Core and Top Trading Cycles in a Market with Indivisible Goods and Externalities

Miho Hong[*] Jaeok Park[†]

December 17, 2017

Abstract

In this paper, we incorporate externalities into Shapley-Scarf housing markets. Agents' preferences are defined over allocations rather than houses, and we focus on preferences that are egocentric in the sense that agents primarily care about their own allotments. When preferences are egocentric, we can apply the top trading cycles (TTC) algorithm using the associated preferences over houses. We propose two solution concepts based on the core. We establish the existence of a solution by showing that the allocation generated by the TTC algorithm is a solution, and we present a further preference restriction under which a solution is unique. We also investigate the properties of the TTC algorithm as a mechanism. Our results extend the existing results on the TTC algorithm to the case of egocentric preferences, and they suggest that the TTC algorithm is useful and has desirable properties even in the presence of externalities.

Keywords: Core, Externalities, Housing markets, Indivisibility, Stability, Top trading cycles.

JEL Classification: C71, C78, D62, D71.

1 Introduction

In this paper, we study a market with indivisibilities and externalities. In their pioneering work, Shapley and Scarf (1974) introduce an exchange economy with a finite number of agents in which each agent owns an indivisible object (say, a house), has use for exactly one object, and looks for trading with other agents. Such an economy is referred to as a housing market in the literature. It is one of the simplest kinds of exchange economies one can imagine, and together with a house allocation problem (Hylland and Zeckhauser, 1979) it

[*]School of Economics, Yonsei University (e-mail: miho.m.hong@gmail.com)

[†]Corresponding author. School of Economics, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea (e-mail: jaeok.park@yonsei.ac.kr)

has served as a basic model for indivisible goods allocation problems and one-sided matching problems (see, for example, Moulin, 1995; Abdulkadiroğlu and Sönmez, 2013). Externalities are abundant in the real world, and a market with indivisible goods is not an exception. For example, in residential areas, a resident of a house cares about the demographics of his neighbors, and in universities, a faculty member's preferences regarding offices may depend on his relationships with colleagues using nearby offices. Despite the ubiquity of externalities, most existing works on housing markets have not taken externalities into consideration, assuming that each agent cares only about his own allotment but not others'. The aim of this paper is to provide an analysis of housing markets with externalities.

Shapley and Scarf (1974) show that the core (defined by strong domination) of a housing market is nonempty. They also introduce an algorithm called the top trading cycles (TTC) algorithm that produces an allocation in the core. Roth and Postlewaite (1977) complement Shapley and Scarf's (1974) results by proving that the core (defined by weak domination) of a housing market is a singleton when agents have strict preferences over houses and that the unique allocation in the core can be obtained by the TTC algorithm. Mumcu and Saglam (2007) incorporate externalities into housing markets, having agents' preferences defined over allocations rather than houses. They show that the core may be empty or contain more than one allocation when there are externalities. Given the nonexistence result of Mumcu and Saglam (2007), it is natural to restrict the preference domain and search for a core-like solution concept that yields an existence result. This is precisely the approach we take in this paper.

We focus on a class of preferences where an agent is primarily interested in his own allotment. That is, each agent has a strict preference order over houses, and whenever any two allocations assign distinct houses to him, his ranking between the two allocations is determined by that between the houses he receives at those allocations.[1] An agent is subject to externalities in the sense that he may have a strict ranking between two allocations that assign the same house to him. We call this kind of preferences egocentric preferences. Though restrictive, it is a reasonable class of preferences over allocations. It expresses the idea that one's own allotment is of higher significance than others' allotments in one's utility. In other words, it reflects a situation in which the size of externalities is small relative to that of utility from one's own allotment.

An advantage of focusing on egocentric preferences is that we can still apply the TTC algorithm because each agent has well-defined preferences over houses. In order to propose

[1] Dutta and Massó (1997) propose similar preference restrictions in two-sided many-to-one matching markets. They consider "firm-lexicographic" (resp. "worker-lexicographic") preferences where a worker's preferences over firms (resp. colleagues) dictate his overall preferences over firm-colleague pairs. In the context of two-sided one-to-one matching markets with contracts, Li (1993) uses weak externalities to refer to a situation where an agent cares about his own partner and contract prior to others'. Sasaki and Toda (1996) call such preferences order preserving.

solution concepts for housing markets with externalities, we investigate the properties of allocations generated by the TTC algorithm. Our solution concepts are based on the core. In the presence of externalities, it matters to a blocking coalition how the residual agents react to its deviation. Following Hart and Kurz (1983), we consider two kinds of behavior of the residual agents, leading to the two concepts of the core called the γ-core and the δ-core. In the γ-model, when a coalition deviates, the remaining agents cannot make any trading arrangement among themselves and end up with their own endowments (see Ehlers, 2014). In contrast, in the δ-model, the remaining agents keep trading their houses as before whenever possible.[2] In the case of the γ-model, a group of agents may form a blocking coalition just to make the outside agents receive their endowments while the allotments of the coalition members remain the same. Obviously, without externalities such an attempt is never profitable for the coalition members, but with externalities it may make them better off. However, if the outside agents prefer the original allocation, they can form a subsequent coalition to return to it. Then the attempt to reset others' allotments as their endowments is ineffective. Thus, we require that, when a coalition attempts to reset outsiders' allotments as their endowments, the proposed allocation should not be reversed to the original allocation by the outsiders. This leads to a weakening of the γ-core, which we call the irreversible γ-core.

Suppose that agents agree on an allocation and execute trade to achieve the allocation. Without externalities, no further improvement by a coalition is possible if the allocation is in the core. However, with externalities, a group of agents may get better off by reallocating their allotments among themselves even if the allocation is in the irreversible γ-core or in the δ-core. This is because reallocation by a coalition can make an outside agent worse off and thus the new allocation does not necessarily dominate the original allocation. We call an allocation that is immune to further coalitional deviations a stable allocation, following the terminology of Roth and Postlewaite (1977). We can expect that an allocation that is not stable would not persist, and thus we focus on stable allocations. In sum, we take stable allocations in the irreversible γ-core and those in the δ-core as our solution concepts.

