a,b}$ be the least preferred matching on $\mathcal{M}(a,b)$ by a. Define a preference relationship $\succ_a^*$ over N by the rule $b \succ_a^* c$ if and only if $\mu_{a,b} \succ_a \mu_{a,c}$. Notice that $\succ_a^*$ is a complete, transitive, and strict relation of preferences.

[4]All through the text, properties in which subscripts are in a set $\{1,\dots,m\}$ hold *modulo m*.

[5]For classical roommate problems, Pęski (2017) describes stable partitions through bijective mappings $f : N \to N$, that he refers as *improper matchings* when $f(f(a)) \neq a$ for some $a \in N$. Underlying his approach is the following one-to-one correspondence between ordered partitions P and bijections $f : N \to N$: $\{a_1,\dots,a_k\} \in P$ if and only if $f(a_i) = a_{i-1}$ for all $i \in \{1,\dots,k\}$ and $k \geq 1$. Since this relationship does not depend of the presence of externalities, we can also describe stable partitions in term of self-bijections of the set of agents. Indeed, a partition P is stable, in the sense of Definition 2, if and only if the bijection $f : N \to N$ induced by P satisfies the following properties:
(i) f associates to each a a potential mate $f(a)$ that is acceptable to her.
(ii) If $f(a) \neq f^{-1}(a)$, then a considers that $f^{-1}(a)$ is more interesting than $f(a)$.
(iii) If a forms a pair with $f(a)$ or $f^{-1}(a)$ in a matching $\mu \in \mathcal{M}(P)$, then a cannot block μ deviating with an agent b that forms a pair with $f(b)$ or $f^{-1}(b)$ in μ.

Given $m \in \mathbb{N}$, let $I_m = \{1, \ldots, m\}$. Since $(N, (\succ^*_a)_{a \in N})$ is a roommate problem without externalities, it follows from Tan and Hsueh (1995, Corollary 3.8) that there is a partition P such that:

(1) If $\{a_1, \ldots, a_k\} \in P$ and $k \geq 3$, then $a_{i+1} \succ^*_{a_i} a_{i-1} \succ^*_{a_i} a_i$ for each $i \in I_k$;

(2) If $\{a_1, a_2\} \in P$, then $a_{i-1} \succ^*_{a_i} a_i$ for each $i \in I_2$;

(3) Given $\{a_1, \ldots, a_k\}$ and $\{b_1, \ldots, b_l\}$ in P, not necessarily different, the following condition holds for each $(i, j) \in I_k \times I_l$ such that $b_j \neq a_{i+1}$: if $b_j \succ^*_{a_i} a_{i-1}$ then $b_{j-1} \succ^*_{b_j} a_i$;

where all subscripts are modulo k or l, when it is appropriate.

The transitivity of preferences implies that $b \succ^*_a c$ if and only if there is $\mu \in \mathcal{M}(a, c)$ such that $\rho \succ_a \mu$ for all $\rho \in \mathcal{M}(a, b)$. Thus, (1) and (2) imply that each element of P is a ring, a pair of mutually acceptable agents, or a singleton. Furthermore, for any pair of sets $\{a_1, \ldots, a_k\}$ and $\{b_1, \ldots, b_l\}$ in P, not necessarily different, and for each $(i, j) \in I_k \times I_l$ such that $b_j \neq a_{i+1}$, we have that:

$$
\begin{aligned}
&\exists \mu_i \in \mathcal{M}(a_i, a_{i+1}) \cup \mathcal{M}(a_i, a_{i-1}), \forall \mu \in \mathcal{M}(a_i, b_j) : \ \mu \succ_{a_i} \mu_i \\
&\quad \Longrightarrow \quad \exists \mu_i \in \mathcal{M}(a_i, a_{i-1}), \forall \mu \in \mathcal{M}(a_i, b_j) : \ \mu \succ_{a_i} \mu_i \\
&\quad \Longleftrightarrow \quad b_j \succ^*_{a_i} a_{i-1} \quad \Longrightarrow \quad b_{j-1} \succ^*_{b_j} a_i \quad \Longleftrightarrow \quad \exists \rho_j \in \mathcal{M}(a_i, b_j), \forall \rho \in \mathcal{M}(b_j, b_{j-1}) : \ \rho \succ_{b_j} \rho_j \\
&\quad \Longrightarrow \quad \exists \rho_j \in \mathcal{M}(a_i, b_j), \forall \rho \in \mathcal{M}(b_j, b_{j-1}) \cup \mathcal{M}(b_j, b_{j+1}) : \ \rho \succ_{b_j} \rho_j.
\end{aligned}
$$

Let $\mu \in \mathcal{M}(P)$ and (a_i, b_j) a pair of agents that block it. If $\{a_i\} \in P$ or $\{a_i, a_{i+1}\} \in P$, then $b_j \neq a_{i+1}$. Hence, the property above implies that, if b_j belongs to a singleton or is paired under μ, then he does not want to form a pair with a_i to block μ. That is, b_j is alone under μ and belongs to a ring of P.

