Characterizing the top trading cycles rule for housing markets with lexicographic preferences when externalities are limited

Bettina Klaus*

February 13, 2021

Abstract

We consider a variation of the housing market model à la Shapley and Scarf (1974) where agents care both about their own consumption via demand preferences and about the agent who receives their endowment via supply preferences (see Klaus and Meo, 2021). Then, if preferences are either all demand lexicographic or all supply lexicographic, we characterize the corresponding top trading cycles rule by *individual rationality*, *Pareto optimality*, and *strategy-proofness*. Since on the lexicographic preference domains the strong core can be multi-valued, our result sheds light on the fact that the properties that also characterized the strong core rule for Shapley-Scarf housing markets (Ma, 1994) characterize the top trading cycles rule and *not* the strong core rule (or correspondence).

Keywords: Core; characterization; externalities; housing markets; top trading cycles rule.

JEL codes: C70, C71, C78, D62, D64.

1 Introduction

In classical *Shapley-Scarf housing markets* each agent is endowed with an indivisible commodity, for instance a house, and wishes to consume exactly one commodity. Agents have complete, reflexive, and transitive preferences over all existing houses and may be better

*Faculty of Business and Economics, University of Lausanne, Internef 538, 1015 Lausanne, Switzerland; *e-mail*: `Bettina.Klaus@unil.ch`. I gratefully acknowledge financial support from the Swiss National Science Foundation (SNSF) through Project 100018_192583. I thank Di Feng, Claudia Meo, Seckin Özbilen, and William Thomson for their valuable comments.

off by trading houses: exchanges do not involve monetary compensations. An outcome for a Shapley-Scarf housing market is a permutation of the endowment allocation.

One of the best known solution concepts for barter economies is the *weak core*, based on the absence of coalitions that may reallocate their endowments among themselves and make all their members strictly better off (i.e., no coalition can *strongly block* a core allocation). The weak core for Shapley-Scarf housing markets is always nonempty (Shapley and Scarf, 1974). If strong blocking is weakened to only require that members of a blocking coalition are not worse off while at least one member is better off, then a stronger solution, the *strong core*, results (i.e., no coalition can *weakly block* a strong core allocation). In contrast to the weak core, the strong core for Shapley-Scarf housing markets can be empty, unless no agent is indifferent between any of houses. Hence, when preferences are strict, also the strong core is nonempty and, in fact, coincides with the unique competitive allocation (Roth and Postlewaite, 1977). Using the so-called top trading cycles (TTC) algorithm (due to David Gale, see Shapley and Scarf, 1974), one can easily determine the unique strong core allocation for any Shapley-Scarf housing market with strict preferences. Then, the solution that always assigns the strong core and the rule that always assigns the TTC allocation are essentially the same.[1]

After the above mentioned seminal papers, a number of studies have analyzed Shapley-Scarf housing markets with strict preferences from a mechanism design perspective. For housing markets with strict preferences, Roth (1982) proved that the rule that assigns the unique strong core allocation is *strategy-proof*, i.e., no agent can benefit from misrepresenting his preferences. Subsequently, Ma (1994) demonstrated that this "strong core rule" is the unique rule satisfying *Pareto optimality*, *strategy-proofness*, and *individual rationality*, i.e., no agent is worse off after trading. For Shapley-Scarf housing markets with strict preferences, the strong core rule equals the top trading cycles rule. Therefore, in Ma's characterization it is not clear whether the properties characterize the strong core solution or the TTC rule. We demonstrate for a slight model variation, housing markets with lexicographic preferences, that the properties *individual rationality*, *Pareto optimality*, and *strategy-proofness*, characterize the TTC rule, which differs from the strong core solution.

Klaus and Meo (2021) built on Shapley and Scarf's classical model by assuming that agents care not only about the object they receive but also about the agent who received their endowment: agents have traditional "demand preferences" as well as less traditional "supply preferences". This form of externality is modelled, for each agent, by a preference relation defined over pairs formed by the object assigned to the agent himself

[1] In our nomenclature, a solution allocates a set of allocations while a rule assigns a single allocation.

and the recipient of his own object. An example would be that of kidney exchange, the agents being formed by recipient-donor pairs and the objects being the kidneys that will be donated. It is clear that each agent (recipient) cares about the kidney he will receive but in addition, each agent (donor) might also care about the recipient of his kidney. More broadly, models with this type of limited externalities also fit well with exchanges that are not permanent, i.e., where the endowments are only temporarily exchanged and eventually return to their original owners: vacation home exchanges such as InterVac or ThirdHome are examples of such temporary exchanges. Aziz and Lee (2020) have introduced the same problem as "temporary exchange problem". General forms of externalities for Shapley-Scarf housing markets have been analyzed before and we refer to Klaus and Meo (2021) where the papers of Mumcu and Sağlam (2007) , Hong and Park (2020), and Graziano et al. (2020) are discussed in more detail.

