Constrained-Optimal Tradewise-Stable Outcomes in the One-Sided Assignment Game: A Solution Concept Weaker than the Core

David Pérez-Castrillo
Marilda Sotomayor

This version: May 2020
(April 2019)

Barcelona GSE Working Paper Series

Working Paper nº 1094

Constrained-optimal tradewise-stable outcomes in the one-sided assignment game: A solution concept weaker than the core*

David Pérez-Castrillo† and Marilda Sotomayor‡

May 2020

Abstract

In the one-sided assignment game any two agents can form a partnership and decide how to share the surplus created. Thus, in this market, an outcome involves a matching and a vector of payoffs. Contrary to the two-sided assignment game, stable outcomes often fail to exist in the one-sided assignment game. We introduce the idea of tradewise-stable (t-stable) outcomes: they are individually rational outcomes where no matched agent can form a blocking pair with any other agent, neither matched nor unmatched. We propose the set of constrained-optimal t-stable outcomes, which is the set of the maximal elements of the set of t-stable outcomes, as a natural solution concept for this game. We prove several properties of t-stable outcomes and constrained-optimal

*We thank the very constructive comments received at SING 15 in Turku, the 30th International Conference on Game Theory at Stony Brook, the ASSET meeting in Athens, as well as at seminars at Aix-Marseille School of Economics, Paris School of Economics, and Shanghai University of Finance and Economics. Marilda Sotomayor acknowledges financial support from CNPq-Brazil. David Pérez-Castrillo acknowledges financial support from the Ministerio de Ciencia, Innovación y Universidades and Feder (PGC2018-094348-B-I00), Generalitat de Catalunya (2017SGR-711), ICREA under the ICREA Academia programme, and Severo Ochoa Programme (SEV-2015-0563). A previous version of this paper has circulated under the title "Conflict-free and Pareto-optimal allocations in the one-sided assignment game: A solution concept weaker than the core."

†Corresponding author. Universitat Autònoma de Barcelona and Barcelona GSE; Dept. Economía e Hist. Económica; Edificio B; 08193 Bellaterra - Barcelona; Spain. Email: david.perez@uab.es

‡Graduate School of Economics, Getulio Vargas Foundation-RJ; Praia de Botafogo,190, 22250-900, Rio de Janeiro, RJ, Brazil; and Universidade de São Paulo-SP; Dep de Economia; Av. Prof. Luciano Gualberto, 908; Cidade Universitária, 5508-900, São Paulo, SP. Email: marildas@usp.br

t-stable outcomes. In particular, we show that each element in the set of constrained-optimal t-stable payoffs provides the maximum surplus out of the set of t-stable payoffs, the set is always non-empty and it coincides with the core when the core is non-empty. The general principle of collective rationality on which our theory is based presupposes that a player only engages in cooperation that is optimal for him/her. That is, whatever the dynamics that underlie the pairwise interactions, any negotiation process in this environment should always arrive to an outcome where every trade is optimal (and stable) for the players involved.

JEL Classification numbers: C78, D78

Keywords: matching, assignment game, stability, core, tradewise-stable, constrained-optimal.

1 Introduction

Interactions among people, firms, and many other agents, often take place in terms of two-agent partnerships. A seller and a buyer meet to realize a transaction that is profitable for both; a firm and a worker sign a contract that benefits both; two firms establish an R&D collaboration agreement; or two roommates agree to share the cost of an apartment. Some of these partnerships take place between pairs of agents from two clearly distinct populations: there is a set of buyers and a set of sellers, as there is a set of firms and a set of workers. A buyer, for instance, is either matched with a seller or he/she does not buy, but he/she is not interested in forming a partnership with another buyer. In other environments, pairs are made between agents who all belong to the same set: a set of innovative firms or a set of tenants. A firm, for instance, may be matched to any other firm or it can do R&D on its own.

Two-sided matching models, pioneered by Gale and Shapley (1962) (the marriage and the college admission models) and Shapley and Shubik (1972) (the assignment game), provide an excellent framework to study pairwise interactions in environments where the players belong to two disjoint sets and interact by pairs. Moreover, the set of stable outcomes for the one-to-one two-sided matching models has a very nice algebraic structure and appealing properties. In particular, this set, the set of pairwise-stable outcomes and the core coincide and they are always non-empty.

In the present paper, we relax the assumption of two sides in the assignment game of Shapley and Shubik and study environments where the agents belong to a single population, and not necessarily to two distinct sets. It is assumed that the players can freely communicate with each other and preferences over outcomes, as well as the rules that govern any coalitional interaction, are common knowledge. The main activity of these agents is to form pairs, whose members endogenously decide on the sharing of the surplus created by the partnership. Every agent can form one partnership at most. Thus, an outcome of our model involves both a one-to-one matching (that is, a partition of the population in either pairs of agents or singletons) and a vector of payoffs (that is, a sharing of the joint surplus for any two-agent partnership). We will refer to this game as the "one-sided assignment game."[1] As in the one-to-one two-sided matching models, *the core, the set of stable payoffs and the*

[1] It is called "the partnership formation problem" in Talman and Yang (2011) and Andersson et al. (2014), "the TU roommate game" in Eriksson and Karlander (2000), and simply "the roommate problem" in Chiappori et al. (2014).

set of pairwise-stable payoffs coincide in the one-sided assignment game (Proposition 1). Furthermore, *any optimal matching*[2] *is compatible with any stable payoff and every matching in a stable outcome is optimal* (Sotomayor, 2005a and 2009a, and Talman and Yang, 2011). Consequently, *every unmatched agent at a stable outcome has a zero payoff at every stable outcome*. However, in the one-sided assignment game the core may be empty (Example 1).[3]

We provide a solution concept for the one-sided assignment that is always non-empty. The general principle of collective rationality on which our theory is based presupposes that *a player only engages in cooperation that is optimal for him/her.* That is, if an agreement between the members of a partnership is reached, then these agents should be sure that more favorable options cannot be obtained elsewhere. Therefore, whatever the dynamics that underlie the pairwise interactions, any negotiation process in this environment should always arrive to an outcome where every trade is optimal (and stable) for the players involved. Any instability (if any) of this outcome should be caused by non-trading agents. We call such an outcome a *tradewise-stable* (*t-stable*) *outcome.* An outcome (that is, a matching and a vector of payoffs) is t-stable if no matched agent can form a blocking pair with any other agent, neither matched nor unmatched. In particular, in a t-stable outcome the set of active players, that is, those players who are matched, is a stable set.

Furthermore, our solution concept requires that the number of trades accomplished is the maximum number that can occur at a t-stable outcome. Thus, it should not be possible to increase the set of active players by adding new trades without violating the optimal behavior principle. We can say that such outcomes are *as stable as possible.* We call them *constrained-optimal tradewise-stable outcomes* (*optimal t-stable outcomes* for short) because they are Pareto optimal among all t-stable outcomes. Clearly, any core outcome is an optimal t-stable outcome. However, there are t-stable outcomes that are not core outcomes because they are blocked by unmatched players.

Within this context, the game theoretic prediction is that an optimal t-stable outcome will occur. If, given such an outcome, no additional interaction can benefit the agents involved, a core outcome is reached, so the traditional game theoretic prediction is maintained. If there are some profitable interactions, but they require that some of the agents involved do not behave optimally, the only conclusion we can arrive is that the new outcome is not an

[2] A matching is optimal if it maximizes the total payoff in the set of feasible matchings.

[3] Gale and Shapley (1962) show that stable matchings may also not exist in the one-sided discrete model, that is, when utility is not transferable. The existence problem for that model was called by these authors "the roommate problem."

equilibrium.

Thus, we provide a solution concept, the set of optimal t-stable outcomes, which is weaker than the core concept and that can be used as a vehicle to identify the outcomes that may occur in this decentralized setting. We will show that the set of active players of an optimal t-stable outcome is endowed with at least one coalitional structure. Such a coalitional structure indicates some possible ordering in the formation of the active coalitions of this outcome. The history that underlies the formation of an outcome that is as stable as possible can be presumed from that ordering.[4]

More specifically, given an optimal t-stable outcome σ, there is a coalitional structure of the set of active players $T(\sigma)$, which constitutes the stable part of σ, given by some partition $(C_1, \ldots, C_k)$ of $T(\sigma)$, where every partition set C_h, with $h = 1, \ldots, k$, has the following property: $C_1 \cup \ldots \cup C_h$ is a stable set and if an agent belongs to C_h, then his/her payoff is greater than or equal to his/her gain from trade with any agent not in $C_1 \cup \ldots \cup C_h$. Furthermore, no proper subset of C_h has this property. From that structure (it may exist more than one) we can presume that the allocation σ is formed along k steps of some negotiation process, operating sequentially. In this *partnership formation process*, which starts with the outcome where every player stands alone, the set C_h is formed at step h, for all $h = 1, \ldots, k$, and it results from the pairwise interactions of current non-trading agents, under the premise of optimal cooperative behavior. As a consequence of this principle, the transactions done at each step h are maintained at the subsequent steps, if any, and the non-trading agents at this stage have no willingness to trade with any subset of players who became active at this or any previous step. Therefore, the stable part of the outcome σ_h, which occurs at step h, is $C_1 \cup \ldots \cup C_h$ and the remaining players stand alone at this step. The optimal t-stable allocation σ occurs when the last term of the sequence $(C_1, \ldots, C_k)$ is formed. Thus, we obtain a sequence of outcomes, $(\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_k = \sigma)$, which are t-stable and where every term σ_h is an extension of the previous terms.

Tradewise-stable outcomes share properties of core outcomes but not necessarily all of them. More specifically, t-stable outcomes are somehow "internally stable." However, they might not be "externality stable," in the sense that there may be a pair of agents not involved in any partnership that could have an incentive to deviate. As it is clear from the

[4] The consideration of the dynamics that may underlie the coalitional interactions among individuals would be a nice exercise. However, the modelling of these dynamics is mathematically untreatable. The existence of a coalition structure for the active coalitions makes the precise model for the dynamics less necessary.

definition, all core outcomes are t-stable outcomes. But the set of t-stable outcomes is always non-empty, since the allocation where everyone is non-trading and receives a zero payoff is t-stable.