Alternatively, suppose that a_i belongs to a ring of P and is paired under μ. If $b_j \neq a_{i+1}$, then the property above implies that b_j does not want to block μ unless he is alone under μ and belongs to a ring of P. If $b_j = a_{i+1}$, then b_j is alone under μ, as otherwise he could block a matching in $\mathcal{M}(b_j, b_{j+1})$ by deviating with b_{j-1}, a contradiction with the fact that b_j considers b_{j+1} more interesting than b_{j-1}.

Therefore, if (a_i, b_j) is a blocking pair of $\mu \in \mathcal{M}(P)$ and either a_i is paired under μ or belongs to a singleton of P, then b_j is alone under μ and belongs to a ring of P. □

We implicitly assume that agents do not have sophisticated conjectures about other players' reactions to deviations. However, to some extent, the existence of a stable partition does not depend on this assumption. Indeed, let $\mathcal{M}_a(b) \subseteq \mathcal{M}(a, b)$ be the set of matchings representing agent a's beliefs about the scenarios that may arise when she decides to form a pair with b. Redefine the set of matchings induced by P to include compatibility with individual beliefs: $\mu \in \mathcal{M}(P)$ if and only if both μ is induced by P and $\mu \in \mathcal{M}_a(\mu(a))$ for all $a \in N$. Also, extend the definitions of *acceptable agent* and *ring* by assuming that to block a matching μ a pair $(a, b) \in N \times N$ requires that both $\rho \succ_a \mu$, $\forall \rho \in \mathcal{M}_a(b)$ and $\rho \succ_b \mu$, $\forall \rho \in \mathcal{M}_b(a)$. Then, the same arguments made in the proof of Theorem 1 ensure that any roommate problem with externalities has a stable partition.[6]

A matching μ is *complete* when there are no agents alone, i.e., $\mu(a) \neq a$ for all $a \in N$, and it is *efficient* when there does not exist $\eta \in \mathcal{M}$ such that $\eta \succ_a \mu$ for all $a \in N$. Notice that, a complete matching exists if and only if $\#N$ is even, while an efficient outcome always exists because $\mathcal{M}$ is finite and individual preferences are transitive.

[6]In the adapted proof, $\mu_{a,b}$ must be defined as the least preferred matching on $\mathcal{M}_a(b)$.

The following result relates the existence of stable partitions with the stability of matchings, showing properties that have analogous counterparts in the literature without externalities (cf. Tan (1991), Tan and Hsueh (1995), Sasaki and Toda (1996)).

Theorem 2. *In any roommate problem with externalities* $(N, (\succ_a)_{a\in N})$ *we have that:*
(i) Any stable partition without odd rings induces a stable matching.
(ii) Any stable partition without odd rings or singletons induces a complete stable matching.
(iii) If there exists a stable matching, then there exists a stable matching that is efficient.

Proof. If P is a stable partition without odd rings, then it follows from Definition 2(ii) that a matching induced by P is stable. Moreover, in the absence of singletons, the partition P only induces complete matchings. These arguments guarantees the validity of properties (i) and (ii).

Item (iii) follows from Theorem 4.3 in Sasaki and Toda (1996). These authors work in a *marriage market* with externalities in which agents never become unpaired. However, their proof of the existence of an efficient and stable matching does not depend on either the two-sided structure or the impossibility of agents to become unpaired. □

When individuals have sophisticated beliefs $\{\mathcal{M}_a(b)\}_{(a,b)\in N\times N}$ (see the remark after Theorem 1) the results of Theorem 2(i)-(ii) do not hold anymore: there are roommate problems without stable matchings that have stable partitions without odd rings and singletons. Intuitively, since a matching μ induced by a partition is required to be compatible with individuals' beliefs (i.e., $\mu \in \mathcal{M}_a(\mu(a))$ for all $a \in N$), it is possible that $\mathcal{M}(P) = \emptyset$. Formally, the property follows from Sasaki and Toda (1996, Proposition 3.1), as the marriage market with externalities is a particular case of our model (see Section 4).

The next result determines conditions on preferences ensuring the existence of an efficient and stable matching. Notice that, as $N \times \mathcal{M}$ is a finite set, a preference profile $(\succ_a)_{a\in N}$ always has a functional representation. That is, there is $\Phi : N \times \mathcal{M} \to \mathbb{R}$ such that $\mu \succ_a \eta$ if and only if $\Phi(a, \mu) > \Phi(a, \eta)$.