The main focus of this paper is on markets where all agents have demand lexicographic preferences,[2] i.e., agents care first about the object they receive before considering who receives their endowment. Then, Klaus and Meo (2021) showed that the strong core is nonempty and possibly multi-valued. Here, we prove how a result similar to Ma (1994) holds for our model: the so-called top trading cycles (TTC) rule is the unique rule satisfying *individual rationality*, *Pareto optimality*, and *strategy-proofness*. Aziz and Lee (2020) show that the TTC rule satisfies *individual rationality*, *Pareto optimality*, and *strategy-proofness*. Our characterization of the TTC rule (Theorem 1) complements this result by showing that it is the only rule satisfying these properties. However, apart from providing a new TTC characterization result, we would like to point out that with the help of this result we can answer the fundamental question whether the properties *individual rationality*, *Pareto optimality*, and *strategy-proofness* characterize the strong core solution or the TTC rule. Since for housing markets with demand lexicographic preferences the strong core can be multi-valued, the properties clearly characterize the TTC rule.

2 The model

We consider an exchange market with indivisibilities formed by n agents and by the same number of indivisible objects, say houses; let $\boldsymbol{N} = \{1, \ldots, n\}$ and $\boldsymbol{H} = \{h_1, \ldots, h_n\}$ denote the **set of agents** and **houses**, respectively. Each agent owns one distinct house when entering the market, desires exactly one house, and has the option to trade the initially owned house in order to get a better one. All trades are made with no transfer of money. We assume that **agent $\boldsymbol{i}$ owns house $\boldsymbol{h_i}$**.

[2]Alternatively, one could assume that all agents have supply lexicographic preferences.

An **allocation** $\boldsymbol{a}$ is an assignment of houses to agents such that each agent receives exactly one house, that is, a bijection $a : N \to H$. Alternatively, we will denote an allocation a as a vector $a = (a_1, \dots, a_n)$ with $a_i \in H$ denoting the house assigned to agent $i \in N$ under allocation a. $\boldsymbol{\mathcal{A}}$ denotes the **set of all allocations** and $\boldsymbol{h} = (h_1, \dots, h_n)$ the **endowment allocation**. Hence, the set of allocations $\mathcal{A}$ is obtained by all permutations of H. A nonempty subset S of N is called a **coalition**. For any coalition $S \subseteq N$ and any allocation $a \in \mathcal{A}$, let $\boldsymbol{a(S)} = \{a_i \in H : i \in S\}$ be the **set of houses that coalition $\boldsymbol{S}$ receives at allocation $\boldsymbol{a}$.**

Up to now we have followed the description of a classical *Shapley-Scarf housing market model* as introduced by Shapley and Scarf (1974). Now, in contrast with that model, we assume that each agent cares not only about the house he receives but also about the recipient of his own house.

Given an allocation $a \in \mathcal{A}$, the **allotment of agent $\boldsymbol{i}$** is the pair $\boldsymbol{(a(i), a^{-1}(h_i))} \in H \times N$, formed by the house $a(i)$ assigned to agent i and the agent who receives agent i's house, i.e., agent $a^{-1}(h_i)$. Note that $a(i) = h(i)$ if and only if $a^{-1}(h_i) = i$, i.e., either both elements of agent i's endowment allotment (h_i, i) occur in his allotment or none. $\boldsymbol{\mathcal{A}_i}$ denotes the **set of all the allotments of agent $\boldsymbol{i}$**.

Each agent $i \in N$ has a preference relation $\succeq_i$ over the set $\boldsymbol{\mathcal{A}_i}$, that is, $\succeq_i$ is a *transitive*, *reflexive*, and *complete* binary relation. As usual, $\succ_i$ and $\sim_i$ denote the asymmetric and symmetric parts of $\succeq_i$, respectively. Here we assume that agents' preferences over allotments are induced lexicographically by their preferences over the houses and the other agents in the market. More specifically, we assume that each agent $i \in N$ has

a "**demand**" strict preference relation $\boldsymbol{\succeq_i^d}$ over the set H of houses and

a "**supply**" strict preference relation $\boldsymbol{\succeq_i^s}$ over the set N of agents.

We say that **house $\boldsymbol{h}$** $\in H \setminus \{h_i\}$ is **acceptable** for agent $i \in N$ if $h \succ_i^d h_i$, otherwise it is **unacceptable**. Symmetrically, we say that **agent $\boldsymbol{j}$** $\in N \setminus \{i\}$ is **acceptable** for agent $i \in N$ if $j \succ_i^s i$, otherwise he is **unacceptable**.

We denote the set of demand preferences over H and the set of demand preference profiles by $\boldsymbol{\mathcal{D}_{\mathrm{d}}}$ and $\boldsymbol{\mathcal{D}_{\mathrm{d}}^N}$, respectively; and the set of supply preferences over N and the set of supply preference profiles by $\boldsymbol{\mathcal{D}_{\mathrm{s}}}$ and $\boldsymbol{\mathcal{D}_{\mathrm{s}}^N}$, respectively. We consider the following domain of strict preferences over allotments.