In our main result (Theorem 1), we show that the allocation generated by the TTC algorithm is a stable allocation in the irreversible γ-core as well as in the δ-core. We further show that, when there are aligned interests among externality creators, a stable allocation in the irreversible γ-core and one in the δ-core are unique and coincide with the allocation obtained by the TTC algorithm. So our proposed solutions are guaranteed to exist, are unique under a further preference restriction, and can be obtained by applying the TTC algorithm. Our result generalizes the existing result that an allocation in the core of a

[2] Existing studies on two-sided matching markets with externalities have made similar assumptions: when some agents deviate, other unaffected agents keep their matches. See, for example, Mumcu and Saglam (2010) for one-to-one matching and Bando (2012, 2014) for many-to-one matching.

housing market without externalities is unique and can be obtained by the TTC algorithm. There are other desirable properties of the TTC algorithm established in the existing literature. For example, Abdulkadiroğlu and Sönmez (1998) show that one can generate all Pareto efficient allocations from the TTC algorithm by varying agents' endowments, Roth (1982) and Moulin (1995) prove that the TTC algorithm as a mechanism is (coalitionally) strategy-proof, and Ma (1994) establishes that a mechanism is individually rational, Pareto efficient, and strategy-proof if and only if it is the TTC mechanism. In Propositions 1–3, we show that these existing results can be generalized to the case of egocentric preferences when we replace Pareto efficiency with stability. This suggests that stability plays a more important role than Pareto efficiency in the presence of externalities, while without externalities these two concepts coincide. Our results also indicate that it is desirable to select the allocation generated by the TTC algorithm when there are multiple stable allocations in the irreversible γ-core or in the δ-core. So our analysis implies that, when externalities are not so large, the TTC algorithm continues to provide the best solution for housing markets. In other words, the desirable properties of the TTC algorithm are robust to the introduction of small externalities.[3]

There is a growing body of literature on matching markets with externalities, especially in two-sided settings (see Bando *et al.*, 2016, for a recent survey). Sasaki and Toda (1996) provide an early analysis on two-sided one-to-one matching problems with externalities. They consider a scenario where a deviating pair holds a pessimistic belief about the reaction of non-deviating agents, and they show that a stable matching exists for all preference profiles if and only if agents estimate that all possible matchings can arise after a deviation. Mumcu and Saglam (2010) also consider two-sided one-to-one matching problems with externalities but a different stability notion analogous to the δ-core, as explained in footnote 2. They provide a sufficient condition on the preference profile for the existence of stable matchings. Bando (2012, 2014) studies two-sided many-to-one matching problems with externalities among firms, and he presents preference restrictions for the existence of quasi-stable matchings as well as a modified deferred acceptance algorithm that generates the worker-optimal quasi-stable matching. Dutta and Massó (1997) analyze two-sided many-to-one matching problems with preferences over colleagues. In this setting, externalities exist only among workers hired by the same firm, and thus, unlike in our model and the aforementioned models, the core concept does not depend on the reaction of residual agents. Dutta and Massó (1997) show among other results that, when preferences are firm-lexicographic, a matching in the core exists and can be obtained by finding a stable matching under the associated preference profile without externalities. Our work is closely related to

[3]In a similar vein, Fisher and Hafalir (2016) consider two-sided one-to-one matching markets with small aggregate externalities and provide a sufficient condition for the existence of stable matchings.

theirs in that not only we adopt a similar preference restriction but also we establish the existence of solutions by using the allocation in the core of the associated housing market without externalities. There are not many studies on one-sided matching problems with externalities. As already mentioned, Mumcu and Saglam (2007) study housing markets with externalities and show that there can be no or more than one allocation in the core. Baccara *et al.* (2012) examine the problem of assigning offices to faculty members, which can be modeled as a house allocation problem, and quantify the effects of network externalities using field data. Sönmez (1999) and Ehlers (2014) consider a class of indivisible goods allocation problems, which includes housing markets as a special case, while allowing externalities. Their main results and comparison with ours are discussed in Section 4.

The rest of this paper is organized as follows. In Section 2, we introduce the model of housing markets, preference restrictions, and our solution concepts. In Section 3, we describe the TTC algorithm and study the properties of allocations it produces. In Section 4, we examine the properties of the TTC algorithm as a mechanism. In Section 5, we conclude. All the proofs are relegated to the Appendix.

2 The Model

2.1 Housing Markets

We consider an exchange economy with a finite number of agents in which each agent owns a house initially, has use for one and only one house, and seeks to trade houses with other agents. Let N and H be finite sets of agents and houses, respectively, with $|N| = |H| \geq 3$. An *allocation* describes the house assigned to each agent and is represented by a one-to-one function from N onto H. Let $\mathcal{A}$ be the set of all allocations. For any allocation $a \in \mathcal{A}$, $a(i)$ denotes the house received by agent i, and we call it the *allotment* of agent i at allocation a. A major departure from most previous work is that an agent cares not only about his own allotment but also about others'. So each agent's preferences are defined over the set of allocations instead of houses. For each agent $i \in N$, R_i denotes his preference relation, which is a complete and transitive binary relation on $\mathcal{A}$. For any $a, b \in \mathcal{A}$, aR_ib means that agent i likes allocation a at least as much as allocation b. Given R_i, we denote its asymmetric and symmetric parts by P_i and I_i, respectively. That is, for any $a, b \in \mathcal{A}$, we write aP_ib if aR_ib and not bR_ia, and we write aI_ib if aR_ib and bR_ia. A profile of agents' preference relations is denoted by $R = (R_i)_{i \in N}$. Agents' ownership is described by an initial endowment allocation $e \in \mathcal{A}$ so that $e(i)$ is the house initially owned by agent i. A *housing market (with externalities)* is a 4-tuple $\langle N, H, R, e \rangle$.

2.2 Preference Restrictions

We introduce several restrictions on preferences that will be used in our analysis. We assume that each agent is not indifferent between any two allocations where he receives different houses. In other words, aI_ib implies $a(i) = b(i)$. We also assume that each agent cares about his own allotment prior to others'. More specifically, aP_ib and $a(i) \neq b(i)$ implies $a'P_ib'$ for any a' and b' such that $a'(i) = a(i)$ and $b'(i) = b(i)$. We call a preference relation satisfying these two assumptions *egocentric*. Note that, given an egocentric preference relation R_i, we can derive a preference relation $\tilde{R}_i$ on H by defining $h\tilde{R}_ih'$ if (i) $h = h'$ or (ii) $h \neq h'$ and aP_ib for all $a, b \in \mathcal{A}$ such that $a(i) = h$ and $b(i) = h'$. By our egocentricity assumption, $\tilde{R}_i$ is a complete, transitive, and antisymmetric binary relation on H, expressing agent i's strict preference ranking of his own allotments at possible allocations. We refer to $\tilde{R}_i$ as the *preference relation over own-allotments associated with* R_i. We write the asymmetric and symmetric parts of $\tilde{R}_i$ as $\tilde{P}_i$ and $\tilde{I}_i$, respectively, which are defined analogously to P_i and I_i.