By its definition, a t-stable allocation is constrained-optimal if and only if it cannot be extended to another t-stable allocation. Indeed, we prove that a t-stable allocation is as stable as possible if and only it is not weakly dominated, via the grand coalition, by any other t-stable allocation. An optimal t-stable payoff is identified with a maximal element of the set of t-stable payoffs (which is proved to be a non-empty compact set), under the partial order relation induced from that of the Euclidean space where this set is immersed. Then, optimal t-stable payoffs always exist. Hence, independently of the emptiness or the non-emptiness of the core, *we can always predict that only optimal t-stable allocations will occur and any optimal t-stable allocation may occur.*

We prove that the set of active players of every t-stable allocation, not necessarily optimal, also has a coalitional structure as described above. Then, it is intuitive that the allocation may be reached at some step of a negotiation process, a process that always culminates with an optimal tradewise allocation. This intuition is confirmed by Proposition 6 that asserts that *every t-stable allocation can be extended to an optimal t-stable allocation.* In particular, *when the core exists, every t-stable allocation can be extended to a core allocation.*

Additionally, we prove that the set of optimal t-stable outcomes always has a structure similar to that of the set of core outcomes. Indeed, *every optimal t-stable outcome provides the maximum total surplus among all t-stable outcomes and no other t-stable outcome can achieve this level of total surplus.* Thus, the matchings that are compatible with optimal t-stable outcomes are "quasi-optimal." Furthermore, they are part of an optimal matching. On the other hand, *the only t-stable allocations compatible with an optimal matching are the corewise-stable payoffs.*

Moreover, as the corewise-stable outcomes, *each quasi-optimal matching is compatible with any optimal t-stable payoff.* Therefore, the set of optimal t-stable outcomes is the Cartesian product of the set of quasi-optimal matchings and the set of optimal t-stable payoffs. Consequently, *if an agent is unmatched at some optimal t-stable allocation, then he gets a zero payoff at every optimal t-stable allocation.*

Since an optimal t-stable allocation is as stable as possible then, by the definition of such allocation, it has the same set of blocking pairs. Therefore, *every blocking pair of an optimal t-stable allocation is unmatched at every optimal t-stable allocation.* As a consequence of the fact that every t-stable allocation is extended by some optimal t-stable allocation, it can be concluded that *every blocking pair of an optimal t-stable allocation is unmatched at every t-stable allocation.*

Our central result states that *the set of optimal t-stable allocations coincides with the core, when this set is non-empty* (Theorem 1). Moreover, *if an optimal t-stable allocation is not in the core, then no optimal t-stable allocation is in the core and this set is empty.* Therefore, if the negotiation process leads to an unstable outcome, which is as stable as possible, then the core is empty.

In sum, the main feature of our solution concept is that it proposes the core when this set is not empty and, when the core is empty, it recommends a set of payoffs as stable as possible that satisfies properties that are similar to the core.

Few papers have studied the one-sided assignment game. Necessary and sufficient conditions for the existence of the core using linear programming are obtained by Talman and Yang (2011). Erikson and Karlander (2000) use graph theory to provide a characterization of the core, and Klaus and Nichifor (2010) provide some properties of this set, when it is not empty. Chiappori et al. (2014) show that stable matchings exist when the economy is replicated an even number of times by "cloning" each individual. Finally, Andersson et al. (2014) propose a dynamic competitive adjustment process that either leads to a stable outcome or disproves the existence of stable outcomes.

The idea of the partnership formation process that leads to an optimal t-stable outcome has some similarities with the "paths to stability" proposed by Roth and Vande Vate (1990). They analyze the marriage market and show that, starting from an arbitrary matching, there always exists a process of myopic blocking pairs that leads to a stable matching. Klaus and Pagot (2015) show that the existence of such paths to stability is not guaranteed in the assignment game and provide a necessary and sufficient condition for the existence. In contrast to the paths to stability that have been considered in various contexts, each step in our partnership formation processes may involve more than one blocking pair. But every single step in the process produces an outcome with nice properties, as it is a t-stable outcome.

The concept of tradewise-stable outcome is an extension of the concept of "simple outcome," introduced in Sotomayor (1996) for the marriage model, and it is a translation of the concept of a "simple outcome," defined in Sotomayor (2005), for the housing market of Shapley and Scarf (1974), with strict or non-strict preferences. Both papers provide a non-constructive and very short proof, which only uses elementary combinatorial arguments, of the existence of the core. Still in environments without transfers, the notion of a simple outcome was used in Sotomayor (1999) for a discrete many-to-many matching model with substitutable and not-necessarily strict preferences, in Sotomayor (2004) where an implementation mechanism for the discrete many-to-many matching model is provided, in Sotomayor (2011) to characterize the set of Pareto-stable matchings in the marriage market and (if the set of stable matchings is not empty) in the discrete roommate model, and in Wu and Roth (2018) for the college admission model. An adaptation of the concept of simple matching was used in Sotomayor (2000) for a unified two-sided matching model, due to Eriksson and Karlander (2000), which includes the marriage and the assignment model, and in Sotomayor (2018) for the two-sided assignment game of Shapley and Shubik (1972). Finally, Sotomayor (2019) introduces the idea of t-stability for one of the sides of the market and provides a framework to treat conjointly stable and instable allocation structures.

The rest of the paper is organized as follows. Section 2 introduces the framework and states some preliminary results for the core. Section 3 introduces the t-stable outcomes and provides some of their properties. Section 4 analyzes the set of optimal t-stable outcomes and presents our results on this set. In section 5, we introduce and discuss a dynamique partnership formation process that ends with an optimal t-stable outcome. In section 6, we prove the links between the set of optimal t-stable outcomes and the set of corewise-stable outcomes. The final remarks are given in section 7. An Appendix includes several lemmas that will be used in the proofs of the results in the main text.

2 Framework and preliminaries

2.1 The framework

The description of the one-sided assignment game follows the one given in Roth and Sotomayor (1990) for the case with two sides, with the appropriate adaptations.

There is a finite set of players, $N = \{1, 2, ..., n\}$. Associated with each partnership $\{i, j\}$ there is a non-negative real number $a_{\{i,j\}}$ which will be denoted a_{ij}. The number a_{ij} represents the surplus that players i and j generate if they form a partnership.

We can represent the environment as a game in coalitional function form (N, v) with side payments determined by (N, a). In this game, the worth $v(i, j)$[5] of a two-player coalition

[5]For notational convenience, we write $v(i, j)$ rather than $v(\{i, j\})$.

$\{i, j\}$ is given by a_{ij}. We will define $v(i) \equiv a_{ii} \equiv 0$ for all $i \in N$. The worth of larger coalitions is entirely determined by the worth of the pairwise combinations that the coalition members can form. That is, $v(S) = max\{v(i_1, j_1) + v(i_2, j_2) + ... + v(i_k, j_k)\}$ for arbitrary coalitions S, where the maximum is taken over all sets $\{i_1, j_1\}, ..., \{i_k, j_k\}$ of two-player disjoint coalitions in S.[6]

Thus, the rules of the game are that any pair of agents $\{i, j\}$ can together obtain a_{ij}, and any larger coalition is valuable only insofar as it can organize itself into such pairs. The members of any coalition may divide their collective worth among themselves in any way they like.

We might think of the two-sided "Assignment Game" of Shapley and Shubik (1972) as a particular case of our model. In the assignment game, there are two disjoint sets P and Q and a pair of players can generate a surplus only if each belongs to a different set. Thus, our model corresponds to an assignment game when $N = P \cup Q$, $P \cap Q = \varnothing$, and $v(S) = 0$ if S contains only agents of P or only agents of Q.

We will represent the set of partnerships that are formed through a matching:

Definition 1 *A **feasible matching** x is a partition of N, where the partition sets are either pairs $\{i, j\}$ or singletons $\{i\}$. If $\{i, j\} \in x$ we can write $x(i) = j$ and we refer to $x(i)$ as the partner of i at x. If $\{i\} \in x$ we can write $x(i) = i$ and we say that i is **unmatched** at x.*

We will use the notation $\sum_A$ to denote the sum over all elements of A. Let x be a feasible matching. If $R \subseteq N$, we denote $\boldsymbol{x(R) \equiv \{j; x(i) = j \text{ for some } i \in R\}}$. If $x(R) = R$, we denote by $x|_R$ the partition of R where the partition sets belong to x. Therefore, $v(R) \geq \sum_{x|_R} a_{ij}$ for all feasible matchings x.

Definition 2 *The feasible matching x is **optimal** if, for all feasible matching x', $\sum_x a_{ij} \geq \sum_{x'} a_{ij}$.*

The set of optimal matchings is always non-empty, since there is a finite number of matchings. Under Definition 2 and since $v(N) \geq \sum_x a_{ij}$ for all feasible matchings x, it follows that the matching **x is optimal if and only if $\boldsymbol{\sum_x a_{ij} = v(N)}$**.

The players' benefit in the game will be represented by a vector of payoffs:

[6] k is an integer number that does not exceed the integer part of $|S|/2$.

Definition 3 *The vector u, with $u \in \mathbb{R}^n$, is called the* ***payoff****. The payoff u is* ***pairwise-feasible*** *for (N, a) if there is a feasible matching x such that*

$$u_i + u_j = a_{ij} \text{ if } x(i) = j \text{ and } u_i = 0 \text{ if } x(i) = i.$$

In this case, we say that (u, x) is a ***pairwise-feasible outcome*** *and x is compatible with u.*

Definition 4 *The payoff u is feasible for (N, a) if $\sum_N u_i \leq v(N)$.*

Remark 1 *Given a coalition R, the definition of v implies that there is some feasible matching x such that $x(R) = R$ and $\sum_{x|R} a_{ij} = v(R)$. Furthermore, $v(R) \geq \sum_{x'|R} a_{ij}$ for all feasible matchings x' such that $x'(R) = R$. Then, it follows from Definition 3 that $\sum_R \boldsymbol{u_i} \leq \boldsymbol{v(R)}$* ***for all $\mathbf{R} \subseteq \mathbf{N}$ and for all pairwise-feasible outcomes $(\boldsymbol{u}, \boldsymbol{x})$ with $\boldsymbol{x(R)} = \mathbf{R}$. In particular, $\sum_N \boldsymbol{u_i} \leq \boldsymbol{v(N)}$****. Therefore,* ***every pairwise-feasible payoff is feasible.*** ■

The natural solution concept is that of stability (the general definition of stability is given in Sotomayor, 2009b). For the one-sided assignment game, stability is equivalent to the concept of **pairwise-stability**.

Definition 5 *The pairwise-feasible payoff u is* ***pairwise-stable*** *if*

(i) $u_i \geq 0$ *for all $i \in N$ and*

(ii) $u_i + u_j \geq a_{ij}$ *for all $\{i, j\} \subseteq N$.*

If x is compatible with u we say that (u, x) is a pairwise-stable outcome.

Condition (i) (individual rationality) means that in a pairwise-stable situation a player always has the option of remaining unmatched. Condition (ii) ensures the stability of the payoff distribution: If it is not satisfied for some agents i and j then it would pay for them to break up their present partnership(s) and form a new one together, as this would give them each a higher payoff. In this case, we say that $\{i, j\}$ **blocks** u.

We now define the **core** of (N, a), which we denote by C:

Definition 6 *We say that $u \in C$ if $\sum_N u_i = v(N)$ and $\sum_S u_i \geq v(S)$ for all $S \subseteq N$.*

The following example shows that the core of this model may be empty.

Example 1 *Consider $N = \{1, 2, 3\}$ and $a_{ij} = 1$ for all $\{i, j\} \subseteq N$. For every feasible payoff u there exist two players i and j such that $u_i + u_j < 1$. Hence, the core of this game is empty.*

Definition 7 *Let* (u, x) *be a pairwise-feasible outcome. Let* $R \subseteq N$*. We say that* **R is a stable coalition for** $(\boldsymbol{u}, \boldsymbol{x})$ *if (a)* $x(R) = R$*, (b)* $u_i + u_{x(i)} = a_{ix(i)}$ *for all* $i \in R$ *and (c)* $u_i + u_j \geq a_{ij}$ *for all* $\{i, j\} \subseteq R$*.*

Remark 2 *Notice that if* R *is stable for* (u, x) *it must be the case that* $\sum_R u_i \geq v(R)$*, according to Definition 7. On the other hand,* $\sum_R u_i \leq v(R)$*, as stated in Remark 1. Therefore,* $\sum_R \boldsymbol{u}_i = \boldsymbol{v}(\boldsymbol{R})$*.* ■

2.2 Preliminary results for the core

In our environment, the concepts of stability and the core are equivalent, as established in the following proposition.