The domain $\boldsymbol{\mathcal{D}_{\mathrm{dlex}}}$ of **demand lexicographic preferences**: an agent $i \in N$ has demand lexicographic preferences $\succeq_i$ if he primarily cares about the house he receives and only secondarily about who receives his house. Formally, $\succeq_i \in \mathcal{D}_{\mathrm{dlex}}$ if for any $(h, j), (h', k) \in \mathcal{A}_i$,

$$(h,j) \succ_i (h',k) \text{ if and only if } h \succ_i^d h' \text{ or } [h = h' \text{ and } j \succ_i^s k].$$

The domain $\boldsymbol{\mathcal{D}_{\text{slex}}}$ of **supply lexicographic preferences**: an agent $i \in N$ has supply lexicographic preferences $\succeq_i$ if he primarily cares about who receives his house and only secondarily about the house he receives. Formally, $\succeq_i \in \mathcal{D}_{\text{slex}}$ if for any $(h,j),(h',k) \in \mathcal{A}_i$,

$$(h,j) \succ_i (h',k) \text{ if and only if } j \succ_i^s k \text{ or } [j = k \text{ and } h \succ_i^d h'].$$

A **housing market with lexicographic preferences**, or **market** for short, is now completely described by the triplet $\boldsymbol{(N, h, \succeq)}$, where N is the set of agents, h is the endowment allocation, and $\succeq \in \mathcal{D}_{\text{dlex}}^N \cup \mathcal{D}_{\text{slex}}^N$ is a preference profile. Since the set of agents and the endowment allocation are fixed, we denote a **market** by its **preference profile** $\boldsymbol{\succeq}$.

For each agent $i \in N$, a preference relation $\trianglerighteq_i$ on the set of allocations $\mathcal{A}$ can be associated with his preferences $\succeq_i$ over $\mathcal{A}_i$. Consider two allocations $a, b \in \mathcal{A}$. Then, we have

$$\begin{aligned} a \vartriangleright_i b \quad &\text{if and only if} \quad (a(i), a^{-1}(h_i)) \succ_i (b(i), b^{-1}(h_i)) \text{ and} \\ a \sim_i b \quad &\text{if and only if} \quad (a(i), a^{-1}(h_i)) = (b(i), b^{-1}(h_i)). \end{aligned}$$

Hong and Park (2020) consider two preference domains over allocations, the domain of *hedonic* and the domain of *order preserving with respect to own-allotments* preferences. Lexicographic preferences in $\mathcal{D}_{\text{dlex}} \cup \mathcal{D}_{\text{slex}}$ are hedonic and demand lexicographic preferences in $\mathcal{D}_{\text{dlex}}$ are order preserving with respect to own-allotments. However, supply lexicographic preferences in $\mathcal{D}_{\text{slex}}$ do not induce preferences over allocations that can be compared with order preserving preferences with respect to own-allotments. See Klaus and Meo (2021, Remark 1) for a more detailed discussion.

We next introduce the strong core, a solution concept that represents the idea of "stable exchange" based on the absence of coalitions that can improve their allotments by reallocating their endowments among themselves.

Definition 1 (**Strong core allocations**).
Let $\succeq \in \mathcal{D}_{\text{dlex}}^N$ and $a \in \mathcal{A}$. Then, **coalition $\boldsymbol{S}$ weakly blocks allocation $\boldsymbol{a}$** if there exists an allocation $b \in \mathcal{A}$ such that

(a) at allocation b agents in S reallocate their endowments, i.e., $b(S) = h(S)$, and

(b) all agents in S are weakly better off with at least one of them being strictly better off, i.e., for all agents $i \in S$,

$$(b_i, b^{-1}(h_i)) \succeq_i ((a_i, a^{-1}(h_i))$$

and for some agent $j \in S$,

$$(b_j, b^{-1}(h_j)) \succ_j ((a_j, a^{-1}(h_j)).$$

Allocation $\boldsymbol{a}$ is a strong core allocation if it is not weakly blocked by any coalition. We denote the **set of strong core allocations** for market $\succeq$ by $\boldsymbol{SC(\succeq)}$.

Klaus and Meo (2021, Proposition 4) (see also Aziz and Lee, 2020) showed that for housing market with lexicographic preferences the strong core is always non-empty, i.e., for each $\succeq \in \mathcal{D}^N_{\text{dlex}} \cup \mathcal{D}^N_{\text{slex}}$, $SC(\succeq) \neq \emptyset$.

Next, a **rule $\boldsymbol{\varphi}$** is a function that associates with each market $(N, h, \succeq)$ an allocation $\varphi(N, h, \succeq) \in \mathcal{A}$.

The first property of a rule we introduce is the well-known condition of *Pareto-optimality*.

Definition 2 (**Pareto optimality**). An allocation $a \in \mathcal{A}$ is **Pareto dominated** by allocation $b \in \mathcal{A}$ if for all agents $i \in N$, $(b_i, b^{-1}(h_i)) \succeq_i ((a_i, a^{-1}(h_i))$ and for some agent $j \in N$, $(b_j, b^{-1}(h_j)) \succ_j ((a_j, a^{-1}(h_j))$. An allocation $a \in \mathcal{A}$ is **Pareto optimal** if it is not Pareto dominated by another allocation. A rule φ is **Pareto optimal** if it only assigns *Pareto-optimal* allocations.

Next, we introduce a voluntary participation conditions based on the idea that no agent can be forced to accept an allotment that is worse than his endowment allotment.

Definition 3 (**Individual rationality**). An allocation $a \in \mathcal{A}$ is **individually rational** if for all agents $i \in N$, $(a_i, a^{-1}(h_i)) \succeq_i (h_i, i)$. A rule φ is **individually rational** if it only assigns *individually rational* allocations.

The well-known non-manipulability property *strategy-proofness* requires that no agent can ever benefit from misrepresenting his preferences.