Most previous studies on housing markets assume that there are no externalities. We define an *egoistic* preference relation as an egocentric preference relation R_i that satisfies aI_ib for any a and b such that $a(i) = b(i)$. That is, an agent with egoistic preferences cares only about his own allotment having strict preferences over houses. We can regard $\tilde{R}_i$ as a binary relation on $\mathcal{A}$ rather than H, interpreting it as an egoistic preference relation. Then, given a housing market $\langle N, H, R, e\rangle$ with externalities, we can consider a related housing market $\langle N, H, \tilde{R}, e\rangle$ without externalities, where $\tilde{R} = (\tilde{R}_i)_{i\in N}$.

We say that a preference profile R has *aligned interests among externality creators* if the following condition is met: for any $a, b \in \mathcal{A}$,

$$\text{if there exists } j \in N \text{ such that } a(j) = b(j) \text{ and } aP_jb,$$
$$\text{then either } aP_ib \text{ for all } i \in N \text{ with } a(i) \neq b(i) \text{ or } bP_ia \text{ for all } i \in N \text{ with } a(i) \neq b(i).$$

The condition says that, if there are agents who can create an externality to another agent by reallocating their allotments, these agents should have common preferences regarding whether to implement the reallocation or not. In other words, if some agents have conflicting interests in reallocating their allotments, then every other agent who receives the same allotment should be indifferent to such a reallocation (that is, no externalities are created by the reallocation).

2.3 Properties of Allocations

We discuss several desirable properties that an allocation can possess. The core is a central solution concept for cooperative games, providing a set of allocations that are immune to coalitional deviations. When there are externalities, the preferences of coalition members depend on the allocation outside of the coalition. Hence, when a coalition plans a deviation, it matters how the residual agents react to the deviation. In this paper, we consider two kinds of behavior of the residual agents following the two models proposed by Hart and Kurz (1983).[4] In the γ-model, it is assumed that the outside agents stay put with their endowments (see Ehlers, 2014), whereas in the δ-model, it is assumed that the outside agents trade their endowments as before whenever possible. Below we formally describe two notions of the core corresponding to these two models.

A nonempty subset of N is called a *coalition*, while N as a coalition is called the grand coalition. For any $a \in \mathcal{A}$ and any $S \subseteq N$, let $a(S) = \{a(i) : i \in S\}$. For a given allocation a, we say that a coalition S is a *trading cycle in* a if $a(S) = e(S)$ and $a(S') \neq e(S')$ for any nonempty $S' \subseteq S$ with $S' \neq S$. That is, a trading cycle in a is a minimal subset of agents who trade houses among themselves to obtain their allotments at a. The elements of a trading cycle in a can be indexed as $\{i_1, \ldots, i_m\}$ such that $a(i_l) = e(i_{l+1})$ for all $l = 1, \ldots, m-1$ and $a(i_m) = e(i_1)$, and we sometimes represent the trading cycle by the sequence $(i_1, \ldots, i_m)$. An allocation a uniquely partitions the set of agents into trading cycles in a, and we use $S_i^{a,e}$ to denote the trading cycle in a to which agent i belongs given the endowment e.

An allocation $b \in \mathcal{A}$ γ-*dominates* an allocation $a \in \mathcal{A}$ via coalition T (or, coalition T γ-*blocks* a via b) in the market $\langle N, H, R, e\rangle$ if (i) $b(T) = e(T)$, (ii) $b(i) = e(i)$ for all $i \notin T$, (iii) bR_ia for all $i \in T$, and (iv) bP_ia for some $i \in T$. The γ-*core* of the market $\langle N, H, R, e\rangle$ is the set of allocations that are not γ-dominated by any allocation in $\langle N, H, R, e\rangle$. An allocation $b \in \mathcal{A}$ δ-*dominates* an allocation $a \in \mathcal{A}$ via coalition T (or, coalition T δ-*blocks* a via b) in the market $\langle N, H, R, e\rangle$ if (i) $b(T) = e(T)$, (ii) $b(i) = e(i)$ for all $i \in (\cup_{j\in T} S_j^{a,e}) \setminus T$, (iii) $b(i) = a(i)$ for all $i \notin \cup_{j\in T} S_j^{a,e}$, (iv) bR_ia for all $i \in T$, and (v) bP_ia for some $i \in T$. The δ-*core* of the market $\langle N, H, R, e\rangle$ is the set of allocations that are not δ-dominated by any allocation in $\langle N, H, R, e\rangle$. In the γ-model, when a coalition deviates from an allocation a, the agreement to select a is nullified and the agents outside of the coalition consume their own endowments. In the δ-model, on the contrary, when a coalition deviates from an allocation a, the agreement to select a is still effective to the outside agents; a trading cycle in a that does not overlap with the coalition executes trades according to a, while outside agents who are in the same trading cycle in a as any agent in the coalition receive their own endowments.

When there are no externalities, agents in a coalition are indifferent to the outsiders' allotments, and thus the two concepts of the γ-core and the δ-core coincide with the usual

[4]Hart and Kurz (1983) consider the case where a coalition breaks up as a result of some members departing. In contrast, we focus on the case where the grand coalition breaks up.

notion of the core (defined by weak domination). Furthermore, it is well-known that, with egoistic preferences, the core is a singleton (Roth and Postlewaite, 1977). However, with egocentric preferences, the γ-core may be empty, as illustrated by the following example.