Proposition 1 *The set of pairwise-stable payoffs coincides with the core of* (N, a)*.*

Proof. Suppose u is a pairwise-stable payoff. Then, u is feasible and so

$$\sum_N u_i \leq v(N) \tag{1}$$

according to Remark 1. Moreover, consider any coalition S and let y be a feasible matching such that $y(S) = S$ and $v(S) = \sum_y a_{ij}$. The pairwise-stability of u implies that $u_i + u_{y(i)} \geq a_{iy(i)}$ for all $i \in S$, so

$$\sum_S u_i \geq v(S) \text{ for all coalition } S. \tag{2}$$

Under (1) and (2) it follows that $\sum_N u_i = v(N)$ and $\sum_S u_i \geq v(S)$ for all $S \subseteq N$, so u is in the core.

Now, suppose u is in the core. Definition 6 implies that $u_i + u_j \geq v(i, j) = a_{ij}$ for every coalition $\{i, j\}$ and $u_i \geq v(i) = 0$ for all $i \in N$, so u does not have any blocking pair and is individually rational. To see that u is pairwise-feasible, let x be a feasible matching such that $v(N) = \sum_x a_{ij}$. Use that $\sum_N u_i = v(N)$ and $u_i + u_{x(i)} \geq a_{ix(i)}$ for all $i \in N$, to get that $\sum_N u_i = \sum_x a_{ij} \leq \sum_{i \leq x(i)} (u_i + u_{x(i)}) = \sum_N u_i$, so the inequality cannot be strict, which implies $u_i + u_{x(i)} = a_{ix(i)}$ for all $i \in N$. Since $u_i \geq 0$, it follows that $u_i = 0$ if $x(i) = i$. Hence, u is pairwise-stable and the proof is complete. ■

In what follows, given its equivalence with the core concept, the concept of pairwise-stability will be called **corewise-stability**.

The following proposition, proven by Sotomayor (2005a, 2009a) and Talman and Yang (2011), makes clear why, similarly to the two-sided assignment game and in contrast to the discrete version (the roommate-problem), we can concentrate on the payoffs to the agents rather than on the underlying matching. Indeed, it shows that the set of corewise-stable outcomes is the Cartesian product of the set of corewise-stable payoffs and the set of optimal matchings. We state the proposition without its proof.

Proposition 2 *(a) If x is an optimal matching then it is compatible with any corewise-stable payoff u.*
(b) If (u, x) is a corewise-stable outcome then x is an optimal matching.

A consequence of Proposition 2 (a) is that, similarly to the two-sided assignment game, every unmatched player in a corewise-stable outcome has a zero payoff at any corewise-stable outcome in the one-sided assignment game. Corollary 1 states this result.

Corollary 1 *Let x be an optimal matching. If i is unmatched at x then $u_i = 0$ for all corewise-stable payoffs u.*[7]

Proof. Let $u \in C$. Under Proposition 2 (a), u is compatible with x, so $u_i = 0$ by the pairwise-feasibility of u. ■

3 Tradewise-stable outcomes

In this section, we introduce a key solution concept for the theory developed in this paper: a tradewise-stable outcome. Tradewise-stable outcomes satisfy properties similar to, but weaker than, stable outcomes.

Definition 8 *The **outcome** (u, x) is **tradewise-stable** (we will often use **t-stable**, for short) if it is pairwise-feasible, individually rational and no blocking pair $\{i, j\}$ exists where either i or j are matched at x. A matching x is tradewise-stable (t-stable) if there is some payoff u such that the outcome (u, x) is t-stable. The payoff u is a **tradewise-stable (t-stable) payoff** if there is some matching x such that (u, x) is a t-stable outcome.*

Clearly, every corewise-stable outcome is t-stable. However, t-stable outcomes are not necessarily stable. For instance, the outcome where every player is unmatched and obtains a payoff of 0 is t-stable, but it is not corewise-stable in any game where at least one partnership

[7] This result was proved in Demange and Gale (1985) for a two-sided matching market where the utilities are continuous, so it applies to the two-sided assignment game of Shapley and Shubik (1972).

creates a positive surplus. Moreover, this example of a t-stable outcome allows us to state that *the set of t-stable outcomes is always non-empty.*

To discuss the difference between corewise-stable and t-stable outcomes, consider a pairwise-feasible and individually rational outcome (u, x). We denote by $T(x)$ the set of all players who are matched at x, and by $U(x)$ the set of players who are unmatched at x. That is,

$$T(x) \equiv \{j \in N; x(j) \neq j\} \text{ and } U(x) \equiv N \backslash T(x).$$

The outcome (u, x) is t-stable if and only if no player in $T(x)$ can form a blocking pair neither with another player in $T(x)$ (which implies that the coalition $T(x)$ is a stable coalition for (u, x)) nor with any player in $U(x)$. In this sense, we could say that a t-stable outcome is "internally stable." We refer to $T(x)$ as the stable active coalition for (u, x). However, to be corewise-stable the outcome also needs to be "externally stable," in the sense that no pair of players in $U(x)$ can block the outcome either. A t-stable outcome might not be "externally stable."

We denote by S **the set of t-stable payoffs**:

$$S \equiv \{u \in \mathbb{R}^n; u \text{ is t-stable}\}.$$

Also, given any t-stable matching x, denote:

$$S(x) \equiv \{u \in S; u \text{ is compatible with } x\}.$$

Before we state results concerning the t-stable outcomes, we make one remark on the set $T(x)$ and another on the set $U(x)$, for any t-stable outcome (u, x).

Remark 3 *Notice that if (u, x) is t-stable then the set of players $T(x)$ who are matched at x and all the subsets $R \subseteq T(x)$ such that $x(R) = R$ are stable coalitions for (u, x). Therefore, by Remark 2, $\sum_{T(x)} u_i = v(T(x))$, and for such coalitions R, we also have $\sum_R u_i = v(R)$.* ■

Remark 4 *Notice that if the t-stable outcome (u, x) is unstable then $v(U(x)) > 0$. In fact, let $\{j, k\}$ be a blocking pair. Then $\{j, k\} \subseteq U(x)$, so*

$$0 = \sum_{U(x)} u_i = (u_j + u_k) + \sum_{U(x)\backslash\{j,k\}} u_i < a_{jk} + \sum_{U(x)\backslash\{j,k\}} a_{ii} \leq v(U(x)).$$

Therefore, $v(U(x)) > 0$. ■

Our next results highlight relationships between t-stable matchings and optimal matchings, as well as between t-stable outcomes and corewise-stable outcomes. Proposition 3 states that a t-stable matching may not be optimal but it is always part of an optimal matching. It uses Lemma 1 (see the Appendix), which shows that $v(T(x)) + v(U(x)) = v(N)$ for any t-stable outcome (u, x).

Proposition 3 *Let (u, x) be a t-stable outcome. Then, the set of active partnerships of x is part of an optimal matching.*

Proof. The proof is immediate after the fact that $v(T(x)) + v(U(x)) = v(N)$ because of Lemma 1, $N = T(x) \cup U(x)$ and $T(x) \cap U(x) = \varnothing$. ■

Proposition 4 states a result complementary to Proposition 3: the only t-stable payoffs compatible with an optimal matching are the corewise-stable payoffs.

Proposition 4 *Let (u, x) be a t-stable outcome. Suppose x is optimal. Then $u \in C$.*

Proof. Denote R a set of pairs $\{i, j\} \subseteq T(x)$ such that $v(T(x)) \equiv \sum_R a_{ij}$. Since x is optimal, then

$$v(N) = \sum_x a_{ij} = \sum_{x|_{T(x)}} a_{ij} + \sum_{x|_{U(x)}} a_{ij} = \sum_{x|_{T(x)}} a_{ij} = v(T(x)),$$

where the last equality follows from Proposition 3. Then, under Lemma 1, we have that $v(U(x)) = 0 = \sum_{U(x)} u_i$ so, as stated in Remark 4, there is no blocking pair in $U(x)$, which implies that (u, x) is corewise-stable. Hence, we have completed the proof. ■

Proposition 4 leads to the following corollary.

Corollary 2 *A t-stable outcome (u, x) is corewise-stable if and only if $\sum_N u_i = v(N)$.*

Proof. Consider the t-stable outcome (u, x). If it is corewise-stable then $\sum_N u_i = v(N)$ according to Definition 6. On the hand, if $\sum_N u_i = v(N)$ then the matching x is necessarily optimal, so u is a corewise-stable payoff as stated in Proposition 4. ■

Finally, we introduce the idea of an extension of a t-stable outcome, which will be useful in the next section. In words, a feasible outcome (w, z) extends the t-stable outcome (u, x) if all the players in the stable active coalition of (u, x) keep their payoff but some players who were unmatched in (u, x) obtain a positive payoff (hence, they are matched) in (w, z).

Definition 9 *Let (u, x) be a t-stable outcome. We say that the feasible outcome* $\boldsymbol{(w, z)}$ ***extends*** $\boldsymbol{(u, x)}$ *if $w_j > u_j$ for some $j \notin T(x)$ and $w_j = u_j$ for all $j \in T(x)$. If (w, z) is t-stable (respectively, corewise-stable) then (w, z) is said to be a t-stable (respectively, corewise-stable) extension of (u, x).*

Sometimes, we will refer to a t-stable outcome that does not have any extension as a *non-extendable outcome*.

Proposition 5 states that any Pareto improvement of a t-stable outcome through another t-stable outcome is necessarily an extension of that outcome.

Proposition 5 *Let (u, x) and (w, y) be t-stable outcomes. Suppose $w > u$.*[8] *Then (w, y) is a t-stable extension of (u, x).*

Proof. To show that (w, y) is a t-stable extension of (u, x), we need to prove that $j \notin T(x)$ for all j such that $w_j > u_j$ (see Definition 9). Consider a player j such that $w_j > u_j$ and suppose, by contradiction, that $j \in T(x)$. Given that $w_j > 0$, we have that $j \in T(y)$. Then, $j \in T(x) \cap T(y)$ and so $j \in M_w$. Denote $k \equiv y(j)$. Lemma 3 implies that $k \in M_u$. Therefore, $u_k > w_k$, which contradicts the assumption that $w > u$. Hence, (w, y) extends (u, x). ■

4 Constrained-optimal tradewise-stable outcomes

Of particular interest for our analysis is the set of the t-stable outcomes that are not dominated, via coalition N, by any other t-stable outcome. This section introduces the set of *constrained-optimal* tradewise-stable outcomes and provides important properties of this set. To introduce the set, let us first formally define the notion of Pareto optimality.

Definition 10 *Let A be a set of payoffs. The payoff u is **Pareto-optimal (PO) in A** (or among all payoffs in A) if it belongs to A and there is no payoff w in A such that $w > u$.*

If u is PO in A and x is compatible with u, we say that (u, x) is a PO outcome in A.

The case in which A corresponds to the set of t-stable payoffs, that is, $A = S$, plays an important role in our theory.