Definition 4 (**Strategy-proofness**). A rule φ is **strategy-proof** if for each market $(N, h, \succeq)$, each agent $i \in N$, and each preference relation $\widetilde{\succeq}_i$,

$$\varphi_i(\succeq) \succeq_i \varphi_i(\widetilde{\succeq}_i, \succeq_{-i}).^3$$

[3] Following standard notation, $(\widetilde{\succeq}_i, \succeq_{-i})$ is the preference profile that is obtained from $\succeq$ when agent i changes his preferences from $\succeq_i$ to $\widetilde{\succeq}_i$.

3 A characterization of the top trading cycles rule

Ma (1994) demonstrated that the allocation rule that always assigns the strong core allocation is the unique rule satisfying *individual rationality*, *Pareto optimality*, and *strategy-proofness*. We show that a similar result holds in our context.

Since we subsequently could symmetrically switch the roles of agents and houses, all results and examples obtained for demand lexicographic markets can be transcribed into corresponding results for supply lexicographic markets.

Consider a housing market $(N, h, \succeq)$, $\succeq \in \mathcal{D}^N_{\text{dlex}}$, and its associated Shapley-Scarf market $(N, h, \succeq^d)$. We then define the **top trading cycles (TTC) allocation for $\boldsymbol{\succeq^d}$** using Gale's **top trading cycles (TTC) algorithm** (Shapley and Scarf, 1974, attributed the TTC algorithm to David Gale) as follows:

Input. A Shapley-Scarf housing market $(N, h, \succeq^d)$.

Step 1. Let $N_1 := N$ and $H_1 := H$. We construct a directed graph with the set of nodes $N_1 \cup H_1$. For each agent $i \in N_1$ we add a directed edge to his most preferred house in H_1. For each directed edge (i, h) we say that agent i points to house h. For each house $h \in H_1$ we add a directed edge to its owner.

A **trading cycle** is a directed cycle in the graph. Given the finite number of nodes, at least one trading cycle exists. We assign to each agent in a trading cycle the house he points to and remove all trading cycle agents and houses. We define N_2 to be the set of remaining agents and H_2 to be the set of remaining houses and, if $N_2 \neq \emptyset$, we continue with Step 2. Otherwise we stop.

In general at Step t we have the following:

Step $\boldsymbol{t}$. We construct a directed graph with the set of nodes $N_t \cup H_t$ where $N_t \subseteq N$ is the set of agents that remain after Step $t-1$ and $H_t \subseteq H$ is the set of houses that remain after Step $t-1$. For each agent $i \in N_t$ we add a directed edge to his most preferred house in H_t. For each house $h \in H_t$ we add a directed edge to its owner.

At least one trading cycle exists and we assign to each agent in a trading cycle the house he points to and remove all trading cycle agents and houses. We define N_{t+1} to be the set of remaining agents and H_{t+1} to be the set of remaining houses and, if $N_{t+1} \neq \emptyset$, we continue with Step $t+1$. Otherwise we stop.

Output. The TTC algorithm terminates when each agent in N is assigned a house in H (it takes at most $|N|$ steps). We denote the house in H that agent $i \in N$ obtains in the TTC algorithm by $\text{TTC}_i(\succeq^d)$ and the final allocation by $\text{TTC}(\succeq^d)$.

The **TTC rule** assigns to each market $(N, h, \succeq)$ $[\succeq \in \mathcal{D}^N_{\text{dlex}}]$ with associated Shapley-Scarf market $(N, h, \succeq^d)$ $[\succeq^d \in \mathcal{D}^N_{\text{d}}]$, the allocation $\text{TTC}(\succeq^d)$, i.e., $\text{TTC}(\succeq) := \text{TTC}(\succeq^d)$. Roth and Postlewaite (1977, Theorem 2) showed that for each Shapley-Scarf market $(N, h, \succeq^d)$,

$$SC\left(\succeq^d\right) = \{\text{TTC}\left(\succeq^d\right)\}.$$

Ma (1994, Theorem 1) characterized the corresponding rule for Shapley-Scarf markets by *individual rationality*, *Pareto optimality*, and *strategy-proofness* (see also Svensson, 1999, Theorem 2). We establish a corresponding characterization for demand lexicographic markets.[4]

Theorem 1. *The TTC rule is the only rule on $\mathcal{D}^N_{\text{dlex}}$ satisfying individual rationality, Pareto optimality, and strategy-proofness.*[5]

Proof. It is easy to see that the TTC rule by definition (and since preferences are demand lexicographic) satisfies *individual rationality* and *Pareto optimality*. Furthermore, the TTC rule satisfies *strategy-proofness* for Shapley-Scarf markets and hence no agent can obtain a better house by changing his preferences. Since agents throughout the TTC algorithm, in line with their demand lexicographic preferences, primarily care about obtaining the best possible house without influencing who receives their house, the TTC rule is also *strategy-proof* for demand lexicographic markets. Aziz and Lee (2020) also prove that the TTC rule satisfies *individual rationality*, *Pareto optimality*, and *strategy-proofness* on $\mathcal{D}^N_{\text{dlex}}$.

Next, let φ be a rule satisfying *individual rationality*, *Pareto optimality*, and *strategy-proofness*. Let $(N, h, \succeq)$ be such that $\succeq \in \mathcal{D}^N_{\text{dlex}}$ and $(N, h, \succeq^d)$ $[\succeq^d \in \mathcal{D}^N_{\text{d}}]$ is the associated Shapley-Scarf market.