Example 1. Consider a housing market with $N = \{i_1, i_2, i_3, i_4\}$, $H = \{h_1, h_2, h_3, h_4\}$, and $e = (h_1, h_2, h_3, h_4)$.[5] Suppose that the agents have egocentric preferences and that the most preferred houses of agents i_1, i_2, i_3, and i_4 are h_2, h_1, h_4, and h_3, respectively, in their associated preferences over own-allotments. Let $a = (h_2, h_1, h_4, h_3)$ and $b = (h_2, h_1, h_3, h_4)$. Suppose that agents i_1 and i_2 (strictly) prefer b to a. Then they can form a coalition to γ-block a via b. Hence, a does not belong to the γ-core. Note that b does not belong to the γ-core either, as it can be γ-blocked by $\{i_3, i_4\}$ via (h_1, h_2, h_4, h_3). Suppose that every agent prefers a to any allocation other than a and b. Then any allocation other than a and b is γ-dominated by a via the grand coalition. Thus, with egocentric preferences, the γ-core may be empty. Note that $\{i_1, i_2\}$ cannot δ-block a via b, because agents i_3 and i_4 will still exchange their endowments. Note also that $\{i_3, i_4\}$ δ-blocks b via a. Hence, a is the unique allocation in the δ-core.

Example 1 suggests that we need to weaken the concept of the γ-core by restricting blocking possibilities in order to guarantee nonemptiness. Consider the situation where agents i_1 and i_2 γ-block a via b in Example 1. They receive the same allotments at the two allocations a and b, but they form a coalition in order to "reset" the allotments of agents i_3 and i_4 as these agents' endowments. However, once b is implemented, agents i_3 and i_4 can benefit mutually by exchanging their allotments h_3 and h_4, which brings the allocation back to a. Hence, we can regard the attempt of the coalition $\{i_1, i_2\}$ to γ-block a via b as ineffective in that the proposed allocation b will be reversed to a by the other agents.

An allocation $b \in \mathcal{A}$ *irreversibly γ-dominates* an allocation $a \in \mathcal{A}$ via coalition T (or, coalition T *irreversibly γ-blocks* a via b) in the market $\langle N, H, R, e\rangle$ if (i) b γ-dominates a via T in $\langle N, H, R, e\rangle$ and (ii) if $b(i) = a(i)$ for all $i \in T$, b is not γ-dominated by a via any $S \subseteq T^c$ in $\langle N, H, R, b\rangle$. The *irreversible γ-core* of the market $\langle N, H, R, e\rangle$ is the set of allocations that are not irreversibly γ-dominated by any allocation in $\langle N, H, R, e\rangle$. The additional requirement (ii) for blocking says that, when a coalition T proposes b in opposition to a, b should not be reversed to a by any coalition consisting of agents outside of T. It limits blocking possibilities when a coalition attempts to reset the outsiders' allotments as their endowments, thereby expanding the γ-core. In Example 1, $T = \{i_1, i_2\}$ does not irreversibly γ-block a via b because $b(i) = a(i)$ for all $i \in T$ while $S = \{i_3, i_4\}$ γ-blocks b via a in $\langle N, H, R, b\rangle$. Thus, a belongs to the irreversible γ-core of $\langle N, H, R, e\rangle$.

[5] When agents are indexed as i_1, i_2, and so on, we often describe an allocation as a list of houses where the first element in the list is assigned to i_1, the second to i_2, and so on.

A *house allocation problem* is a triple $\langle N, H, R\rangle$, which is the same as a housing market except that agents own houses collectively rather than individually. An allocation $a \in \mathcal{A}$ is *stable* in the problem $\langle N, H, R\rangle$ if a belongs to the γ-core of the market $\langle N, H, R, a\rangle$ (Roth and Postlewaite, 1977). That is, no coalition can block a stable allocation once it is implemented. Suppose that a is not stable. Then there exists an allocation b and a coalition T such that b γ-dominates a via T in $\langle N, H, R, a\rangle$. If $b(i) = a(i)$ for all $i \in T$, then $b = a$, a contradiction. Thus, there exists $i \in T$ such that $b(i) \neq a(i)$.[6] Let $T' = \{i \in N : b(i) \neq a(i)\}$. Then $T' \subseteq T$, and thus bR_ia for all $i \in T'$. With egocentric preferences, we have bP_ia for all $i \in T'$. Note that T' is the union of all non-singleton trading cycles in b with endowment a. Hence, the concept of stability can be equivalently defined as follows: An allocation a is stable if there exists no allocation b such that $b \neq a$ and $b(i)\tilde{P}_ia(i)$ for all $i \in N$ with $b(i) \neq a(i)$. Thus, stability is a concept that can be checked with preferences over own-allotments.

Remark 1. Our notion of stability is slightly different from that of Roth and Postlewaite (1977) in that we use a core concept defined by weak domination while Roth and Postlewaite (1977) adopt strong domination in which every coalition member gains strictly.[7] Kawasaki (2015) shows that the set of stable allocations in the sense of Roth and Postlewaite (1977) is the unique von Neumann–Morgenstern (vNM) stable set with respect to some domination relation. Although we use weak domination to define the γ-core, our assumption of egocentric preferences allows us to use strong domination for agents whose allotments differ at the two compared allocations in the alternative definition of stability, and we can obtain a similar result to that of Kawasaki (2015). For any $a, b \in \mathcal{A}$, let us define $b \rhd a$ if $b \neq a$ and $b(i)\tilde{P}_ia(i)$ for all $i \in N$ with $b(i) \neq a(i)$. Then the set of stable allocations is the unique vNM stable set with respect to the domination relation $\rhd$.[8]

An allocation $a \in \mathcal{A}$ is *Pareto efficient* in the problem $\langle N, H, R\rangle$ if there exists no allocation $b \in \mathcal{A}$ such that bR_ia for all $i \in N$ and bP_ia for some $i \in N$. Note that, unlike stable allocations, we need preferences over allocations in order to find Pareto efficient allocations and those in the irreversible γ-core and in the δ-core, as there can be an agent who receives the same allotment in the two compared allocations. Stable allocations as well as those in the irreversible γ-core and in the δ-core are Pareto efficient because the grand

[6] This explains that the concept of stable allocations is the same regardless of whether we use the γ-core or the irreversible γ-core in the definition. Note also that using the δ-core instead does not change the concept either because the endowment coincides with the allocation that a coalition attempts to block.

[7] Another difference is that we use the γ-core while Roth and Postlewaite (1977) use the core to define stable allocations.