Definition 11 *The payoff u is a **constrained-optimal tradewise-stable (optimal t-stable**, for short) **payoff** if it is a t-stable payoff and it is PO in the set of t-stable payoffs. The outcome (u, x) is an **optimal t-stable outcome** if (u, x) is a t-stable outcome and u is an optimal t-stable payoff.*

[8]Given two vectors $w, v \in \mathbb{R}^n$, we will denote $w > u$ if $w_j \geq u_j$ for all players $j \in N$ and $w_j > u_j$ for at least one player $j \in N$.

The **set of optimal t-stable payoffs** will be denoted by S^*:

$$S^* \equiv \{u \in S; u \text{ is Pareto optimal in } S\}.$$

We notice that, given Definition 9, every optimal t-stable outcome is non-extendable.

Similarly, if u is Pareto optimal in A and A is the set of individually rational and feasible payoffs we will refer to u as a PO feasible payoff.

Remark 5 *It follows from Definition 10 that an individually rational and feasible payoff u is PO feasible if and only if $\sum_N u_i = v(N)$. Thus, every corewise-stable payoff is PO feasible. However, the Pareto optimality of a payoff is not enough to guarantee its corewise-stability. For instance, in Example 1, the payoff $u = (1, 0, 0)$ is not in the core but it is PO feasible, since $\sum_N u_i = v(N)$.* ■

Optimal t-stable payoffs are, by definition, undominated in the set S. Next result shows that every t-stable payoff which is not a optimal t-stable payoff is necessarily dominated for some optimal t-stable payoff. It uses Lemma 5 (see the Appendix), which shows that the set of optimal t-stable payoffs S^* is a non-empty and compact set of $\mathbb{R}^n$.[9]

Proposition 6 *Let u be a t-stable payoff which is not optimal t-stable, that is, $u \in S \backslash S^*$. Then there is some optimal t-stable payoff w_u such that $w_u > u$.*

Proof. Suppose, by way of contradiction, that there is no payoff w in S^* such that $w > u$. Since $u \notin S^*$, there is some $w^1 \in S$ such that $w^1 > u$. Then, by contradiction, $w^1 \notin S^*$, so there is some $w^2 \in S$ such that $w^2 > w^1 > u$. Again, w^2 cannot be in S^*. By repeating this procedure, we obtain an infinite sequence $(w^t)_{t=1,2,\ldots}$ of t-stable payoffs with distinct terms. On the other hand, there is a finite number of t-stable matchings, so there is some t-stable matching x which is compatible with infinitely many terms of the sequence. Denote $(v^t)_{t=1,2,\ldots}$ that subsequence. All the members of the subsequence are in S so $\sum_{T(x)} v_j^t = v(T(x))$ for all $t = 1, 2, \ldots$. However, $v^1 < v^2$, so $v(T(x)) = \sum_{T(x)} v_j^1 = \sum_N v_j^1 < \sum_N v_j^2 = \sum_{T(x)} v_j^2 = v(T(x))$, which is an absurd. Hence, there is some $w_u \in S^*$ such that $w_u > u$. ■

Proposition 6 allows us to establish an interesting corollary: there is some payoff in S^* that dominates every t-stable payoff outside S^*.

Corollary 3 *There is some optimal t-stable payoff $w^* \in S^*$ such that $\sum_N w_j^* > \sum_N u_j$ for all $u \in S \backslash S^*$.*

[9] Lemma 2 shows that the set S is also a non-empty and compact set of $\mathbb{R}^n$.

Proof. Since S^* is a non-empty and compact set of $\mathbb{R}^n$ (see Lemma 5 in the Appendix) and every continuous function defined in a compact set has a maximum in this set, there is some $w^* \in S^*$ such that $\sum_N w_j^* \geq \sum_N w_j$ for all $w \in S^*$. Now use Proposition 6 to get that $\sum_N w_j^* > \sum_N u_j$ for all $u \in S \backslash S^*$. ■

Proposition 6, together with Proposition 5, also implies that not only are the optimal t-stable outcomes non-extendable but they are the only non-extendable outcomes. This result is stated in Corollary 4.

Corollary 4 *The set of optimal t-stable outcomes equals the set of non-extendable outcomes.*

Proof. Let $(u, x) \in S^*$. Then (u, x) cannot have any t-stable extension, as stated in Definition 9. The other direction is immediate from propositions 6 and 5. ■

The next property that we prove is that all optimal t-stable outcomes are equally efficient, in the sense that the players' total payoff is the same in every optimal t-stable payoff. This property will be proven in Proposition 7, which requires a lemma (Lemma 7 in the Appendix) that states an interesting property: If (u, x) and (w, y) are optimal t-stable outcomes, then every unmatched player at x has zero payoff at y. Equivalently, if j has a positive payoff under an optimal t-stable outcome (u, x) then j is matched under every optimal t-stable outcome; in particular, j is matched under every corewise-stable outcome.

Proposition 7 *Let (u, x) and (w, y) be optimal t-stable outcomes. Then, $\sum_N u_j = \sum_N w_j$.*

Proof. Set

$$B_1(x) = \{j \in T(x) \cap T(y); x(j) \in T(x) \cap T(y)\};$$

$$B_1(y) = \{j \in T(x) \cap T(y); y(j) \in T(x) \cap T(y)\};$$

$$B_2(x) = \{j \in T(x) \cap T(y); x(j) \in T(x) \backslash T(y)\};$$

$$B_2(y) = \{j \in T(x) \cap T(y); y(j) \in T(y) \backslash T(x).$$

Clearly,

$$T(x) \cap T(y) = B_1(x) \cup B_2(x) = B_1(y) \cup B_2(y). \tag{3}$$

Under Remark 3,

$$\sum_{B_1(x)} u_j = v(B_1(x)) \text{ and } \sum_{B_1(y)} w_j = v(B_1(y)). \tag{4}$$

On the other hand, according to Lemma 7, $u_j = w_j$ for all $j \in B_2(x)$ and $w_j = u_j$ for all $j \in B_2(y)$, so

$$\sum_{B_2(x)} u_j = \sum_{B_2(x)} w_j \text{ and } \sum_{B_2(y)} w_j = \sum_{B_2(y)} u_j. \tag{5}$$

Moreover, Lemma 7 implies that

$$\sum_{T(x)\setminus T(y)} u_j = 0 \text{ and } \sum_{T(y)\setminus T(x)} w_j = 0. \tag{6}$$

Therefore, we can write,

$$\sum_N u_j = \sum_{T(x)} u_j + \sum_{N\setminus T(x)} u_j = \sum_{T(x)} u_j = \sum_{T(x)\cap T(y)} u_j + \sum_{T(x)\setminus T(y)} u_j = \sum_{T(x)\cap T(y)} u_j =$$
$$\sum_{B_1(x)} u_j + \sum_{B_2(x)} u_j = v(B_1(x)) + \sum_{B_2(x)} w_j \leq \sum_{B_1(x)} w_j + \sum_{B_2(x)} w_j =$$
$$\sum_{T(x)\cap T(y)} w_j = \sum_{T(x)\cap T(y)} w_j + \sum_{T(y)\setminus T(x)} w_j = \sum_{T(y)} w_j = \sum_{T(y)} w_j + \sum_{N\setminus T(y)} w_j = \sum_N w_j,$$

where the fourth equality uses (6); the fifth equality follows from (3); the sixth equality follows from (4) and (5); the inequality follows from the fact that $B_1(x) \subseteq T(y)$ and y is a t-stable matching, and so $B_1(x)$ cannot block y; the seventh equality follows from (3); and the eighth equality follows from (6).

Then,

$$\sum_N u_j \leq \sum_N w_j. \tag{7}$$

By reverting the roles between (u, x) and (w, y) in the expression (7) we obtain

$$\sum_N w_j \leq \sum_N u_j. \tag{8}$$

According to (7) and (8) we get that $\sum_N u_j = \sum_N w_j$ and we have completed the proof.

■

Proposition 7 implies that every payoff in S^* reaches the maximum total payoff among all t-stable payoffs. Together with Corollary 3, Proposition 7 also implies that every payoff in S^* dominates every t-stable payoff not in S^*.

We will refer to a matching that is compatible with an optimal t-stable payoff as **quasi-optimal**. Proposition 8 asserts that every quasi-optimal matching is compatible with any optimal t-stable outcome. That is, Proposition 8 states for the set of optimal t-stable outcomes a property similar to that stated in Proposition 2 (a) for the set of corewise-stable outcomes. Indeed, the set of optimal t-stable outcomes is the Cartesian product of the set of optimal t-stable payoffs and the set of quasi-optimal matchings.

Proposition 8 *Let (u, x) be an optimal t-stable outcome. Then, x is compatible with any optimal t-stable payoff.*

Proof. Let (w, y) be any optimal t-stable outcome. We want to show that $w_i + w_j = a_{ij}$ if $x(i) = j$ and $w_i = 0$ if $i \in U(x)$. Notice that

$$\sum_{T(x)} u_i = \sum_{T(x)} u_i + \sum_{U(x)} u_i = \sum_{N} u_i = \sum_{N} w_i = \sum_{T(x)} w_i + \sum_{U(x)} w_i =$$
$$\sum_{T(x)} w_i + \sum_{T(y)\backslash T(x)} w_i + \sum_{U(y)\backslash T(x)} w_i = \sum_{T(x)} w_i,$$

where the third equality is due to Proposition 7; and Lemma 7 was used in the last equality to conclude that $\sum_{T(y)\backslash T(x)} w_i = 0$. Then,

$$\sum_{T(x)} u_i = \sum_{T(x)} w_i. \tag{9}$$

We can write $T(x) = (T(x) \cap T(y)) \cup (T(x) \backslash T(y))$. From Lemma 7 it follows that $\sum_{T(x)\backslash T(y)} u_i = \sum_{T(x)\backslash T(y)} w_i = 0$. Then, $\sum_{T(x)} u_i = \sum_{T(x)\cap T(y)} u_i$ and $\sum_{T(x)} w_i = \sum_{T(x)\cap T(y)} w_i$. Therefore, using (9) we obtain

$$\sum_{T(x)\cap T(y)} u_i = \sum_{T(x)\cap T(y)} w_i. \tag{10}$$

To prove that $w_i + w_j = a_{ij}$ if $x(i) = j$, set $G \equiv \{\{i, j\} \subseteq T(x) \cap T(y); x(i) = j\}$ and $H \equiv \{\{i, j\} \subseteq T(x); i \in T(x) \cap T(y), j \in T(x) \backslash T(y) \text{ and } x(i) = j\}$. Lemma 7 implies that

$$\sum_{H} u_i = \sum_{H} w_i. \tag{11}$$

Moreover, since w is t-stable, we must have that $w_i + w_j \geq a_{ij}$ for all $\{i, j\} \subseteq T(x) \cap T(y)$. Then, in particular,

$$\sum_{G} a_{ij} \leq \sum_{G} (w_i + w_j). \tag{12}$$

Therefore,

$$\sum_{T(x)\cap T(y)} u_i = \sum_{G} (u_i + u_j) + \sum_{H} u_i = \sum_{G} a_{ij} + \sum_{H} w_i \leq \sum_{G} (w_i + w_j) + \sum_{H} w_i = \sum_{T(x)\cap T(y)} w_i,$$

where we used (11) in the second equality and (12) in the inequality. According to (10), the inequality must be an equality, and so we have proved that $w_i + w_j = a_{ij}$ for all $\{i, j\}$ such that $\{i, j\} \subseteq (T(x) \cap T(y))$ and $x(i) = j$. Next, consider $\{i, j\}$ with $x(i) = j$ such that either $i \in (T(x) \backslash T(y))$ or $j \in (T(x) \backslash T(y))$. Without loss of generality suppose that $i \notin T(y)$. Then, $w_i = 0$ and, under Lemma 7, we have that $u_i = 0$ and $w_j = u_j$, from which follows that $w_i + w_j = u_i + u_j = a_{ij}$, so $w_i + w_j = a_{ij}$.