We show that $\text{TTC}(\succeq) = \varphi(\succeq)$. We first explain the intuition of the proof.

Consider a trading cycle that forms in the first step of the TTC algorithm for $(N, h, \succeq)$. Note that an agent who points at his own house will receive it due to *individual rationality*. For larger trading cycles, agents in the trading cycle receive their most preferred houses, which are different from their own houses. The proof that agents in that trading cycle receive their TTC houses under φ as well proceeds as follows. By *individual rationality*, trading cycle agents receive a house that is at least as preferred as their own house. Now,

[4] A corresponding characterization can then symmetrically be derived (by changing the roles of houses and agents) for the domain of supply lexicographic preferences.

[5] Note that the TTC does satisfy the stronger property of *group strategy-proofness* on $\mathcal{D}^N_{\text{dlex}}$: a rule is **group strategy-proof** if no group of agents can ever benefit by jointly misrepresenting their preferences, i.e., for each market $(N, h, \succeq)$, there is no set of agents $S \subseteq N$ and no preference profile $\widetilde{\succeq}$ such that for each $i \in N \setminus S$, $\widetilde{\succeq}_i = \succeq_i$, for each $j \in S$, $\varphi_j(\widetilde{\succeq}) \succeq_i \varphi_j(\succeq)$, and for some $k \in S$, $\varphi_k(\widetilde{\succeq}) \succ_k \varphi_k(\succeq)$. The *group strategy-proofness* of the TTC rule follows from Hong and Park (2020, Proposition 10), a result that was established for a larger preference domain.

imagine that demand preferences for trading cycle agents were such that each of them ranks their own house just below their TTC house. Then, preferences being demand lexicographic, by *individual rationality*, each trading cycle agent either receives their TTC or their own house. However, then the only two feasible allotments for trading cycle agents are either the endowment allotments or the TTC allotments. By *Pareto optimality*, trading cycle agents then all must receive their TTC house. Of course, original demand lexicographic preferences might not have been based on the specific demand preferences we just assumed; this is where *strategy-proofness* is used to change trading cycle agents' preferences one by one back to their original preferences without changing the allotments of trading cycle agents under φ.

Once we have shown that agents who trade in the first step of the TTC algorithm always receive their TTC allotments under φ, we can consider agents who trade in the second step of the TTC. If such a trading cycle consists of only one agent, then that agent can never receive his most preferred house (since it is traded in Step 1) and, by *individual rationality*, now receives his second most preferred own house. For larger trading cycles, we first consider demand preferences where trading cycle agents rank their "TTC house" first and their own house second and follow the same proof steps as for Step 1 trading cycle agents; etc. The formal proof for agents trading in the first step of the TTC algorithm now follows.

Consider a trading cycle that forms in the first step of the TTC algorithm for $(N, h, \succeq)$. If the trading cycle consists of only one agent, then that agent most prefers his own house and, by *individual rationality*, receives it. Hence, consider a trading cycle consisting of agents $i_0, \ldots, i_K$, $K \geq 1$, and houses $h_{i_0}, \ldots, h_{i_K}$. Note that each agent $i_k \in \{i_0, \ldots, i_K\}$, according to his demand preferences $\succeq^d_{i_k}$, prefers house $h_{i_{k+1}}$ most among houses in H.

For every $i_k \in \{i_0, \ldots, i_K\}$ we define preferences $\tilde{\succeq}_{i_k} \in \mathcal{D}_{\text{dlex}}$ based on new demand preferences $\tilde{\succeq}^d$ and the original supply preferences $\succeq^s$ such that

- $h_{i_{k+1}} \tilde{\succ}^d_{i_k} h_{i_k} \tilde{\succ}^d_{i_k} \ldots$ (modulo K),[6]

e.g., by moving h_{i_k} just after $h_{i_{k+1}}$ in the demand preferences (without changing the ordering of other houses) and by moving i_k just after i_{k-1} in the supply preferences (without changing the ordering of other agents). We omit the mention of "modulo K" in the sequel.

Following standard notation, for any set $S \subseteq N$, $(\tilde{\succeq}_S, \succeq_{-S})$ is the preference profile that is obtained from $\succeq$ when all agents $i \in S$ change their preferences from $\succeq_i$ to $\tilde{\succeq}_i$. Consider the preference profile

[6]Note that the we could add a transformation of supply preferences from $\succeq^s$ to $\tilde{\succeq}^s$ such that $i_{k-1} \tilde{\succ}^s_{i_k} i_k \tilde{\succ}^s_{i_k} \ldots$ (modulo K), but it is not necessary for the proof.

$$\succeq^0 = (\widetilde{\succeq}_{\{i_0,\ldots,i_K\}}, \succeq_{-\{i_0,\ldots,i_K\}}).$$

For each $0 \le k \le K$, $\text{TTC}_{i_k}(\succeq^0) = (h_{i_{k+1}}, i_{k-1})$. Furthermore, by *individual rationality*, for each $0 \le k \le K$, $\varphi_{i_k}(\succeq^0) \in \{(h_{i_k}, i_k), (h_{i_{k+1}}, i_{k-1})\}$. So, the set of houses allocated to agents in $\{i_0, \ldots, i_K\}$ at allocation $\varphi(\succeq^0)$ equals $\{h_{i_0}, \ldots, h_{i_K}\}$. Hence, by *Pareto-optimality*, for each $i_k \in \{i_0, \ldots, i_K\}$, $\varphi_{i_k}(\succeq^0) = (h_{i_{k+1}}, i_{k-1}) = \text{TTC}_{i_k}(\succeq^0)$.