Theorem 3. *A roommate problem with externalities* $(N, (\succ_a)_{a\in N})$ *has an efficient and stable matching when any of the following conditions hold:*
(i) The preference profile $(\succ_a)_{a\in N}$ *has no odd rings.*
(ii) $(\succ_a)_{a\in N}$ *has a representation satisfying* $\Phi(a, \mu) = \Phi(\mu(a), \mu), \forall (a, \mu) \in N \times \mathcal{M}$.

Proof. It follows from Theorem 1 that $(N, (\succ_a)_{a\in N})$ has a stable partition P. If (i) holds, then P does not have odd rings, and the result follows from Theorem 2. If (ii) holds, then the result follows from the previous case as $(\succ_a)_{a\in N}$ has no rings. Indeed, by contradiction, suppose that $\{a_1, \ldots, a_k\}$ is a ring. Since for every $i \in I_k$ the agent a_i considers a_{i+1} more interesting than a_{i-1}, there is $\mu_i \in \mathcal{M}(a_i, a_{i-1})$ such that $\Phi(a_i, \mu_i) < \Phi(a_i, \mu)$ for all $\mu \in \mathcal{M}(a_{i+1}, a_i)$. Hence, as $\mu_{i+1}(a_{i+1}) = a_i$, $\Phi(a_i, \mu_i) < \Phi(a_i, \mu_{i+1}) = \Phi(a_{i+1}, \mu_{i+1})$. It follows that $\Phi(a_1, \mu_1) < \Phi(a_k, \mu_k) < \Phi(a_k, \mu) = \Phi(a_1, \mu), \forall \mu \in \mathcal{M}(a_1, a_k)$, which contradicts the fact that $\mu_1 \in \mathcal{M}(a_1, a_k)$. □

Theorem 3(i) extends Chung (2000, Corollary 2) to an environment with externalities. Theorem 3(ii) guarantees that a stable matching exists when two agents matched under some $\mu \in \mathcal{M}$ give the same utility to it. This hypothesis can be viewed as an extension to the framework with externalities of the *symmetric utilities hypothesis* introduced by Rodrigues-Neto (2007).

3. The Effects of Externalities on Roommate Problems

In this section, we show through several examples that a variety of properties of the classical roommate problem are lost when externalities are included, even when individuals have prudent expectations about the reactions of other agents to deviations.

For instance, without external effects, the existence of a stable matching is incompatible with the existence of a stable partition with odd rings (see Tan and Hsueh (1995, Theorem 3.10 + Corollary 3.8) and Tan (1991, Theorem 3.3)). However, as the following example points out, stable matchings and stable partitions with odd rings may coexist in our framework.

Example 2. Following the notation of Example 1, let $(N, (\succ_a)_{a\in N})$ be a roommate problem with three agents, $N = \{1, 2, 3\}$, with preferences characterized by

$$\mu_3 \succ_1 \mu_2 \succ_1 \mu_0 \succ_1 \mu_1, \qquad \mu_1 \succ_2 \mu_0 \succ_2 \mu_3 \succ_2 \mu_2, \qquad \mu_0 \succ_3 \mu_2 \succ_3 \mu_1 \succ_3 \mu_3.$$

In this context, μ_0 is the only stable matching, because $(1, 3)$ blocks μ_1, $(1, 2)$ blocks μ_2, and $(2, 3)$ blocks μ_3. Also, Definition 2 guarantees that $\{\{1\}, \{2\}, \{3\}\}$ and $\{\{1, 2, 3\}\}$ are stable partitions. □

In the proof of Theorem 1 we have shown that $(N, (\succ_a)_{a\in N})$ has a stable partition by considering a roommate problem without externalities $(N, (\succ^*)_{a\in N})$ that has a partition satisfying the properties of stability required by Tan (1991) and Tan and Hsueh (1995). The Example 2 proves that neither the sets of stable partitions nor the sets of stable matchings of these problems coincide. Indeed, $\{\{1\}, \{2\}, \{3\}\}$ is unstable in the roommate problem without externalities induced by the problem of Example 2, which in turn does not have stable matchings.[7]

Example 2 also shows that the assumptions on preference profiles stated in Theorem 3 are not necessary to guarantee the existence of an efficient and stable matching. On the one hand, the preference profile described in Example 2 has odd rings. On the other hand, if Φ is a representation of $(\succ_a)_{a\in N}$ satisfying the condition (ii) of Theorem 3, then $\Phi(2, \mu_1) = \Phi(3, \mu_1) < \Phi(3, \mu_2) = \Phi(1, \mu_2) < \Phi(1, \mu_3) = \Phi(2, \mu_3) < \Phi(2, \mu_1)$, which is a contradiction.