It remains to show that $w_i = 0$ for all $i \in U(x)$. But this is immediate from the fact that if $i \in T(y) \backslash T(x)$, then Lemma 7 implies that $w_i = 0$. Hence, the matching x is compatible with w and we have completed the proof. ■

Our final result in this section provides another feature that is shared by all optimal t-stable outcomes. It also helps us better understand the structure of the optimal t-stable and that of the t-stable outcomes. The result states that if an optimal t-stable outcome is not corewise-stable then the set of pairs of blocking agents is the same for every optimal t-stable outcome. Moreover, each of those pairs also blocks any t-stable outcome. This result implies, in particular, that if an agent is unmatched at some optimal t-stable outcome but he is matched with zero payoff at another optimal t-stable outcome then that agent will never be part of a blocking pair in an optimal t-stable outcome. And since every t-stable outcome is extended by an optimal t-stable outcome, any blocking agent of an optimal t-stable outcome is unmatched at any t-stable outcome.

Proposition 9 *Let $(u, x) \in S^* \backslash C$ and let $\{j, k\}$ be a blocking pair for (u, x). Then, $\{j, k\}$ blocks (w, y), for any $(w, y) \in S$. In particular, $\{j, k\} \subseteq U(y)$, for any $(w, y) \in S$.*

Proof. Notice first that, given that $\{j, k\}$ blocks (u, x), it is the case that j and k are unassigned at x, so $0 = u_j + u_k < a_{jk}$. Moreover, j and k have a zero payoff at any optimal t-stable outcome, under Lemma 7 (even if j or k were matched in an optimal t-stable outcome, they would obtain a zero payoff). Therefore, the sum of the payoffs of j and k in any optimal t-stable outcome is less than a_{jk}, so $\{j, k\}$ blocks any optimal t-stable outcome. Now, use propositions 4 and 6 to get that any (w, y) is extended by some optimal t-stable outcome, so $\{j, k\}$ blocks any t-stable (w, y). In particular, j and k are unassigned at y. Hence, the proof is complete. ■

The properties that we have proved in this section suggest that the set of optimal t-stable outcomes constitutes a natural solution concept if one cares about payoffs that are "as stable as possible." First, every optimal t-stable outcome is internally stable, so no active player has an incentive to look for other partners inside or outside the set of active players. Second, each optimal t-stable outcome provides the maximum surplus out of the set of internally stable outcomes; and all the internally stable outcomes outside the set provide less surplus. Third, any internally stable outcome that is not optimal t-stable can be naturally extended to an optimal t-stable outcome. Fourth, all optimal t-stable outcomes are compatible with the same matchings. Fifth, all the previous properties of optimal t-stable outcomes replicate properties that are satisfied by the corewise-stable outcomes. Finally, the set of optimal t-stable outcomes is always non-empty and, as will be proved in section 6, the set of optimal t-stable payoffs coincides with the core, when the core is non-empty.

5 A dynamic partnership formation process

The intuitive idea of an optimal t-stable outcome is that it corresponds to an outcome that we can expect to occur in an idealized environment where agents take decisions under the assumption of cooperative behavior. As stated in the Introduction, our theory is based on the idea of collective rationality, in the sense that a player only engages in cooperation that is optimal for him/her. Once an agreement between a pair of agents is reached, then they must be certain than there will be no more favorable option elsewhere.

In fact, the properties of the t-stable outcomes allow us to envision a dynamic and finite partnership formation process of t-stable outcomes that ends with an optimal t-stable outcome. At every step t of this process, the current unmatched agents work among themselves to form partnerships and to split the gains obtained in these partnerships. Once the vector of payoffs of any new partnership of matched agents is established, the unmatched agents are not interested in trading with any matched agent, currently and previously formed. Because of the properties concerning the extensions of t-stable outcomes (Proposition 6 and Corollary 4) and the fact that the optimal t-stable payoffs extend the t-stable payoffs (Proposition 4), this process always ends in an optimal t-stable outcome, which provides further support for the set of optimal t-stable outcomes as a natural solution concept for the one-sided assignment game.

To describe the sequential process, consider any t-stable outcome (u, x) such that $T(x) \neq \varnothing$. Define

$$A(u, x) \equiv \{(u^p, x^p) \text{ is t-stable}; T(x^p) \subseteq T(x), \text{ and } x^p(j) = x(j) \text{ and } u_j^p = u_j \text{ for all } j \in T(x^p)\}.$$

That is, $A(u, x)$ is the set of t-stable outcomes in which the pairs that are matched do it according to x and have the same payoffs as u.

Denote $B_1(u, x) \equiv \{S^p \subseteq T(x); S^p \neq \varnothing \text{ and } S^p = T(x^p) \text{ for some } (u^p, x^p) \in A(u, x)\}$. That is, $B_1(u, x)$ is the set of the non-empty stable active coalitions of the t-stable outcomes in $A(u, x)$. The set $B_1(u, x)$ is non-empty, since $T(x) \in B_1(u, x)$. Furthermore, $B_1(u, x)$ is finite and is endowed with the partial order defined by the set inclusion relation. Then $B_1(u, x)$ has a minimal element, that is, there exists some coalition that does not have any sub-coalition in $B_1(u,x)$. Set $D_1(u,x)$ any such coalition[10] and let (u^1,x^1) be the corresponding t-stable outcome in $A(u,x)$. Moreover, define $C_1(u,x) \equiv D_1(u,x)$.

If $D_1(u,x) \neq T(x)$, that is, if $(u^1,x^1) \neq (u,x)$, we denote $B_2(u,x) \equiv \{S^p \subseteq T(x); S^p = T(x^p)$ for some $(u^p,x^p) \in A(u,x)$ such that (u^p,x^p) is an extension of $(u^1,x^1)\}$. The set $B_2(u,x)$ is also non-empty because $T(x) \in B_2(u,x)$. By using similar arguments as above, we obtain the existence of a minimal element of $B_2(u,x)$. Set $D_2(u,x)$ any such coalition and let (u^2,x^2) be the corresponding t-stable outcome in $A(u,x)$. By construction, (u^2,x^2) is a t-stable outcome in $A(u,x)$ and it extends (u^1,x^1). Also, by defining $C_2(u,x) \equiv D_2(u,x) \backslash D_1(u,x)$, we obtain a partition $\{C_1(u,x), C_2(u,x)\}$ of the set $T(u^2)$ of active player in the t-stable outcome (u^2,x^2).

By continuing this procedure, we obtain a finite sequence of t-stable outcomes in $A(u,x)$: $(u^1,x^1),(u^2,x^2),\ldots,(u^k,x^k)$, where $(u^k,x^k)=(u,x)$ and (u^{p+1},x^{p+1}) extends (u^p,x^p) for all $p=1,\ldots,k-1$. Then, $T(x)=D_k(u,x)=C_1(u,x)\cup\ldots\cup C_k(u,x)$.

We can describe the sequential process generated by $(D_p(u,x))_{p=1,\ldots,k}$ as follows. The first step of such a process yields (u^1,x^1), the second step yields (u^2,x^2), and so on. At every step t, no agent j in $N\backslash D_t(u,x)$ is willing to pay any agent i in $D_t(u,x)$ more than u_i^t. Thus, at any step t, the current outcome (u^t,x^t) is t-stable and is an extension of the current outcome (u^{t-1},x^{t-1}).

If the t-stable outcome (u,x) is not an optimal t-stable outcome then there is another t-stable outcome that extends (u,x) and the procedure could continue. It only ends when no interaction is able to benefit the agents involved, in which case the core is reached, or when any new interaction leads to a set of matched agents which is not internally stable. In any case, the final outcome is an optimal t-stable outcome. Since any optimal t-stable outcome can be formed this way and those outcomes cannot be extended by another t-stable outcome, we have that the set of optimal t-stable outcomes is the set of all the outcomes which constitute the final steps of such procedures.

The following example shows that the number of processes that reach an optimal t-stable outcome may vary inside the set of optimal t-stable outcomes. This happens even though, in the example there is only one matching which is compatible with all the optimal t-stable payoffs.

Example 2 *The set of players is $N=\{1,2,3,4\}$ and the surplus of the partnerships is $a_{12}=10$, $a_{13}=4$, $a_{34}=12$, and $a_{ij}=0$ for the other partnerships. The set of optimal t-stable*

[10]There can be several minimal coalitions.

outcomes, which coincides with the corewise-stable outcomes, is the set of outcomes (u, x) *that satisfy* $x_{12} = 1$*,* $x_{34} = 1$ *and the payoffs are non-negative numbers with* $u_1 + u_2 = 10$*,* $u_3 + u_4 = 12$*, and* $u_1 + u_3 \geq 4$*.*

(i) For the optimal t-stable outcome with $u = (6, 4, 5, 7)$*, there are two different processes: either (*$u^1 = (6, 4, 0, 0)$*,* x^1 *with* $x^1_{12} = 1$ *and the other entries are* 0*) or (*$u^1 = (0, 0, 5, 7)$*,* x^1 *with* $x^1_{34} = 1$ *and the other entries are* 0*); in both cases* $(u^2, x^2) = (u, x)$*.*

(ii) For the optimal t-stable outcome with $u = (6, 4, 3, 9)$*, there is only one process: (*$u^1 = (6, 4, 0, 0)$*,* x^1 *with* $x^1_{12} = 1$ *and the other entries are* 0*) and* $(u^2, x^2) = (u, x)$*. Note that (*$u^1 = (0, 0, 3, 9)$*,* x^1 *with* $x^1_{34} = 1$ *and the other entries are* 0*) cannot be part of the process because it is not a t-stable outcome: players* 1 *and* 3 *block this outcome.*

(iii) Finally, for the optimal t-stable outcome with $u = (3, 7, 3, 9)$*, the only process is the trivial one-step process:* $(u^1, x^1) = (u, x)$*.* ■

We should emphasize that the dynamic partnership formation process that we have described is not an algorithm to find the set of optimal t-stable allocations. Also, we do not provide a history behind the coalitional interactions among individuals that would lead to the formation of an optimal t-stable allocation. The process only postulates a plausible theory according to which every optimal t-stable allocation may arise. It is based on the property that, for any optimal t-stable allocation, there is a coalitional structure that points to a possible ordering in the formation of these coalitions and which ends in that allocation. Moreover, this coalitional structure naturally defines a dynamic procedure where players only trade under the assumption of optimal cooperative behavior and where some players end up unmatched. Only optimal t-stable allocations occur at the end of such a partnership formation process (this does not imply that every such allocation necessarily occurs). Thus, the process provides a justification of the concept of optimal t-stability as the natural cooperative solution concept for this market.