Next, let $l_1 \in \{i_0, \ldots, i_K\}$ such that $l_1 + 1$ is the successor and $l_1 - 1$ is the predecessor in the trading cycle and $\text{TTC}_{l_1}(\succeq^0) = (h_{l_1+1}, l_1 - 1)$. Consider the preference profile

$$\succeq^1 = (\widetilde{\succeq}_{\{i_0,\ldots,i_K\}\setminus\{l_1\}}, \succeq_{-(\{i_0,\ldots,i_K\}\setminus\{l_1\})}) = (\succeq_{l_1}, \succeq^0_{-l_1}).$$

Assume that starting from preference profile $\succeq^0$, agent l_1 changes his preferences from $\widetilde{\succeq}_{l_1}$ to $\succeq_{l_1}$. Then, since agent l_1's trading cycle did not change, $\text{TTC}_{l_1}(\succeq^0) = \text{TTC}_{l_1}(\succeq^1) = (h_{l_1+1}, l_1 - 1)$. Considering the same preference change under rule φ, by *strategy-proofness* we have $\varphi_{l_1}(\succeq^1) \succeq_{l_1} \varphi_{l_1}(\succeq^0)$. Recall that at $\varphi_{l_1}(\succeq^0) = (h_{l_1+1}, l_1 - 1)$, agent l_1 receives his favorite house according to demand preferences $\succeq^d_{l_1}$. Hence, $\varphi_{l_1}(\succeq^1) = (h_{l_1+1}, \cdot)$. Next, by *individual rationality*, for each $i_k \in \{i_0, \ldots, i_K\} \setminus \{l_1\}$, $\varphi_{i_k}(\succeq^1) \in \{(h_{i_k}, i_k), (h_{i_{k+1}}, i_{k-1})\}$. So, the set of houses allocated to agents in $\{i_0, \ldots, i_K\}$ at allocation $\varphi(\succeq^1)$ equals $\{h_{i_0}, \ldots, h_{i_K}\}$. Then, by *Pareto optimality*, for each $i_k \in \{i_0, \ldots, i_K\}$, $\varphi_{i_k}(\succeq^1) = (h_{i_{k+1}}, i_{k-1}) = \text{TTC}_{i_k}(\succeq^1)$.

Now, let $l_2 \in \{i_0, \ldots, i_K\} \setminus \{l_1\}$ such that $l_2 + 1$ is the successor and $l_2 - 1$ is the predecessor in the trading cycle and $\text{TTC}_{l_2}(\succeq^1) = (h_{l_2+1}, l_2 - 1)$. Consider the preference profile

$$\succeq^2 = (\widetilde{\succeq}_{\{i_0,\ldots,i_K\}\setminus\{l_1,l_2\}}, \succeq_{-(\{i_0,\ldots,i_K\}\setminus\{l_1,l_2\})}) = (\succeq_{l_2}, \succeq^1_{-l_2}).$$

Assume that starting from preference profile $\succeq^1$, agent l_2 changes his preferences from $\widetilde{\succeq}_{l_2}$ to $\succeq_{l_2}$. Then, since agent l_2's trading cycle did not change, $\text{TTC}_{l_2}(\succeq^1) = \text{TTC}_{l_2}(\succeq^2) = (h_{l_2+1}, l_2 - 1)$. Considering the same preference change under rule φ, by *strategy-proofness* we have $\varphi_{l_2}(\succeq^2) \succeq_{l_2} \varphi_{l_2}(\succeq^1)$. Recall that at $\varphi_{l_2}(\succeq^1) = (h_{l_2+1}, l_2 - 1)$, agent l_2 receives his favorite house according to demand preferences $\succeq^d_{l_2}$. Hence, $\varphi_{l_2}(\succeq^2) = (h_{l_2+1}, \cdot)$. Since the choice of agents $\{l_1, l_2\} \subseteq \{i_0, \ldots, i_K\}$ was arbitrary, we obtain by the same argument changing the roles of l_1 and l_2 that $\varphi_{l_1}(\succeq^2) = (h_{l_1+1}, \cdot)$

Next, by *individual rationality*, for each $i_k \in \{i_0, \ldots, i_K\} \setminus \{l_1, l_2\}$, $\varphi_{i_k}(\succeq^1) \in \{(h_{i_k}, i_k), (h_{i_{k+1}}, i_{k-1})\}$. So, the set of houses allocated to agents in $\{i_0, \ldots, i_K\}$ at allocation $\varphi(\succeq^2)$ equals $\{h_{i_0}, \ldots, h_{i_K}\}$. Then, by *Pareto optimality*, for each $i_k \in \{i_0, \ldots, i_K\}$, $\varphi_{i_k}(\succeq^2) = (h_{i_{k+1}}, i_{k-1}) = \text{TTC}_{i_k}(\succeq^2)$.

We continue to replace the preferences of agents in $\{i_0, \ldots, i_K\} \setminus \{l_1, l_2\}$ one at a time as above until we reach the preference profile $\succeq$ with the conclusion that for each $i_k \in \{i_0, \ldots, i_K\}$, $\varphi_{i_k}(\succeq) = \text{TTC}_{i_k}(\succeq)$. $\square$

In the following remarks, we discuss various important aspects of Theorem 1.