[7] *On stable algorithms.* It is not difficult to verify that a stable matching of the problem without externalities $(N, (\succ^*)_{a\in N})$ is also stable in $(N, (\succ_a)_{a\in N})$. Hence, when $(N, (\succ^*)_{a\in N})$ is solvable, the classical algorithm proposed by Irving (1985) can be applied to it in order to find a stable matching for $(N, (\succ_a)_{a\in N})$. Alternatively, if $(N, (\succ^*)_{a\in N})$ has no stable matching, we can apply any of the algorithms proposed by Tan (1991) and Tan and Hsueh (1995) in order to find a stable partition, which is also stable for $(N, (\succ)_{a\in N})$ (see the proof of Theorem 1). However, the stable partitions of $(N, (\succ^*)_{a\in N})$ do not necessarily induce stable matchings of $(N, (\succ)_{a\in N})$. For instance, in Example 2 the matchings induced by the stable partitions of $(N, (\succ^*)_{a\in N})$ are $\{\mu_1, \mu_2, \mu_3\}$, while μ_0 is the only stable matching of $(N, (\succ)_{a\in N})$.

REMARK 1 (THE EXISTENCE OF PATHS TO STABILITY)

In the class of roommate problems with a non-empty set of stable matchings it is natural to analyze the existence of *paths to stability*. Thta is, to determine if a stable outcome can be reached from any unstable matching through a finite sequence of blockings. This issue was extensively discussed in the previous literature without externalities and positive results were shown (see, Roth and Vande Vate (1990), Chung (2000), and Diamantoudi, Miyagawa, and Xue (2004)). Unfortunately, the Example 2 shows that the presence of externalities compromises the existence of paths to stability. Indeed, in the roommate problem described in that example any unstable matching has only one pair of agents that wants to block it: $(1,3)$ wants to block μ_1, $(1,2)$ wants to block μ_2, and $(2,3)$ wants to block μ_3. Therefore, from any unstable matching, a sequence of blocking pairs always enters into the cycle $\mu_1 \to \mu_2 \to \mu_3 \to \mu_1 \to \cdots$, which never leads to the unique stable matching μ_0. □

Given a matching μ, P is a *partition induced by* μ if $P = \{\{a,b\} : \mu(a) = b\} \cup \{\{c\} : \mu(c) = c\}$. The following example shows that in the presence of externalities a stable matching does not necessarily induce a stable partition (see Theorem 2(i)-(ii)).

EXAMPLE 3. Let $(N, (\succ_a)_{a\in N})$ be a roommate problem in which $N = \{1,2,3,4\}$ and

$$\begin{aligned}
&((1,2),(3,4)) \succ_1 ((1,2),3,4) \succ_1 ((1,3),2,4) \succ_1 ((1,3),(2,4)) \succ_1 \cdots \succ_1 \mu_0,\\
&((1,2),(3,4)) \succ_2 ((2,3),1,4) \succ_2 ((1,4),(2,3)) \succ_2 ((1,2),3,4) \succ_2 \cdots \succ_2 \mu_0,\\
&((1,2),(3,4)) \succ_3 ((1,3),2,4) \succ_3 ((1,3),(2,4)) \succ_3 ((2,3),1,4) \succ_3 ((1,4),(2,3)) \succ_3 \cdots \succ_3 \mu_0,\\
&((1,2),(3,4)) \succ_4 \cdots\cdots \succ_4 (1,2,(3,4)).
\end{aligned}$$

We affirm that $((1,2),(3,4))$ is the only stable matching and, in spite of what happens for classical roommate problems, the partition $\{\{1,2\},\{3,4\}\}$ is unstable. Indeed, $((1,2),(3,4))$ is stable since it is top ranked for all agents, while the other matchings are unstable because the following properties hold: μ_0 is blocked by $(1,2)$, $((1,4),2,3)$ and $((1,4),(2,3))$ are blocked by $(1,3)$, $((2,4),1,3)$ and $((2,4),(1,3))$ are blocked by $(1,2)$, $((3,4),1,2)$ is blocked by $(1,2)$, $((2,3),1,4)$ is blocked by $(1,3)$, $((1,3),2,4)$ is blocked by $(1,2)$, and $((1,2),3,4)$ is blocked by $(2,3)$. The partition $\{\{1,2\},\{3,4\}\}$ is unstable because $\{3,4\}$ are not mutually acceptable agents.

Furthermore, it is not difficult to verify that $\{\{1,2,3\},\{4\}\}$ is a stable partition. Thus, when externalities are allowed, complete stable matchings and stable partitions with odd rings may coexist. □

The next example shows another interesting effect of the presence of externalities in models with an even number of agents: complete and incomplete stable matchings may coexist.