We will come back to the partnership formation process in the next section, once we analyze the relationship between the set of optimal t-stable outcomes and the core.

6 Constrained-optimal tradewise-stable outcomes and corewise-stable outcomes

Section 4 states several appealing properties of the set of optimal t-stable outcomes. They allowed us to propose this set as a natural solution concept for the one-sided assignment game. In the current section, we further support our proposal as a new stability concept by showing that the set of optimal t-stable payoffs and the core coincide, when the core is not empty. Moreover, the relationship between the core and the set of optimal t-stable outcomes constitutes a useful tool to establish conditions under which the core is non-empty in these environments.

Proposition 10 proves two important properties of the t-stable outcomes when the core is non-empty. First, it states that for any t-stable outcome which is not corewise-stable, it is always possible to construct a new outcome that keeps the payoff of each matched player and is corewise-stable. Second, it shows that the sum of the payoffs of the set of agents that are matched in a t-stable outcome is always maintained in any corewise-stable outcome.

Proposition 10 *Let (u, x) be a t-stable outcome which is not corewise-stable. Suppose the set of corewise-stable outcomes is non-empty. Then:*
(a) there exists a corewise-stable outcome (u^*, z) that extends (u, x), and
(b) $\sum_{T(x)} u_i = \sum_{T(x)} w_i$ for all $w \in C$.

Proof. According to Proposition 3, the set of active partnerships of x is part of some optimal matching. Therefore, there is some optimal matching z such that $z(i) = x(i)$ for all $i \in T(x)$. Let (w, z) be any corewise-stable outcome. Construct the outcome (u^*, z) such that $u_i^* = u_i$ for all $i \in T(x)$ and $u_i^* = w_i$ for all $i \in N \backslash T(x)$. The outcome (u^*, z) is feasible. We claim that $u^* \in C$. In fact, suppose $\{i, j\}$ blocks u^*. Then, $u_i^* + u_j^* < a_{ij}$. Notice that, by construction, $u^* \geq u$, so $u_i + u_j < a_{ij}$. Since x is t-stable, we must have that $\{i, j\} \subseteq N \backslash T(x)$. On the other hand, the corewise-stability of w implies that $u_i^* + u_j^* = w_i + w_j \geq a_{ij}$, which contradicts the assumption that $\{i, j\}$ blocks u^*. Then, u^* does not have any blocking pair.

The property that (u^*, z) is individually rational is immediate from the individual rationality of (u, x) and (w, z). According to Definition 6, it remains to show that $v(N) = \sum_N u_i^*$. Write:

$$v(N) \leq \sum_N u_i^* = \sum_{T(x)} u_i + \sum_{N \backslash T(x)} w_i = v(T(x)) + \sum_{N \backslash T(x)} w_i \leq \sum_{T(x)} w_i + \sum_{N \backslash T(x)} w_i = \sum_N w_i = v(N), \tag{13}$$

where in the first inequality we used the fact that u^* does not have any blocking pair, in the second equality we used Remark 3, and the second inequality follows from the corewise-stability of w. Then, the inequalities in (13) must be equalities, so $v(N) = \sum_N u_i^*$. Therefore, we have proved that (u^*, z) is corewise-stable.

To see that (u^*, z) extends (u, x), use that $u_i^* \geq u_i$ for all $i \in N$. Given that (u^*, z) is corewise-stable and that (u, x) is unstable, we have that $\{j \in N; u_j^* > u_j\} \neq \varnothing$. On the other hand, since $u_j^* = u_j$ for all $j \in T(x)$, it follows that $\{j \in N; u_j^* > u_j\} \subseteq N \backslash T(x)$. Then, according to Definition 9, (u^*, z) extends (u, x), and we have proved part (a) of the proposition.

Now use that the inequalities in (13) must be equalities, so $\sum_{T(x)} u_i = v(T(x)) = \sum_{T(x)} w_i$, which proves assertion (b). Hence, we have completed the proof. ■

It is worth mentioning that while Proposition 10 states that any t-stable outcome which is not in the core can be extended to a corewise-stable payoff, it does not ensure that it is possible to "shrink" any corewise-stable payoff. For example, in Case (iii) of Example 2, the payoff vector $u = (3, 7, 3, 9)$ is a corewise-stable payoff but $(3, 7, 0, 0)$ is not a t-stable payoff, even though $x_{12} = 1$ in the corewise-stable matching.

The set of optimal t-stable payoffs provides a set of solutions for every game. Theorem 1 states that this set coincides with the core, which is equivalent to the set of pairwise-stable payoffs, if and only if the core is not empty.

Theorem 1 *The set of corewise-stable payoffs is non-empty if and only if* $S^* = C$.

Proof. Suppose the core is non-empty. Let (u, x) be an optimal t-stable outcome. We are going to show that (u, x) is corewise-stable. In fact, suppose by way of contradiction, that (u, x) is unstable. Under Proposition 10, there is some corewise-stable outcome (u^*, z) which extends (u, x). Then, $u_j^* \geq u_j$ for all $j \in N$ and $u_j^* > u_j$ for at least one player j. But this contradicts the fact that (u, x) is Pareto optimal in S. Hence, (u, x) is corewise-stable. In the other direction, let (u, x) be a corewise-stable outcome. Then, (u, x) is PO feasible. Since every t-stable outcome is feasible, it follows that there is no t-stable Pareto improvement of (u, x). Given that (u, x) is t-stable, it must be an optimal t-stable outcome. Hence, $S^* = C$.

The proof that $S^* = C$ implies that the core is not-empty is immediate from the fact that $S^* \neq \varnothing$. ■

To emphasize that the set of optimal t-stable outcomes is the natural extension of the set of corewise-stable outcomes when the last set is empty, let us mention that we can use properties for the set of optimal t-stable outcomes, together with Theorem 1, to obtain the corresponding properties of the core as immediate corollaries. In particular, the result that an optimal matching is compatible with any corewise-stable payoff (Proposition 2) is a corollary of Proposition 8 and Theorem 1 for the environments where the core is not empty, once we realize that the quasi-optimal matchings are optimal if the core exists. And Corollary 1 is just a corollary of that result.

Going back to the partnership formation process discussed in the previous section, Theorem 1 ensures that it ends when a core outcome is reached, whenever the core is non-empty. If the final outcome of some sequential process is corewise-stable then the final outcome of every sequential process is corewise-stable. On the other hand, whatever sequence is formed, if the final outcome of some coalition formation process is corewise-unstable then the final outcome of every coalition formation process is corewise-unstable and the core is empty.

We can use the relationship between the core and the set of optimal t-stable outcomes established in Theorem 1 to obtain conditions under which the core exists. First, Theorem 2 uses the properties of the set of optimal t-stable payoffs to provide a necessary and sufficient condition for the core to be non-empty based on the examination of optimal t-stable payoffs.

Theorem 2 *The set of corewise-stable outcomes is non-empty if and only if every optimal t-stable payoff is PO feasible.*

Proof. Suppose first that the set of corewise-stable outcomes is non-empty and let $u \in S^*$. Theorem 1 implies that $u \in C$, so u is PO feasible.

In the other direction, take an optimal t-stable payoff u, which is also PO feasible, and let x be a t-stable matching compatible with u. Then, $\sum_N u_i = \sum_x (u_i + u_j) = \sum_x a_{ij}$. Since u is PO feasible then $v(N) = \sum_N u_i$. Therefore, $\sum_x a_{ij} = v(N)$, so x is an optimal matching. Proposition 4 then implies that $u \in C$, so $C \neq \varnothing$. Hence, the proof is complete. ■

Our final theorem provides a necessary and sufficient condition for the emptiness of the core based on the idea of "non-solvable blocking pairs."

Some of the blocking pairs of a t-stable outcome "vanish" along the partnership formation process that we have described at the end of section 5, in the sense that they do not block some t-stable outcomes that extend the original t-stable outcome. Other blocking pairs "persist" along the process as they block all the t-stable extensions of the original outcome, including the optimal t-stable outcomes that can be obtained in the last term of the sequences. As we will show in Theorem 3, the last type of blocking pairs play a fundamental role in the emptiness of the core. We will call them "non-solvable blocking pairs."

Definition 12 *Let (u, x) be a t-stable outcome and let $\{i, j\} \subseteq U(x)$, with $a_{ij} > 0$ (i.e., $\{i, j\}$ is a blocking pair). We say that* $\boldsymbol{\{i, j\}}$ ***is a non-solvable blocking pair of*** $\boldsymbol{(u, x)}$

if either $u \in S^*$ or $\{i,j\} \subseteq U(x')$ for every t-stable extension (u',x') of (u,x). Also, we say that ***$\{i,j\}$ is a non-solvable blocking pair*** *if it is a non-solvable blocking pair for some t-stable outcome (u,x).*

Therefore, since $a_{ij} > 0$, **if $\{i,j\}$ is a non-solvable blocking pair for (u,x)**, then $\{i,j\}$ blocks every t-stable extension of (u,x), if any. In this case, **$\{i,j\}$ also blocks the optimal t-stable outcome which extends (u,x)**, and that optimal t-stable outcome is corewise-unstable which, under Theorem 1, implies $C = \varnothing$. In fact, **every blocking pair $\{i,j\}$ of an optimal t-stable outcome is a non-solvable blocking pair of all t-stable outcomes**. This is because, under Proposition 9, the pair $\{i,j\}$ blocks every t-stable outcome (including all optimal t-stable outcomes). Thus, $\{i,j\}$ is a non-solvable blocking pair of every t-stable outcome. Hence, **the set of non-solvable blocking pairs of a given t-stable outcome is the same as that of every t-stable outcome and**, in particular, **it coincides with the set of blocking pairs of any optimal t-stable outcome**. These conclusions are formalized in the following results.

Proposition 11 *Let (u,x) be a t-stable outcome and let $\{i,j\}$ be a non-solvable blocking pair for (u,x). Then, $\{i,j\}$ is a non-solvable blocking pair of every t-stable outcome.*

Proof. As stated in Definition 12, $\{i,j\}$ is a blocking pair of the optimal t-stable outcome that extends (u,x), in case $u \notin S^*$. Then, in any case, $\{i,j\}$ is a blocking pair of an optimal t-stable outcome. According to Proposition 9, $\{i,j\}$ is a blocking pair of every t-stable outcome, and is so for every extension of any t-stable outcome. Definition 12 then implies that $\{i,j\}$ is a non-solvable blocking pair of every t-stable outcome. Hence, the proof is complete. ■

Corollary 5 uses Proposition 11 to characterize the non-solvable blocking pairs as the blocking pairs of an optimal t-stable outcome.

Corollary 5 *The pair $\{i,j\}$ is a non-solvable blocking pair if and only if it is a blocking pair of an optimal t-stable outcome.*

Proof. According to Definition 12, the non-solvable blocking pairs of an optimal t-stable outcome are its blocking pairs. On the other hand, under Proposition 11, the set of non-solvable blocking pairs of a given t-stable outcome is the same as that for every t-stable outcome, in particular for every optimal t-stable outcome. Then, the set of non-solvable blocking pairs of a given t-stable outcome coincides with the set of blocking pairs of any optimal t-stable outcome. ■

We can now state the theorem that provides the conditions for the existence of the core.