Remark 1 (On the proof of the TTC rule characterization). Note that in the proof of Theorem 1, we make explicit use of the steps used by the TTC algorithm to compute the TTC allocation. In contrast, Ma's (Ma, 1994, Theorem 1) original proof is based on the fact that the TTC allocation equals the unique strong core allocation; thus, his proof is not based on the TTC algorithm but on the absence of weak blocking coalitions. To the best of our knowledge, Svensson (1999, Lemma 3 and Theorem 2) was the first with a TTC algorithm based proof strategy to characterize the TTC rule. Note that in comparison to Svensson's proof of the TTC rule characterization, we have to make sure that all preference transformations are in subdomain $\mathcal{D}_{\text{dlex}}$. □

Remark 2 (Our TTC characterization in comparison with Hong and Park, 2020, Proposition 11). Hong and Park (2020, Proposition 11) prove that on the domain of order-preserving preferences with respect to own-allotments, the TTC rule is the only rule satisfying *individual rationality*, *stability*,[7] and *strategy-proofness*. *Stability* implies *Pareto optimality* and Hong and Park (2020, Example 2) show that *stability* is a stronger requirement than *Pareto optimality*. Hence, in our Theorem 1, we use a weaker property to characterize the TTC rule. On the other hand, we define the TTC rule on a smaller preference domain than Hong and Park (2020) do. Thus, one assumption in our characterization result is stronger while another is weaker than in Hong and Park (2020, Proposition 11) and our results are logically independent. □

Remark 3 (Strategy-proofness and essentially single-valued cores). Sönmez (1999) proved for so-called generalized indivisible goods allocation problems, which include Shapley-Scarf housing markets, that if there exists a rule φ that is *individually rational*, *Pareto optimal*, and *strategy-proof*, then the strong core solution is essentially single-valued (i.e., all agents are indifferent between any pair of allocations in the strong core) and the rule φ is a selection of the strong core solution itself.

First, for housing markets with lexicographic preferences, the strong core solution is not essentially single valued. We can demonstrate this with the following example (see Klaus and Meo, 2021, Example 2). Let $N = \{1, 2, 3\}$ and $h = (h_1, h_2, h_3)$. We assume that $\succeq \in \mathcal{D}^N_{\text{dlex}}$ with the following demand and supply preferences.

[7] An allocation $a \in \mathcal{A}$ is **stable** if there exists no allocation $b \in \mathcal{A}$ such that $b \neq a$ and for all agents $i \in N$ with $b(i) \neq a(i)$, $b \succ_i a$.

Agent 1		Agent 2		Agent 3	
$\succeq_1^d$	$\succeq_1^s$	$\succeq_2^d$	$\succeq_2^s$	$\succeq_3^d$	$\succeq_3^s$
h_2	3	h_1	1	h_2	·
h_3	2	h_3	3	h_1	·
h_1	1	h_2	2	h_3	·

The empty column means that any linear order $\succeq_3^s$ can be considered.

The strong core for the Shapley-Scarf market $\succeq^d$ is formed by the unique allocation (h_2, h_1, h_3), which can easily be computed by Gale's top trading cycles (TTC) algorithm based on $\succeq^d$. Then, $(h_2, h_1, h_3) \in SC(\succeq)$. Next, $(h_2, h_3, h_1) \in SC(\succeq)$ because agent 1 gets his most preferred allotment $(h_2, 3)$ and coalition $S = \{2, 3\}$ cannot block by swapping their endowments (at allocation (h_1, h_3, h_2) agent 2 would be worse off since $(h_3, 1) \succ_2 (h_3, 3)$).[8] Hence, (h_2, h_1, h_3), $(h_2, h_3, h_1) \in SC(\succeq)$ but $(h_2, h_1, h_3) \succ_2 (h_2, h_3, h_1)$. Thus, $SC(\succeq)$ is not essentially single-valued.

Second, for housing markets with lexicographic preferences the richness condition on agents' preferences as required in Sönmez (1999) is violated:[9] consider the following demand lexicographic preferences for agent 1,

$$\succeq_1\colon (h_2, 2) \succ_1 (h_2, 3) \succ_1 \boldsymbol{(h_3, 2)} \succ_1 (h_3, 3) \succ_1 \boldsymbol{(h_1, 1)}.$$

Then, domain richness Sönmez (1999, Assumption B) would require that agent 1's demand lexicographic preferences $\succeq_1$ can be transformed either into preferences

$$\widetilde{\succeq}_1 : (h_2, 2) \,\widetilde{\succ}_1\, (h_2, 3) \,\widetilde{\succ}_1\, \boldsymbol{(h_3, 2)} \,\widetilde{\succ}_1\, \boldsymbol{(h_1, 1)} \,\widetilde{\succ}_1\, (h_3, 3)$$

or into preferences

$$\widetilde{\widetilde{\succeq}}_1 : (h_2, 3) \,\widetilde{\widetilde{\succ}}_1\, (h_2, 2) \,\widetilde{\widetilde{\succ}}_1\, \boldsymbol{(h_3, 2)} \,\widetilde{\widetilde{\succ}}_1\, \boldsymbol{(h_1, 1)} \,\widetilde{\widetilde{\succ}}_1\, (h_3, 3).$$

However, neither of the latter preferences are demand lexicographic.