[8] The set of stable allocations satisfies internal stability by definition. We can check that the relation $\rhd$ is transitive. Hence, if an allocation a is not stable, we can eventually obtain a stable allocation b such that $b \rhd a$, which proves external stability. Furthermore, we can show that there is no other vNM stable set than the set of stable allocations, following the argument in the proof of Theorem 1 of Kawasaki (2015).

coalition cannot block these allocations. When there are no externalities, a Pareto efficient allocation is stable,[9] and thus the concepts of stability and Pareto efficiency coincide. In contrast, stability is stronger than Pareto efficiency when there are externalities, as shown in the following example.

Example 2. Consider a housing market with $N = \{i_1, i_2, i_3\}$, $H = \{h_1, h_2, h_3\}$, and $e = (h_1, h_2, h_3)$. Suppose that the agents have egocentric preferences, that their associated preferences over own-allotments are given by $\tilde{P}_{i_1} : h_2, h_3, h_1$, $\tilde{P}_{i_2} : h_3, h_1, h_2$, $\tilde{P}_{i_3} : h_1, h_3, h_2$, and that agent i_1's preferences satisfy $(h_2, h_1, h_3) P_{i_1} (h_2, h_3, h_1)$. Let $b = (h_2, h_1, h_3)$. Since any reallocation of b makes agent i_1 worse off, b is Pareto efficient. However, b is not stable because agents i_2 and i_3 can get better off by exchanging their allotments from b.

Example 2 suggests that there can be too many Pareto efficient allocations when there are externalities: if an agent has a unique most preferred allocation, it is Pareto efficient even when some other agents can become better off by reallocating their allotments from the allocation. Our analysis in the next two sections shows that stability plays a more important role than Pareto efficiency in the presence of externalities.

3 Top Trading Cycles in a Housing Market with Externalities

3.1 Top Trading Cycles Algorithm

Without externalities, the unique allocation in the core can be obtained by the top trading cycles (TTC) algorithm developed by David Gale and introduced in Shapley and Scarf (1974). Given a housing market $\langle N, H, R, e \rangle$ with egocentric preferences, we can apply the TTC algorithm to the related market $\langle N, H, \tilde{R}, e \rangle$ with egoistic preferences. With egoistic preferences, the TTC algorithm proceeds as follows (see, for example, Roth, 1982 and Abdulkadiroğlu and Sönmez, 1998).

- *Step 1.* Each agent points to the owner of the house he prefers most. Since there is a finite number of agents, there exists at least one cycle of agents pointing to one another. Each agent in a cycle is assigned the endowment of the agent he points to and removed from the market. If there is at least one remaining agent, proceed to the next step. Otherwise, stop.

  In general, at:

- *Step k.* Each remaining agent points to the owner of the house he prefers most among the remaining houses. Each agent in a cycle is assigned the endowment of the agent

[9] As noted, if a is not stable, there exists another allocation b such that $bP_i a$ for all $i \in N$ with $b(i) \neq a(i)$. Without externalities, $bI_i a$ for all $i \in N$ with $b(i) = a(i)$. Thus, a is not Pareto efficient.

he points to and removed from the market. If there is at least one remaining agent, proceed to the next step. Otherwise, stop.

Since there is a finite number of agents, the algorithm terminates within a finite number of steps. With strict preferences over houses, each agent at any step has a unique most preferred house. It is possible that multiple cycles form at a step, and in this case agents in all the cycles are removed simultaneously. Let K be the total number of steps in the algorithm, and let S^k be the set of agents removed at step k, for $k = 1, \dots, K$. The algorithm proceeds in a unique way, producing an allocation as well as a partition $\{S^1, \dots, S^K\}$ of N. We call the allocation obtained by the TTC algorithm the *TTC allocation* of the market $\langle N, H, R, e\rangle$. Note that any cycle formed at any step in the TTC algorithm is a trading cycle in the TTC allocation, and thus we call it a *top trading cycle*. We can regard S^k as the union of top trading cycles formed at step k.[10]

3.2 Properties of TTC Allocations

As already mentioned, when there are no externalities, the TTC allocation is the unique allocation in the core (which is also the unique competitive allocation). Our main result below presents the properties that the TTC allocation possesses when there are externalities.

Theorem 1. *Consider a housing market $\langle N, H, R, e\rangle$ with egocentric preferences. The TTC allocation of $\langle N, H, R, e\rangle$ is stable in $\langle N, H, R\rangle$ and is in the irreversible γ-core of $\langle N, H, R, e\rangle$. If R has aligned interests among externality creators, then the TTC allocation of $\langle N, H, R, e\rangle$ is the unique allocation that is stable in $\langle N, H, R\rangle$ and is in the irreversible γ-core of $\langle N, H, R, e\rangle$. The same statements hold when the irreversible γ-core is replaced with the δ-core.*

Theorem 1 shows that the TTC allocation is stable and is in the irreversible γ-core as well as in the δ-core, and that under a further preference restriction, it is the unique stable allocation in either of the two cores. This generalizes the existing result with egoistic preferences that the TTC allocation is the unique allocation in the core. With egoistic preferences, the aligned interests condition is satisfied trivially, the irreversible γ-core and the δ-core coincide with the core, and any allocation in the core is stable.

When there are no externalities and agents may be indifferent between distinct houses, the core defined by strong domination is nonempty and may contain unstable allocations, but it contains at least one stable allocation, which can be obtained by the TTC algorithm

[10] More precisely, a top trading cycle for a housing market can be defined as a trading cycle in which each agent obtains his favorite house. Then S^k is the union of top trading cycles for the housing market with the remaining agents at step k.

(see Shapley and Scarf, 1974; Roth and Postlewaite, 1977). When externalities are introduced, we obtain similar results even with strict preferences over own-allotments and a core notion defined by weak domination.

We illustrate the result of Theorem 1 with the two previous examples. First, consider Example 1. The TTC allocation is $a = (h_2, h_1, h_4, h_3)$. We have already seen that a is in the irreversible γ-core as well as in the δ-core. Since it is not in the γ-core, we cannot strengthen Theorem 1 by replacing the irreversible γ-core with the γ-core. We can also check that a is stable because any agent receiving an allotment different from that at a gets worse off. Next, consider Example 2. The TTC allocation is $a = (h_2, h_3, h_1)$. It can be checked that both a and $b = (h_2, h_1, h_3)$ are in the irreversible γ-core[11] as well as in the δ-core. Also, a is stable while b is not. So we can interpret that the TTC algorithm selects a stable allocation among allocations in the irreversible γ-core or in the δ-core. However, there does not need to be a unique stable allocation in the irreversible γ-core or in the δ-core, as shown by the following example.