Theorem 3 *The following conditions are equivalent:*
(i) $C = \varnothing$*;*
(ii) every t-stable outcome has a non-solvable blocking pair;
(iii) there is a t-stable outcome that has a non-solvable blocking pair.

Proof. Suppose $C = \varnothing$. Let (w, y) be an optimal t-stable outcome. Since $C = \varnothing$ we have that (w, y) is corewise-unstable according to Theorem 1. Let $\{i, j\} \subseteq U(y)$ with $a_{ij} > 0$. It follows from Proposition 9 that $\{i, j\}$ blocks every t-stable outcome, in particular it blocks every extension of any t-stable outcome, if any. Then, under Definition 12, $\{i, j\}$ is a non-solvable blocking pair of every t-stable outcome. Then (i) implies (ii). Clearly, (ii) implies (iii).

Now, let (u, x) be a t-stable outcome and suppose $\{i, j\}$ is a non-solvable blocking pair for (u, x). As shown in the proof of Proposition 11, $\{i, j\}$ is a blocking pair of an optimal t-stable outcome. Theorem 1 implies that $C = \varnothing$, so (iii) implies (i). Hence, we have completed the proof. ■

Remark 6 *From the results above we can conclude that* $\{i, j\}$ *is a non-solvable blocking pair for some t-stable outcome if and only if the pair* $\{i, j\}$ *blocks every t-stable outcome. Then, if two t-stable outcomes have disjoint sets of blocking pairs, the core is non-empty.* ■

Our final remark makes clear the extent to which a non-solvable blocking pair is distinct from the other blocking pairs.

Remark 7 *Our previous results imply that if* $\{i, j\}$ *is a non-solvable blocking pair for* (u, x) *then there is no t-stable extension* (w, y) *of* (u, x) *such that* $y(i) = j$*.* ■

7 Concluding remarks

Our paper studies the one-sided assignment game, which is the generalization of the two-sided assignment game of Shapley and Shubik (1972) to the case where any two agents can form a partnership. It provides a new point of view about stability through the concepts of tradewise-stable outcome and the constrained-optimal tradewise-stable outcome.

Tradewise-stable outcomes capture some notion of internal stability: In a tradewise-stable outcome, a matched agent cannot block the situation deviating with either another matched or unmatched agent. In that sense, the set of matched agents (the members of the "club of active agents") is in a stable situation as none of its members can deviate. The properties of the set of tradewise-stable outcomes allow us to propose a dynamic "partnership formation" process. The set of partners enlarges at each step of the process, but the payoff of the old partners does not change with the arrival of new members. At the end of the process, we always obtain an outcome which is constrained-optimal in the set of t-stable outcomes. The process suggests that when there is no stable allocation, the constrained-optimal t-stable allocation is the most stable outcome that we can expect.

We view the set of constrained-optimal tradewise-stable outcomes as a natural solution concept for the one-sided assignment game. Each of them generates (and they are the only ones that do so) the highest possible total surplus in the set of t-stable outcomes. And, as the previous dynamic process suggests, these outcomes are "as stable as possible," in the sense that any matching involving a larger set of matched agents will necessarily be unstable; the club of active agents would be too large. In fact, the set of constrained-optimal tradewise-stable payoffs coincides with the core when the core is not empty. Thus, the solution concept keeps all the good properties of the core when it exists, but it also provides a prediction for those markets where the core does not exist. Moreover, several of the nice properties of the core, when it is non-empty, are extended to the set of Pareto-optimal tradewise-stable outcomes.

Bondareva (1963) and Shapley (1967) proved that the core of a transferable utility game is non-empty if and only if the game is balanced. Thus, for the game considered here, the condition that every optimal tradewise-stable payoff is Pareto-optimal feasible is equivalent to balancedness. This suggests the question of whether this equivalence persists in all transferable-utility (TU) games. The answer to this question is not easy. Our results strongly rely on the existence of a feasible matching underlying every feasible outcome. However, players do not necessarily form partnerships in the general TU game. On the other hand, the intuition behind a tradewise-stable outcome is not related to a matching and seems to be quite general: if all "interactions" are made under the premise of optimal behavior, a tradewise-stable outcome results. In a subsequent work, Pérez-Castrillo and Sotomayor (2019) consider the extension of the present investigation to the coalitional games with transfers. This extension is not straightforward because our concepts make use of the fact that every feasible allocation is compatible with a feasible matching. The solution of this problem is possible by the identification, for every feasible allocation, of some convenient coalitional structure that, restricted to the one-sided matching model, coincides with a feasible matching.

8 Appendix

Lemma 1 *Let (u, x) a t-stable outcome. Then, $v(T(x)) + v(U(x)) = v(N)$.*

Proof. Let y be an optimal matching. Then, $v(N) = \sum_y a_{ij}$. Set

$$\alpha \equiv \{\{i,j\} \in y; \{i,j\} \cap T(x) \neq \varnothing \text{ and } \{i,j\} \cap U(x) \neq \varnothing\},$$
$$\beta \equiv \{\{i,j\} \in y; \{i,j\} \subseteq T(x)\} \text{ and}$$
$$\gamma \equiv \{\{i,j\} \in y; \{i,j\} \cap T(x) = \varnothing\}.$$

Also, denote

$$\alpha_x \equiv \{i \in T(x); \{i,j\} \in \alpha \text{ for some } j\}$$

and

$$R \text{ is a set of pairs } \{i,j\} \subseteq U(x) \text{ such that } v(U(x)) \equiv \sum_R a_{ij},$$
$$R' \equiv U(x) \backslash \cup_\gamma \{i,j\} \text{ and}$$
$$R'' \text{ is a set of pairs } \{i,j\} \subseteq T(x) \text{ such that } v(T(x)) \equiv \sum_{R''} a_{ij}.$$

Then, $\sum_{T(x)} u_i = \sum_{\alpha_x} u_i + \sum_\beta a_{ij} \geq \sum_\alpha a_{ij} + \sum_\beta a_{ij}$, where the inequality is due to the fact that $(u, x) \in S$, $i \in T(x)$, and $u_{x(i)} = 0$ for all $i \in \alpha_x$, and so $u_i = u_i + u_{x(i)} \geq a_{ix(i)}$ for all $i \in \alpha_x$.

Also, $R' \cup \gamma$ is a partition of $U(x)$, so $v(U(x)) = \sum_R a_{ij} \geq \sum_\gamma a_{ij} + \sum_{R'} a_{ii} = \sum_\gamma a_{ij}$. Then,

$$v(N) = \sum_y a_{ij} = \left(\sum_\alpha a_{ij} + \sum_\beta a_{ij} \right) + \sum_\gamma a_{ij} \leq \sum_{T(x)} u_i + v(U(x)) = v(T(x)) + v(U(x)),$$

where in the last equality it was used that $T(x)$ is a stable coalition. Therefore,

$$v(N) \leq v(T(x)) + v(U(x)). \tag{14}$$

On the other hand, y is an optimal matching, so $v(N) = \sum_y a_{ij} \geq \sum_{R \cup R''} a_{ij} = v(T(x)) + v(U(x))$. Then,

$$v(N) \geq v(T(x)) + v(U(x)). \tag{15}$$

Hence, $v(T(x)) + v(U(x)) = v(N)$ and the proof is complete. ■

Lemma 2 *The set of t-stable payoffs S is a compact set of $\mathbb{R}^n$.*

Proof. The set S is bounded because $0 \leq u_j \leq v(N)$ for all $j \in N$ and for all t-stable payoffs u. To prove that it is also closed, take any sequence $(u^t)_{t=1,2,...}$ of t-stable payoffs, with $u^t \to u$ when t tends to infinity. Since the set of matchings is finite, there is some matching x which is compatible with infinitely many terms of the sequence $(u^t)_{t=1,2,...}$. Denote $(v^t)_{t=1,2,...}$ this subsequence. Then, if $x(j) = k$, $u_j + u_k = lim_{t\to\infty}(v_j^t + v_k^t) = lim_{t\to\infty}a_{jk} = a_{jk}$. Similarly, if $x(j) = j$ then $u_j = lim_{t\to\infty}v_j^t = 0$. Thus, x is compatible with u, so (u, x) is feasible. We claim that if j is matched at x then j is not part of a blocking pair of (u, x). In fact, $u_j + u_k = lim_{t\to\infty}(v_j^t + v_k^t) \geq lim_{t\to\infty}a_{jk} = a_{jk}$ for any $k \in N\backslash\{j\}$, where the inequality holds because (v^t, x) is a t-stable outcome for all t. Therefore, (u, x) is a t-stable outcome, so u is a t-stable payoff. Hence, the set of t-stable payoffs is bounded and closed, so it is compact. ■

Next result is a *Decomposition Lemma* for the set of t-stable outcomes which has similarities with other decomposition lemmas in matching models (see, for instance, Gale and Sotomayor, 1985). The lemma states that, for any two t-stable outcomes, a player who is matched at both outcomes and obtains a higher payoff in the first is necessarily matched, at both outcomes, to a player who obtains a higher payoff in the second.

Lemma 3 *Let (u, x) and (w, y) be t-stable outcomes. Let $M_u \equiv \{j \in T(y); u_j > w_j\}$ and $M_w \equiv \{j \in T(x); w_j > u_j\}$. Then $x(M_u) = y(M_u) = M_w$ and $x(M_w) = y(M_w) = M_u$.*[11]

Proof. We first prove that $x(M_u) \subseteq M_w$. Take $j \in M_u$; then j is matched under x since $u_j > w_j \geq 0$. We show by contradiction that $k \equiv x(j)$ is in M_w. Suppose $k \notin M_w$, then

$$a_{jk} = u_j + u_k > w_j + w_k$$

which implies that (j, k) blocks (w, y). However, $j \in M_u$ so it is matched at y, which contradicts that (w, y) is t-stable.

A similar argument leads to $y(M_w) \subseteq Mu$.

[11] The decomposition lemma applies, in particular, to core outcomes. Then, an immediate consequence of the lemma is a polarization of interests between the partners along the core: If (u, x) and (w, y) are corewise-stable outcomes, j is matched to k under x or under y, and $u_j > w_j$, then $w_k > u_k$. This is because both payoffs are compatible with the same optimal matching; therefore, if j is matched to k under (u, x) then j is also matched to k under (w, x), so Lemma 3 applies.

Moreover, $x(M_u) \subseteq M_w$ implies $M_u \subseteq x(M_w)$ and $y(M_w) \subseteq M_u$ implies $M_w \subseteq y(M_u)$. Since all the players in M_u and in M_w are matched at x and y, it follows that $|M_u| = |x(M_u)|$, $|M_w| = |y(M_w)|$, $|y(M_u)| = |M_u|$ and $|y(M_u)| = |M_u|$. Therefore,

$$|M_u| = |x(M_u)| \leq |M_w| = |y(M_w)| \leq |M_u|$$

and

$$|M_w| \leq |y(M_u)| = |M_u| \leq |x(M_w)| = |M_w|,$$

which imply $x(M_u) = M_w$, $y(M_w) = M_u$, $y(M_u) = M_w$, and $x(M_w) = M_u$. ■

We notice that, in the proof of Lemma 3, it is shown that $u_j > 0$ for all $j \in M_u$ and $w_j > 0$ for all $j \in M_w$. Therefore, we can write $M_u = \{j \in T(x) \cap T(y); u_j > w_j\}$ and $M_w \equiv \{j \in T(x) \cap T(y); w_j > u_j\}$.