EXAMPLE 4. Let $(N, (\succ_a)_{a\in N})$ be a roommate problem in which $N = \{1,2,3,4\}$ and

$$\begin{aligned}
&((1,2),(3,4)) \succ_1 \cdots\cdots \succ_1 (1,2,3,4) \succ_1 ((1,2),3,4) \succ_1 ((1,4),2,3),\\
&((1,2),(3,4)) \succ_2 \cdots\cdots \succ_2 (1,2,3,4) \succ_2 ((2,3),1,4),\\
&((1,2),(3,4)) \succ_3 \cdots\cdots \succ_3 (1,2,3,4) \succ_3 ((1,3),2,4) \succ_3 (1,2,(3,4)),\\
&((1,2),(3,4)) \succ_4 \cdots\cdots \succ_4 (1,2,3,4) \succ_4 ((2,4),1,3).
\end{aligned}$$

It follows that $((1,2),(3,4))$ is stable, although $\{\{1,2\},\{3,4\}\}$ is not a stable partition because 2 is not acceptable to 1. Notice that agents 1 and 3 have only one acceptable partner and, therefore, $(\succ_a)_{a\in N}$ has no odd rings. Moreover, since there are no pairs of mutually acceptable agents, $\{\{1\},\{2\},\{3\},\{4\}\}$ is the only stable partition and Theorem 2(i) guarantees that $(1,2,3,4)$ is a stable matching.

If we modify the example exchanging the position of matchings $((1,2),(3,4))$ and $(1,2,(3,4))$ in the preference profile above, we guarantee that $(1,2,(3,4))$ is an *incomplete* stable matching that cannot be induced by a stable partition. □

Remark 2 (The Lone Wolf Theorem)

In the absence of externalities, the so-called *Lone Wolf Theorem* holds: an agent that is alone in a stable matching will remain alone in any other stable matching (see Gusfield and Irving (1989), Klaus and Klijn (2010)). This property does not hold in the presence of externalities, as in Example 4 we have that $(1,2,3,4)$ and $((1,2),(3,4))$ are stable matchings. Furthermore, Tan and Hsueh (1995) showed that without externalities all stable partitions have the same odd rings and singletons. This property is not fulfilled in our framework, as $\{\{1\},\{2\},\{3\}\}$ and $\{\{1,2,3\}\}$ are stable partitions in Example 2. □

The next example shows that there are solvable roommate problems with externalities that do not have a complete stable matching although the number of agents is even.

Example 5. Let $(N,(\succ_a)_{a\in N})$ such that $N=\{1,2,3,4\}$ and

$$\begin{aligned}
&((1,2),3,4) \succ_1 \cdots\cdots \succ_1 ((1,2),(3,4)) \succ_1 ((1,4),2,3),\\
&((2,3),1,4) \succ_2 \cdots\cdots \succ_2 ((2,3),(1,4)) \succ_2 ((2,4),1,3),\\
&((1,3),2,4) \succ_3 \cdots\cdots \succ_3 ((1,3),(2,4)) \succ_3 ((3,4),1,2).
\end{aligned}$$

Since $(\succ_a)_{a\in N}$ has no odd rings and there are no pairs of agents mutually acceptable, it follows from Theorem 1 that, independently of 4's preferences, $\{\{1\},\{2\},\{3\},\{4\}\}$ is the only stable partition. Hence, $(1,2,3,4)$ is a stable matching. Also, all complete matchings are unstable. □

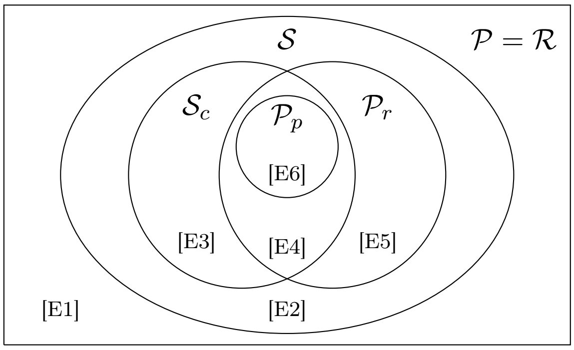
In previous examples all stable partitions have either a ring or a singleton. The next one shows that the set of roommate problems that have a stable partition without odd rings or singletons is non-empty.

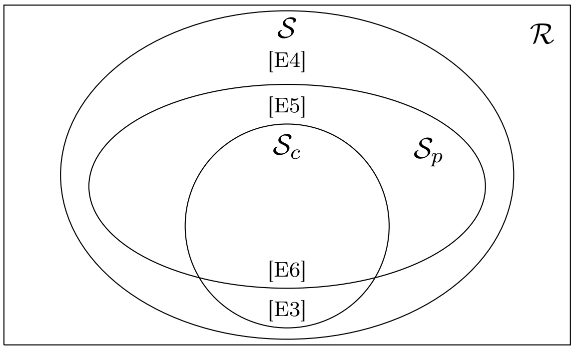
Example 6. Let $(N,(\succ_a)_{a\in N})$ be a roommate problem in which $N=\{1,2,3,4\}$ and