Example 3. Consider a housing market with $N = \{i_1, i_2, i_3\}$, $H = \{h_1, h_2, h_3\}$, and $e = (h_1, h_2, h_3)$. Suppose that the agents have egocentric preferences, that their associated preferences over own-allotments are given by $\tilde{P}_{i_1} : h_2, h_3, h_1$, $\tilde{P}_{i_2} : h_1, h_3, h_2$, $\tilde{P}_{i_3} : h_2, h_1, h_3$ and that agent i_2's preferences satisfy $(h_3, h_1, h_2) P_{i_2} (h_2, h_1, h_3)$. The TTC allocation is $a = (h_2, h_1, h_3)$, and thus a is stable in $\langle N, H, R \rangle$ and is in the irreversible γ-core as well as in the δ-core of $\langle N, H, R, e \rangle$ by Theorem 1. Let $b = (h_3, h_1, h_2)$. It can be checked that b is also stable and is in the irreversible γ-core as well as in the δ-core. Hence, there can be multiple stable allocations that are in the irreversible γ-core and in the δ-core.

In all the three examples so far, a problem arises when there is an agent who benefits from others' reallocating their allotments while maintaining his allotment. In the first two examples, the other agents do not gain by reallocating their allotments, and we can preclude such a reallocation by imposing irreversibility or stability requirements. In Example 3, however, there are conflicting interests among the other agents; agent i_3 benefits from moving from a to b while agent i_1 loses. In this situation, we cannot preclude one or the other of a and b because there will be at least one agent who objects to a change, and it creates the possibility of multiple stable allocations in the irreversible γ-core and in the δ-core.

In Theorem 1, in order to obtain a uniqueness result, we impose a further restriction that the preference profile has aligned interests among externality creators. To get a sense of the restriction, consider the preference relations over own-allotments in Example 2. Agents i_1 and i_2 have common interests in reallocating h_2 and h_3 between themselves: both prefer

[11] The allocation a is also in the γ-core, while b is in the γ-core if $(h_2, h_1, h_3) R_{i_3} (h_1, h_2, h_3)$.

$(a(i_1), a(i_2)) = (h_2, h_3)$ to (h_3, h_2). In this case, agent i_3 can have a strict preference between (h_2, h_3, h_1) and (h_3, h_2, h_1). In contrast, agents i_1 and i_2 have conflicting interests in reallocating h_1 and h_3 because both prefer h_3 to h_1. Then agent i_3 should be indifferent between (h_1, h_3, h_2) and (h_3, h_1, h_2). In this way, we can test whether a given preference profile has aligned interests among externality creators. The aligned interests condition limits the extent to which a group of agents exerts externalities to the other agents.[12] For instance, it precludes the preference profile in Example 3: by reallocating their allotments from $a = (h_2, h_1, h_3)$, agents i_1 and i_3 can make agent i_2 better off, but they have conflicting interests regarding the change.

The role of the aligned interests condition in obtaining the uniqueness result can be interpreted as follows. Note that TTC allocations and stable allocations depend only on preferences $\tilde{R}$ over own-allotments while the irreversible γ-core and the δ-core use preferences R over allocations. So in general, the equivalence between the TTC allocation and a stable allocation in the irreversible γ-core (and in the δ-core) cannot be expected. But under the aligned interests condition, it suffices to know $\tilde{R}$ to determine whether a stable allocation is in the irreversible γ-core or not. Suppose that an allocation a is stable in $\langle N, H, R\rangle$ and that an allocation b irreversibly γ-dominates a via coalition T in $\langle N, H, R, e\rangle$. Let $S = \{i \in N : b(i) \neq a(i)\}$. First, consider the case where bP_ia for some $i \in S$ and aP_ib for some $i \in S$. Then by the aligned interests condition, bI_ia for all $i \notin S$, and thus in this case, only the preference profile $\tilde{R}$ over own-allotments plays a role. Next, consider the case where bP_ia for all $i \in S$ or aP_ib for all $i \in S$. Since a is stable, the possibility that bP_ia for all $i \in S$ is excluded. Then the blocking coalition T does not overlap with S, and having aP_ib for all $i \in S$ violates the irreversibility requirement. Hence, the second case cannot arise. If a coalition T δ-blocks a via b, then T must contain an agent i such that $b(i) \neq a(i)$. Thus, we can use a similar argument to show that, under the aligned interests condition, it suffices to know $\tilde{R}$ to determine whether a stable allocation is in the δ-core or not.

The aligned interests condition does not imply that the irreversible γ-core and the δ-core are singletons. In Example 2, a preference profile over allocations can be constructed to satisfy the aligned interests condition, but the two allocations a and b are in the irreversible γ-core and in the δ-core. So we need to impose stability in order to obtain the uniqueness result.

In our next result, we characterize the set of TTC allocations obtained by varying the endowment allocation e for a given house allocation problem $\langle N, H, R\rangle$. It is known that, with egoistic preferences, the set of TTC allocations is equal to the set of Pareto efficient

[12] Roughly speaking, under the aligned interests condition, it is easier to have externalities when agents have diverse preferences over own-allotments. In the extreme case where every agent has identical preferences over own-allotments, the aligned interests condition permits no externalities. It can also be shown that the condition precludes the case where every agent has strict preferences over allocations.

allocations (see Lemma 1 of Abdulkadiroğlu and Sönmez, 1998). However, we can expect that this result does not generalize to the case of egocentric preferences for the following two reasons. First, TTC allocations depend only on preferences $\tilde{R}$ over own-allotments, whereas Pareto efficiency requires the knowledge of preferences R over allocations. Second, any TTC allocation is stable as shown in Theorem 1, while the set of stable allocations can be strictly smaller than that of Pareto efficient allocations as shown in Example 2. The following result shows that we can generate any stable allocation as a TTC allocation for some endowment allocation.