To conclude, for our model, it is still true that if a rule is *individually rational*, *Pareto optimal*, and *strategy-proof*, then it is a selection of the strong core solution. More precisely, for each market $(N, h, \succeq)$, $\succeq \in \mathcal{D}_{\text{dlex}}^N$ with associated Shapley-Scarf market $(N, h, \succeq^d)$,

[8] Note that $(h_2, h_3, h_1) \notin SC(\succeq^d)$ because it is weakly blocked by $S = \{2, 3\}$ through (h_1, h_3, h_2) (agent 2 receives the same house and agent 3 a better house).

[9] The domain richness condition of Sönmez (1999, Assumption B) for strict preferences over allotments reads as follows: for each preference relation $\succeq_i$ and each allotment $(h, j) \in \mathcal{A}_i$ such that $(h, j) \succ_i (h_i, i)$, there exists a preference relation $\widetilde{\succeq}_i$ such that (i) the weak upper contour sets at (h, j) under $\succeq_i$ and $\widetilde{\succeq}_i$ are the same, i.e., for all $(h', k) \in \mathcal{A}_i$, $[(h', k) \succeq_i (h, j)$ if and only if $(h', k) \,\widetilde{\succeq}_i\, (h, j)]$, (ii) the weak lower contour sets at (h, j) under $\succeq_i$ and $\widetilde{\succeq}_i$ are the same, i.e., for all $(h', k) \in \mathcal{A}_i$, $[(h, j) \succeq_i (h', k)$ if and only if $(h, j) \,\widetilde{\succeq}_i\, (h', k)]$, and (iii) the endowment allotment ranks right after (h, j), i.e., if $(h, j) \succ_i (h', k)$, then $(h, j) \,\widetilde{\succ}_i\, (h_i, i) \,\widetilde{\succeq}_i\, (h', k)$.

$$\{\text{TTC}(\succeq)\} = SC(\succeq^d) \subseteq SC(\succeq),$$

where $SC(\succeq^d) \subseteq SC(\succeq)$ follows from Klaus and Meo (2021, Proposition 4). Our example above (Klaus and Meo, 2021, Example 2) shows that $\{\text{TTC}(\succeq)\} = SC(\succeq^d) \subsetneq SC(\succeq)$ is possible. □

4 Conclusions

On the demand lexicographic preference domain, the top trading cycles (TTC) rule, is the unique rule satisfying *individual rationality*, *Pareto optimality*, and *strategy-proofness*.

For classical Shapley-Scarf housing markets (with strict preferences), the TTC rule is also characterized by these three properties and the TTC allocation equals the unique strong core allocation. This gives rise to two proof strategies for the characterization of the TTC rule: one using the structure of the unique strong core allocation (Ma, 1994, Theorem 1), the other the structure imposed by the TTC algorithm (Svensson, 1999, Lemma 3 and Theorem 2).

φ is the TTC rule

$\Updownarrow$ Svensson (1999)

φ satisfies *individual rationality*, *Pareto optimality*, and *strategy-proofness*

$\Updownarrow$ Ma (1994)

φ assigns the unique strong core allocation

We use the latter proof strategy and in fact can show that in our model with lexicographic preferences, the properties *individual rationality*, *Pareto optimality*, and *strategy-proofness* characterize the TTC rule but *not* the strong core (which can be multi-valued).

φ is the TTC rule

$\Updownarrow$ Theorem 1

φ satisfies *individual rationality*, *Pareto optimality*, and *strategy-proofness*

$\not\Uparrow\Downarrow$ Klaus and Meo (2021, Proposition 4)

φ assigns a strong core allocation

Hence, our characterization of TTC rules sheds some important light on the classical characterization result in that the properties *individual rationality*, *Pareto optimality*, and *strategy-proofness* primarily induce the TTC allocation and that the strong core allocation is only secondarily (or coincidentally) pinned down by the three properties.

References

Aziz, H. and Lee, E. J. (2020): "The temporary exchange problem." AAMAS2020 Conference Presentation.

Graziano, M. G., Meo, C., and Yannelis, N. C. (2020): "Shapley and Scarf housing markets with consumption externalities." *Journal of Public Economic Theory*, 22: 1481–1514.

Hong, M. and Park, J. (2020): "Core and top trading cycles in a market with indivisible goods and externalities." Working Paper.

Klaus, B. and Meo, C. (2021): "The core for housing markets with limited externalities." Cahier de recherches économiques 21.03, University of Lausanne, HEC.

Ma, J. (1994): "Strategy-proofness and the strict core in a market with indivisibilities." *International Journal of Game Theory*, 23: 75–83.

Mumcu, A. and Sağlam, I. (2007): "The core of a housing market with externalities." *Economics Bulletin*, 55: 1–5.

Roth, A. E. (1982): "Incentive compatibility in a market with indivisible goods." *Economics Letters*, 9: 127–132.

Roth, A. E. and Postlewaite, A. (1977): "Weak versus strong domination in a market with indivisible goods." *Journal of Mathematical Economics*, 4: 131–137.

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

Sönmez, T. (1999): "Strategy-proofness and essentially single-valued cores." *Econometrica*, 67,3: 677–689.

Svensson, L.-G. (1999): "Strategy-proof allocation of indivisible goods." *Social Choice and Welfare*, 16: 557–567.