$$\begin{aligned}
&((1,4),(2,3)) \succ_1 ((1,3),2,4) \succ_1 ((1,3),(2,4)) \succ_1 ((1,2),(3,4)) \succ_1 \cdots\cdots \succ_1 ((1,4),2,3) \succ_1 \mu_0,\\
&((1,2),(3,4)) \succ_2 ((2,4),1,3) \succ_2 ((2,4),(1,3)) \succ_2 ((1,4),(2,3)) \succ_2 \cdots\cdots \succ_2 ((1,2),3,4) \succ_2 \mu_0,\\
&((1,4),(2,3)) \succ_3 ((1,3),2,4) \succ_3 ((1,3),(2,4)) \succ_3 ((1,2),(3,4)) \succ_3 \cdots\cdots \succ_3 ((2,3),1,4) \succ_3 \mu_0,\\
&((1,2),(3,4)) \succ_4 ((2,4),1,3) \succ_4 ((2,4),(1,3)) \succ_4 ((1,4),(2,3)) \succ_4 \cdots\cdots \succ_4 ((3,4),1,2) \succ_4 \mu_0
\end{aligned}$$

Since $((1,3),(2,4))$ is a stable matching, it follows from Definition 2 that $\{\{1,3\},\{2,4\}\}$ is a stable partition. Notice that, $(N,(\succ_a)_{a\in N})$ does not have odd rings and there are no stable matching with two or more agents alone. Hence, $(N,(\succ_a)_{a\in N})$ has no stable partitions with odd rings or singletons. □

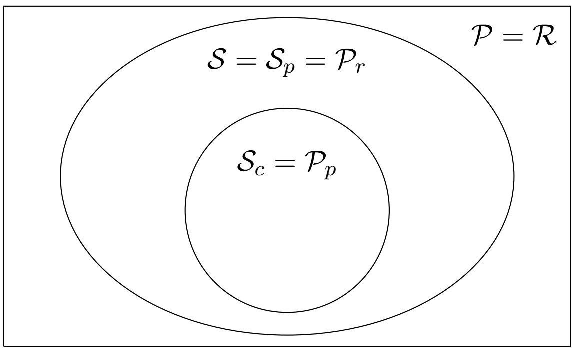
The following Venn diagrams summarize some of our results:





$\mathcal{R}$	: Roommate problems with externalities $(N, (\succ_a)_{a \in N})$.
$\mathcal{P}$	: Problems in $\mathcal{R}$ with a stable partition.
$\mathcal{S}$	: Problems in $\mathcal{R}$ with a stable matching.
$\mathcal{S}_c$	: Problems in $\mathcal{R}$ with a complete stable matching.
$\mathcal{P}_r$	: Problems in $\mathcal{R}$ where stable partitions do not have odd rings.
$\mathcal{P}_p$	: Problems in $\mathcal{R}$ with a stable partition without odd rings or singletons.
$\mathcal{S}_p$	: Problems in $\mathcal{S}$ where all stable matchings are induced by stable partitions.
[En]	: Example n.

Note that, abusing the notation slightly, in models *without externalities* we have that:



Remark 3 (The Core with Externalities)

Given a roommate problem with externalities $(N,(\succ_a)_{a\in N})$, a matching μ *can be blocked by a coalition* $S\subseteq N$ if there exists $f:S\to S$ such that $f(f(S))=S$ and $\rho\succ_a\mu$ for all $a\in S$ and $\rho\in\mathcal{M}$ such that $\rho(b)=f(b)$, $\forall b\in S$. The *core* of $(N,(\succ_a)_{a\in N})$ is the set of matchings that cannot be blocked by any coalition. Notice that, the core is contained in the set of efficient and stable matchings.

In spite of what happens in roommate problems without externalities (cf. Alcalde (1994)), in our framework the core may be an empty set even when there are efficient and stable matchings. For instance, following the notation of Example 1, consider a problem where agents are characterized by preferences $\mu_2\succ_1\mu_0\succ_1\mu_1\succ_1\mu_3$, $\mu_0\succ_2\mu_1\succ_2\mu_2\succ_2\mu_3$, and $\mu_1\succ_3\mu_2\succ_3\mu_0\succ_3\mu_3$. In this context, μ_1 is the unique (efficient) stable matching.[8] However, it is blocked by the coalition $\{1,2\}$ through $f:\{1,2\}\to\{1,2\}$ such that $(f(1),f(2))=(1,2)$. Thus, the core is empty. □

4. Insertion in the Literature

We will show that well-known characterizations of stability for marriage markets with externalities and for classical roommate problems can be recovered as particular cases of our previous results.