Proposition 1. *Consider a house allocation problem $\langle N, H, R \rangle$ with egocentric preferences. An allocation $a \in \mathcal{A}$ is stable in $\langle N, H, R \rangle$ if and only if it is the TTC allocation of $\langle N, H, R, e \rangle$ for some endowment allocation $e \in \mathcal{A}$.*

From Proposition 1, we can see that TTC allocations are more closely related to stable allocations than to Pareto efficient allocations. As mentioned following Example 2, the concept of Pareto efficiency is too weak when there are externalities. In the proof of Proposition 1, in order to show that any stable allocation a is a TTC allocation for some endowment allocation, we set the endowment allocation as a itself. Abdulkadiroğlu and Sönmez (1998) take exactly the same approach to prove that any Pareto efficient allocation is in the core for some endowment allocation when there are no externalities.

4 TTC Mechanism and Strategy-Proofness

In this section, we fix N, H, and e, and consider a scenario where each agent i reports his preferences R_i and an allocation is chosen based on agents' reports. Let $\mathcal{R}$ and $\tilde{\mathcal{R}}$ be the set of all egocentric and egoistic preference relations, respectively, on $\mathcal{A}$. A *mechanism* is a rule that selects an allocation for each egocentric preference profile. That is, a mechanism can be represented by a mapping $\varphi : \mathcal{R}^N \to \mathcal{A}$. We can consider a mechanism that selects the TTC allocation for each preference profile. We call it the *TTC mechanism* and denote it by φ^T. Note that the TTC mechanism utilizes only preferences over own-allotments, and so we can think of $\tilde{\mathcal{R}}^N$ instead of $\mathcal{R}^N$ as the domain of φ^T.[13] We can also consider a mechanism that selects an allocation in the core (resp. the irreversible γ-core and the δ-core) for each preference profile and call it a core mechanism (resp. an irreversible γ-core mechanism and a δ-core mechanism). With egoistic preferences, a core mechanism is unique, coinciding with the TTC mechanism. However, with egocentric preferences, there can be multiple allocations in the irreversible γ-core and in the δ-core, and thus there can be an irreversible γ-core mechanism and a δ-core mechanism different from the TTC mechanism.

[13] This feature can be regarded as an informational advantage of the TTC mechanism, as it suffices for agents to report their preferences over own-allotments rather than those over allocations.

For any coalition $T \subseteq N$, let $R_T = (R_i)_{i \in T} \in \mathcal{R}^T$ and $R_{-T} = (R_i)_{i \notin T} \in \mathcal{R}^{N \setminus T}$ denote the preference profiles of agents in T and those not in T, respectively. When T consists of a single agent i, we write R_{-i} instead of $R_{-\{i\}}$. A mechanism φ is *strategy-proof* if for any $R \in \mathcal{R}^N$, for any $i \in N$, and for any $R'_i \in \mathcal{R}$, $\varphi(R) R_i \varphi(R'_i, R_{-i})$. That is, when a mechanism is strategy-proof, no agent can gain by misreporting his preferences. A mechanism φ is *coalitionally strategy-proof* if for any $R \in \mathcal{R}^N$ and for any nonempty $T \subseteq N$, there exists no $R'_T \in \mathcal{R}^T$ such that $\varphi(R'_T, R_{-T}) R_i \varphi(R)$ for all $i \in T$ and $\varphi(R'_T, R_{-T}) P_i \varphi(R)$ for some $i \in T$. In other words, when a mechanism is coalitionally strategy-proof, no coalition of agents can make at least one member better off and none worse off by jointly misreporting their preferences. Obviously, if a mechanism is coalitionally strategy-proof, it is strategy-proof.

When the domain of preferences consists of egoistic preferences, it has been shown that the TTC mechanism is strategy-proof (Roth, 1982) and coalitionally strategy-proof (Moulin, 1995, Lemma 3.3).[14] In the following proposition, we extend these existing results to the case where the domain of preferences consists of egocentric preferences.

Proposition 2. *The TTC mechanism φ^T is coalitionally strategy-proof.*

Roth (1982) proves that no agent can obtain a more preferred house by misreporting his preferences in the TTC mechanism. Similarly, Moulin (1995, Lemma 3.3) shows that, when a group of agents misreports their preferences in the TTC mechanism, no member can obtain a more preferred house without making another member receive a less preferred house. Thus, there is no profitable manipulation by an individual agent or a group of agents in which a misreporting agent receives a better house. Without externalities, this is sufficient for (coalitional) strategy-proofness. However, when there are externalities, there is another kind of potentially profitable manipulation in which misreports by an agent or a group of agents influence others' allotments while misreporting agents receive the same allotments. In the proof of Proposition 2, we show that this second kind of manipulation cannot be profitable. When a group of agents misreports their preferences in the TTC mechanism in a way that they maintain the same allotments, the allotments of the remaining agents are not affected either.[15]

We say that an allocation $a \in \mathcal{A}$ is *individually rational* in the housing market $\langle N, H, R, e \rangle$ if $a(i) \tilde{R}_i e(i)$ for all $i \in N$.[16] Note that individual rationality can be checked only with pref-

[14] Bird (1984) proves that the TTC mechanism is coalitionally strategy-proof in a weaker sense that there is no joint misreport that makes every member better off, but in a more general setting where indifferences between distinct houses are allowed. Roth (1982) also considers a setting that allows indifferences.

[15] A mechanism is called *nonbossy* if no agent can influence others' allotments without changing his own allotment (Satterthwaite and Sonnenschein, 1981). Miyagawa (2002) argues that the TTC mechanism is nonbossy. It is easy to see that, if a mechanism is nonbossy, no group of agents can influence the outsiders' allotments without changing the members' allotments.

[16] Note that this is weaker than the condition that $a R_i e$ for all $i \in N$. Any allocation a in the γ-core

erences over own-allotments. A mechanism φ is *individually rational* if the allocation $\varphi(R)$ is individually rational for every $R \in \mathcal{R}^N$. A mechanism φ is *stable* if the allocation $\varphi(R)$ is stable for every $R \in \mathcal{R}^N$.

Ma (1994) shows that, when the preference domain is egoistic, a mechanism φ is individually rational, Pareto efficient, and strategy-proof if and only if $\varphi = \varphi^T$. As we have observed in Proposition 1, stability is a more relevant concept than Pareto efficiency for TTC allocations under egocentric preferences. The next proposition shows that the same result extends to the egocentric preference domain when we replace Pareto efficiency with stability.