Lemma 4 *Let A be a non-empty and compact set of $\mathbb{R}^n$, ordered with the partial order relation $\geq$ induced by $\mathbb{R}^n$. Then, the set of maximal elements of A with respect to $\geq$ is a non-empty and compact set of $\mathbb{R}^n$.*

Proof. Denote $A^* \equiv \{u \in A; u$ is a maximal element of $A\}$. It is known that every non-empty, compact and partially ordered set has a maximal element, so $A^* \neq \varnothing$. The set A^* is clearly bounded, since A is bounded. To see that A^* is closed, take any sequence of vectors $(u^t)_{t=1,2,\ldots}$, with $u^t \in A^*$ for all t, which converges to some vector u. Suppose, by way of contradiction, that $u \notin A^*$. Then, there exists some vector $w \in A$ such that $w > u$. If this is the case, there is some neighborhood V of the vector u and some integer k such that $u^t \in V$ for all $t \geq k$ and $w > u'$ for all $u' \in V$. In particular, $w > u^k$, which contradicts the assumption that $u^k \in A^*$. Hence, A^* is a compact set of $\mathbb{R}^n$. ■

Lemma 5 *The set of optimal t-stable payoffs S^* is a non-empty and compact set of $\mathbb{R}^n$.*

Proof. According to Lemma 2, S is compact and non-empty. Moreover, S is an ordered set by the partial order relation $\geq$ induced by $\mathbb{R}^n$. Then, Lemma 4 applies and so S^*, the set of maximal elements of S, is a non-empty and compact set of $\mathbb{R}^n$. ■

Lemma 6 *Let (u, x) and (w, y) be optimal t-stable outcomes. Let $j^* \in T(x) \backslash T(y)$ with $a_{j^* x(j^*)} > 0$. Then $x(j^*) \in T(y)$.*

Proof. Suppose, by way of contradiction, that $x(j^*) \in T(x) \backslash T(y)$. Denote $A \equiv \{t \in T(x) \backslash T(y); x(t) \in T(x) \backslash T(y)\}$. We have that $j^* \in A$, so $A \neq \varnothing$.

We first show that

$$w_{q^*} + u_{t^*} = u_{t^*} < a_{q*t^*} \text{ for some } t^* \in A \text{ and } q^* \in N \backslash (T(y) \cup A). \tag{16}$$

Suppose, by way of contradiction, that there are no such players t^* and q^*. Then, either $N\backslash(T(y)\cup A) = \varnothing$ or (given that $w_q = 0$ for all $q \notin T(y)$) $u_t \geq a_{qt}$ for all $q \in N\backslash(T(y) \cup A)$ and $t \in A$. In any case, we can construct the outcome (w', y') as follows: the matching y' agrees with x on A and it agrees with y on $N\backslash A$, hence, $T(w') = T(y) \cup A$; the payoff vector satisfies $w'_j = w_j$ for all $j \in T(y)$, $w'_j = u_j$ for all $j \in A$ and $w'_j = 0$ for all $j \in N\backslash(T(y) \cup A)$. Since (u, x) is t-stable, there is no pair blocking (w', y') among the agents of A. Because (w, y) is t-stable, there is no blocking pair formed by two agents in $T(y)$ or an agent in A and an agent in $T(y)$. Finally, if $N\backslash(T(y) \cup A) \neq \varnothing$ then first, no blocking pair exists between an agent in $T(y)$ and an agent in $N\backslash(T(y)\cup A)$ because w' coincides with w for these agents and (w, y) is t-stable and second, by using the contradiction assumption, $0 + w'_t = u_t \geq a_{qt}$ for all $t \in A$ and for all $q \in N\backslash(T(y) \cup A)$. Therefore, the outcome (w', y') is t-stable. Since $a_{j^*x(j^*)} > 0$, it follows that either $u_{j^*} > 0$ or $u_{x(j^*)} > 0$. Hence, the outcome (w', y') is a t-stable extension of (w, y), which is a contradiction because (w, y) is an optimal t-stable outcome.

Once we have shown that there exist some $t^* \in A$ and $q^* \in N\backslash(T(y) \cup A)$ such that $w_{q^*} + u_{t^*} = u_{t^*} < a_{q^*t^*}$, we claim that such q^* necessarily satisfies $q^* \in T(x)$ and $u_{q^*} > 0$. Otherwise, $u_{q^*} = 0$, in which case $u_{q^*} + u_{t^*} < a_{q^*t^*}$ by (16), and then $\{q^*, t^*\}$ would block (u, x), which is not possible because $t^* \in T(x)$ and (u, x) is a t-stable outcome. Therefore, $q^* \in T(x)\backslash(T(y) \cup A)$. Since $q^* \notin A$, we must have that $p \equiv x(q^*) \in T(x) \cap T(y)$, so

$$u_p = w_p \tag{17}$$

because otherwise we should have that $q^* \in T(x) \cap T(y)$ according to Lemma 3, which would be a contradiction. Furthermore, the fact that $u_{q^*} > 0$ implies that $u_p < a_{pq^*}$, and so $w_p < a_{pq^*}$ by (17), which implies that $\{q^*, p\}$ blocks (w, y) (because $w_{q^*} = 0$), which is a contradiction since $p \in T(y)$. Hence, $x(j^*) \in T(x) \cap T(y)$, and the proof is complete. ■

Lemma 7 *Let (u, x) and (w, y) be optimal t-stable outcomes. Let $j^* \in T(x)\backslash T(y)$. Then $u_{j^*} = w_{j^*} = 0$ and $u_{x(j^*)} = w_{x(j^*)}$.*

Proof. Denote $k^* \equiv x(j^*)$. If $a_{j^*k^*} = 0$, then $u_{j^*} = w_{j^*} = 0$ and $u_{k^*} = 0$. If $k^* \in T(x)\backslash T(y)$ then $w_{k^*} = 0$. Otherwise, we cannot have that $u_{k^*} \neq w_{k^*}$, because Lemma 3 would imply that $j^* \in T(x) \cap T(y)$, which would be a contradiction. Therefore, it is always the case that $u_{x(j^*)} = w_{x(j^*)}$.

Suppose now that $a_{j^*k^*} > 0$. Under Lemma 6, $k^* \in T(x) \cap T(y)$. Then, since $j^* \notin T(y)$ we have that $w_{j^*} = 0$. In addition, we cannot have that $u_{k^*} \neq w_{k^*}$, according to Lemma 3 and the assumption that $j^* \in T(x) \backslash T(y)$, so $u_{k^*} = w_{k^*}$. Now, suppose by way of contradiction that $u_{j^*} > 0$. Then, $u_{k^*} < a_{j^*k^*}$. Therefore, $w_{j^*} + w_{k^*} = u_{k^*} < a_{j^*k^*}$, so $\{j^*, k^*\}$ blocks (w, y), which is a contradiction because $k^* \in T(y)$ and (w, y) is a t-stable outcome. Hence, $u_{j^*} = w_{j^*} = 0$ and $u_{x(j^*)} = w_{x(j^*)}$, and the proof is complete. ■

References

[1] Andersson, T., Gudmundsson, J., Talman, D. and Yang, Z. (2014) "A competitive partnership formation process," *Games and Economic Behavior* 86, 165–177.

[2] Bondareva, O. N. (1963) "Some applications of linear programming methods to the theory of cooperative games," *Problemy Kybernetiki* 10, 119–139 [in Russian].

[3] Chiappori, P.-A., Galichon, A. and Salanié, B. (2014) "The roommate problem is more stable than you think," CESifo WP 4676.

[4] Demange, G. and Gale, D. (1985) "The strategy structure of two-sided matching markets," *Econometrica* 53, 873–888.

[5] Eriksson, K. and Karlander, J. (2000) "Corewise-stable matching in a common generalization of the marriage and assignment models," *Discrete Mathematics* 217, 135–156.

[6] Gale, D. and Shapley, L. (1962) "College admissions and the stability of marriage," *American Mathematical Monthly* 69, 9–15.

[7] Gale, D. and Sotomayor M. (1985) "Some remarks on the corewise-stable matching problem," *Discrete Applied Mathematics* 11, 223–32.

[8] Klaus, B. and Nichifor, A. (2010) "Consistency in one-sided assignment problems," *Social Choice and Welfare* 35, 415–433.

[9] Klaus, B. and Payot, F. (2015) "Paths to stability in the assignment problem," *Journal of Dynamics & Games*, 2, 257-287.

[10] Koopmans, T.C., and M. Beckmann (1957) "Assignment problems and the location of economic activities," *Econometrica* 25, 53–76.

[11] Pérez-Castrillo, D. and Sotomayor, M. (2019) "Constrained-optimal tradewise-stable outcomes in coalitional games with transferable payoff: a solution concept weaker than the core," mimeo.

[12] Roth, A. and Sotomayor, M (1990) "*Two-sided matching. A study in game-theoretic modeling and analysis*," Econometric Society Monographs, 2nd ed., Vol. 18, Cambridge University Press.

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

[14] Shapley, L, S. (1967) "On balanced sets and cores," *Naval Research Logistics Quarterly* 14, 453–460.

[15] Shapley, L. and Scarf, H. (1974) "On cores and indivisibility," *Journal of Mathematical Economics* 1, 23–8.

[16] Shapley, L. and Shubik, M. (1972) "The assignment game I: The core," *International Journal of Game Theory* 1, 111–130.

[17] Sotomayor, M. (1996) "A non-constructive elementary proof of the existence of corewise-stable marriages," *Games and Economic Behavior* 13, 135–137.

[18] ______ (1999) "Three remarks on the many-to-many stable matching problem," *Mathematical Social Sciences* 38, 55–70.

[19] ______ (2000) "Existence of stable outcomes and the lattice property for a unified matching market," *Mathematical Social Science* 39 , 119–132.

[20] ______ (2004) "Implementation in the many-to-many matching market," *Games and Economic Behavior* 46, 199–212.

[21] ______ (2005a) "On the core of the one-sided assignment game," WP Universidade de São Paulo.

[22] ______ (2005b) "An elementary non-constructive proof of the non-emptiness of the core of the housing market of Shapley and Scarf," *Mathematical Social Sciences* 50, 298–303.

[23] ______ (2009a) "The role played by simple outcomes in coalition formation process of core outcomes,", WP Universidade de São Paulo.

[24] ______ (2009b) "Coalitional interactions leading to new solution concepts in the time-sharing assignment game," mimeo.

[25] ______ (2011) "The Pareto-stability concept is a natural solution concept for discrete matching markets with indifferences," *International Journal of Game Theory* 40, 631–644.

[26] ______ (2018) "Simple allocations in the assignment game," WP Universidade de São Paulo.

[27] ______ (2019) "Matching with stable trades," mimeo.

[28] Talman, D. and Yang, Z. (2011) "A model of partnership formation," *Journal of Mathematical Economics* 47, 206–212.

[29] Wu, Q. and Roth, A. (2018) "The lattice of envy-free matchings," *Games and Economic Behavior* 109, 201–211.