Given non-empty, finite, and disjoint sets M_1 and M_2, the roommate problem with externalities $(M_1\cup M_2,(\succ_a)_{a\in M_1\cup M_2})$ is an *instance of the marriage market* when it is a two-sided problem in which agents in one side are only interested in forming pairs with agents on the other side. That is, for each $i\in\{1,2\}$ and $a,b\in M_i$, with $a\neq b$, we have that $\mu\succ_a\eta$ for all $(\mu,\eta)\in\mathcal{M}(a,a)\times\mathcal{M}(a,b)$. A problem satisfying the conditions above is an *instance of the strict marriage market* when $\#M_1=\#M_2$ and all agents prefer to form a pair than being alone. Hence, for each $i\in\{1,2\}$ and $(a,b)\in M_i\times M_j$, with $i\neq j$, we have that $\mu\succ_a\eta$ for all $(\mu,\eta)\in\mathcal{M}(a,b)\times\mathcal{M}(a,a)$.

Corollary (Sasaki and Toda (1996), Mumcu and Saglam (2010))

Any problem $(M_1\cup M_2,(\succ_a)_{a\in M_1\cup M_2})$ that is an instance of the (strict) marriage market has an efficient and (complete) stable matching.

Proof. Let $(M_1\cup M_2,(\succ_a)_{a\in M_1\cup M_2})$ be an instance of the marriage market. It follows from Theorem 3 that it is sufficient to prove that $(\succ_a)_{a\in M_1\cup M_2}$ has no odd rings. Let $\{a_1,\dots,a_{2k+1}\}$ be an odd ring and assume, without loss of generality, that $a_1\in M_1$. Since for every $i\in I_{2k+1}$, a_{i+1} and a_{i-1} are acceptable to a_i, we conclude that $\{a_2,a_4,\dots,a_{2k},a_1\}\subseteq M_2$, which is a contradiction.

Suppose that $(M_1\cup M_2,(\succ_a)_{a\in M_1\cup M_2})$ is an instance of the strict marriage market and let μ be a stable matching, which exists as a consequence of the previous case. If under $\mu\in\mathcal{M}$ there is an agent alone, then there is an even number of agents alone and, therefore, any pair of them can block the matching. □

Externalities are neutral in $(N,(\succ_a)_{a\in N})$ when, for each $a,b,c\in N$ and $(\tilde\mu,\tilde\eta)\in\mathcal{M}(a,b)\times\mathcal{M}(a,c)$, we have that $\tilde\mu\succ_a\tilde\eta\implies\mu\succ_a\eta$, $\forall(\mu,\eta)\in\mathcal{M}(a,b)\times\mathcal{M}(a,c)$. That is, if externalities are neutral, an agent never considers beliefs about other individuals' decisions to form a pair.

[8] Indeed, μ_2 is blocked by $(2,3)$, while μ_3 and μ_0 are blocked by $(1,3)$.

COROLLARY (TAN (1991), TAN AND HSUEH (1995))
For any roommate problem in which externalities are neutral we have that:
(i) There is a stable partition.
(ii) There is a stable matching if and only if there is a stable partition without odd rings.
(iii) There is a complete stable matching iff there is a stable partition without odd rings or singletons.

PROOF. Item (i) follows from Theorem 1. Properties (ii) and (iii) are a consequence of Theorem 2 and the fact that, when externalities are neutral, any stable matching (respectively, any complete stable matching) induces a stable partition without odd rings (respectively, without odd rings or singletons). Indeed, the neutrality of externalities guarantees that two agents forming a pair in a stable matching are mutually acceptable and, therefore, the requirements of Definition 2 hold. □

5. CONCLUDING REMARKS

In this work, we extended the classical roommate problem to include externalities in such form that individuals have strict preferences over the set of matchings. Assuming that agents have prudent expectations about other agents' reactions to their deviations, we proposed stability concepts for partitions of the set of agents and for matchings. Since previous results of the literature without externalities were obtained as particular cases of our main findings, our concepts of stability can be viewed as natural extensions of those that were studied in models without external effects.

We have shown that in the presence of externalities a stable partition always exists and a stable matching can be induced by a stable partition without odd rings. We also provided extensions to a framework with externalities of the no-odd-rings condition of Chung (2000) and the symmetric utilities hypothesis of Rodrigues-Neto (2007).

However, as the following table summarizes, some properties were lost by including externalities:

	Externalities	*Classical*
A stable partition always exists	✓	✓
A stable partition without odd rings induces a stable matching	✓	✓
Stable matchings and stable partitions with odd rings never coexist	×	✓
All stable partitions have the same (number of) odd rings and singletons	×	✓
There is a path to stability from any matching	×	✓
The core coincides with the set of stable matchings	×	✓
The Lone Wolf Theorem holds	×	✓

Therefore, as a matter of future research, it is interesting to discuss if the inclusion of restrictions on the space of individuals' preferences, or the sophistication of individuals' expectations about other agents deviations, or the focus on particular types of externalities, allow us to recover some of the properties described above (cf., Bando (2012), Hafalir (2008), Fisher and Hafalir (2016)).

It is important to remark that there are solution concepts that make all classical roommate problems solvable and guarantee that any stable matching, when it exists, belongs to the solution set. For instance, the *P-stable matchings* defined by Iñarra, Larrea, and Molis (2008), the *absorbing sets* proposed by

Iñarra, Larrea, and Molis (2013), and the *Q-stable matchings* introduced by Biró, Iñarra, and Molis (2016). Unfortunately, if stable matchings are still required to be part of the outcomes determined by a solution concept, then these previous approaches of the literature cannot be adapted to our framework. Indeed, our Examples 2 and 3 show that the natural extensions to the framework with externalities of the definitions of *P-stable matchings*, *absorbing sets*, and *Q-stable matchings* are not compatible with the inclusion of stable outcomes.[9]

REFERENCES

[1] Alcalde, J. (1994): "Exchange-Proofness or Divorce-Proofness? Stability in one-sided matching markets," *Economic Design,* 1, 275-287.

[2] Bando, K. (2012): "Many-to-one matching markets with externalities among firms," *Journal of Mathematical Economics,* 48, 14-20.

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

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

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

[6] Fisher, J. and I. Hafalir (2016): "Matching with aggregate externalities," *Mathematical Social Sciences,* 81, 1-7.

[7] Gale, D., and L. Shapley (1962): "Collegue admisions and the stability of marriage," *The American Mathematical Monthly,* 69, 9-15.

[8] Gusfield, D. and R. Irving (1989): "The stable marriage problem: structure and algorithms," The MIT Press, Cambridge.

[9] Hafalir, I. (2008): "Stability of marriage with externalities," *International Journal of Game Theory,* 37, 353-369.

[10] Iñarra, E., C. Larrea, and E. Molis (2008): "Random paths to P-stability in the roommate problem," *International Journal of Game Theory,* 36, 461-471.

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

[12] Irving, R. (1985): "An efficient algorithm for the stable roommates problem," *Journal of Algorithms,* 6, 577-595.

[13] Klaus, B. and F. Klijn (2010): "Smith and Rawls share a room: stability and medians," *Social Choice and Welfare,* 25, 647-667.

[14] Mumcu, A., and I. Saglam (2010): "Stable one-to-one matching with externalities," *Mathematical Social Sciences,* 60, 154-159.

[9] By analogy with the definition introduced by Iñarra, Larrea, and Molis (2008), the set of *P-stable* matchings is given by $\mathcal{M}(P)$. Thus, although $((1,2),(3,4))$ is the only stable matching in Example 3, it is not a P-stable matching.

Inspired by Iñarra, Larrea, and Molis (2013, Definition 2), we can define an *absorbing set* $\mathcal{A} \subseteq \mathcal{M}$ as a non-empty set that satisfies the following conditions: (i) for any $\mu, \mu' \in \mathcal{A}$, μ can be reached from μ' through a finite sequence of blockings; and (ii) departing from any $\mu \in \mathcal{A}$ we cannot reach, through a finite sequence of blockings, a matching that not belongs to $\mathcal{A}$. Hence, in the context of Example 2, $\{\mu_1, \mu_2, \mu_3\}$ is an absorbing set that not include stable matchings. Notice that, without externalities, if a roommate problem has a non-empty set of stable matchings, then $\mathcal{A}$ is an absorbing set if and only if $\mathcal{A} = \{\mu\}$ for some stable matching μ (see Iñarra, Larrea, and Molis (2013, Proposition 2)).

Using our blocking concept, we can extend to our framework the definitions of *maximum internally stable matching* and *maximum irreversible matching* that were introduced by Biró, Iñarra, and Molis (2016, Sections 3.1 and 3.2) in the context of classical roommate problems. Hence, defining a *Q-stable matching* as a matching that is both maximum internally stable and maximum irreversible, it is not difficult to verify that in the context of Example 2 the set of Q-stable matchings is given by $\{\mu_1, \mu_2, \mu_3\}$ and, therefore, it does not include the stable outcome μ_0.

[15] Pęski, M. (2017): "Large roommate problem with non-transferable random utility," *Journal of Economic Theory,* 168, 432-471.

[16] Rodrigues-Neto, J. (2007): "Representing roommates' preferences with symmetric utilities," *Journal of Economic Theory,* 135, 545-550.

[17] Sasaki, H., and M. Toda (1996): "Two-sided matching problems with externalities," *Journal of Economic Theory,* 70, 93-108.

[18] Roth, A., and J. Vande Vate (1990): "Random paths to stability in two-sided matching," *Econometrica,* 58, 1475-1480.

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

[20] Tan, J., and Y. Hsueh (1995): "A generalization of the stable matching problem," *Discrete Applied Mathematics,* 59, 87-102.