Proposition 3. *A mechanism φ is individually rational, stable, and strategy-proof if and only if it is the TTC mechanism φ^T.*

The key idea in proving Proposition 3 is as follows. Suppose that an individually rational and stable mechanism selects an allocation b different from the TTC allocation a for some preference profile. Then there must be some agent i who ranks $b(i)$ between $a(i)$ and $e(i)$. Such an agent can obtain $a(i)$ by reporting that he ranks $e(i)$ immediately after $a(i)$, which makes the mechanism selecting b not strategy-proof. Hence, in order for a mechanism to satisfy all the three properties, it must be the TTC mechanism. In particular, any stable and irreversible γ-core (or δ-core) mechanism different from the TTC mechanism is not strategy-proof. Thus, selecting TTC allocations is appealing from an incentive point of view in the presence of externalities where there can be multiple stable allocations in the irreversible γ-core and in the δ-core.

Sönmez (1999) considers a class of indivisible goods allocation problems, including housing markets, where agents have preferences over allocations. He shows that, if there exists a mechanism that is individually rational, Pareto efficient, and strategy-proof, then every agent is indifferent between any two allocations in the core and the mechanism selects an allocation in the core whenever the core is nonempty. Ehlers (2014) strengthens Sönmez's (1999) result by showing that the same result holds with the γ-core and that the γ-core is the largest core concept for such a result to hold. In our setting, there exists a mechanism that is individually rational, stable (thus Pareto efficient), and strategy-proof, as shown in Proposition 3, while there can be multiple allocations in the γ-core between which some agent is not indifferent, as can be seen in Example 3.[17] The results of Sönmez (1999) and

satisfies the latter condition, but there can be some allocation a in the irreversible γ-core or in the δ-core that violates the condition. For example, suppose that agents i_1, i_2, and i_3 own houses h_1, h_2, and h_3, respectively, and that their most preferred houses are h_2, h_1, and h_3, respectively. Then the TTC allocation is $a = (h_2, h_1, h_3)$, but it is possible that agent i_3 prefers $e = (h_1, h_2, h_3)$ to a. If we impose Assumption A of Sönmez (1999) (i.e., aI_ie if and only if $a(i) = e(i)$) in our model, then the two conditions are equivalent.

[17] In Example 3, the allocation $b = (h_3, h_1, h_2)$ is in the γ-core, while the TTC allocation $a = (h_2, h_1, h_3)$ is in the γ-core if agent i_3 weakly prefers a to $e = (h_1, h_2, h_3)$.

Ehlers (2014) do not apply to our model because the domain of egocentric preferences does not satisfy the conditions in their models (called Assumptions A and B). Thus, we provide an alternative model where it is possible to obtain an individually rational, Pareto efficient, and strategy-proof mechanism.

Remark 2. A widely studied class of mechanisms for house allocation problems is the class of simple serial dictatorships (Abdulkadiroğlu and Sönmez, 1998). We can apply a simple serial dictatorship to a house allocation problem with egocentric preferences using the associated preferences over own-allotments. Since the set of allocations obtained from simple serial dictatorships is equal to that of TTC allocations (Lemma 1 of Abdulkadiroğlu and Sönmez, 1998), any allocation obtained from a simple serial dictatorship is stable. It is easy to see that any simple serial dictatorship, as a mechanism, is strategy-proof. Also, when a group of agents misreports their preferences, no member can obtain a more preferred house without having another member receive a less preferred house. This together with nonbossiness proves that the serially dictatorial mechanism is coalitionally strategy-proof as well.

5 Conclusion

In this paper, we have studied housing markets with externalities. We have restricted preferences to egocentric preferences and proposed two solution concepts of stable allocations in the irreversible γ-core and those in the δ-core. Using the TTC algorithm, we have shown that a solution always exists and that it is unique under a further preference restriction. We have also established that we can generate any stable allocation as the TTC allocation for some endowment allocation, that the TTC mechanism is coalitionally strategy-proof, and that a mechanism is individually rational, stable, and strategy-proof if and only if it is the TTC mechanism. Our results provide desirable properties of the TTC algorithm, extending the existing results to the case of egocentric preferences. Although our analysis represents a significant step toward understanding housing markets with externalities, further studies on this topic are needed for a more comprehensive understanding. We mention two directions for future research. First, we can relax the restriction of egocentric preferences so that we can cover situations where the size of externalities is large. Second, instead of assuming that the residual agents react passively when a coalition deviates, we can model that they also behave rationally and strategically (for example, as in the recursive core proposed by Kóczy, 2007).

A Proofs

Proof of Theorem 1

Proof. Let $a \in \mathcal{A}$ be the TTC allocation of $\langle N, H, R, e\rangle$, and let $\{S^1, \ldots, S^K\}$ be the partition of N obtained from the TTC algorithm.

First, we show that a is stable in $\langle N, H, R\rangle$. Suppose to the contrary that there exists an allocation b such that $b \neq a$ and $b(i)\tilde{P}_i a(i)$ for all $i \in T := \{i \in N : b(i) \neq a(i)\}$. Let k' be the smallest k such that $S^k \cap T \neq \varnothing$. Choose any $i' \in S^{k'} \cap T$. Then $b(i')\tilde{P}_{i'} a(i')$. Since $a(i')$ is agent i''s most preferred house at step k', $i'' := b(i')$ belongs to $S^{k''}$ for some $k'' < k'$.[18] Since $i'' \in S^{k''}$, there is some $j \in S^{k''}$ such that $a(j) = i''$. Since $S^{k''} \cap T = \varnothing$, we have $b(j) = a(j)$. Then we have $b(i') = i'' = b(j)$ and $i' \neq j$, contradicting the one-to-one property of the allocation b.

Next, we show that a is in the irreversible γ-core of $\langle N, H, R, e\rangle$. Suppose to the contrary that there exists an allocation b and coalition T such that b irreversibly γ-dominates a via T in $\langle N, H, R, e\rangle$. Consider the case where there exists $i \in T$ such that $b(i) \neq a(i)$. Let $T' = \{i \in T : b(i) \neq a(i)\}$. Then T' is nonempty, and let k' be the smallest k such that $S^k \cap T' \neq \varnothing$. Choose any $i' \in S^{k'} \cap T'$. Since $b(i') \neq a(i')$ and $bR_{i'}a$, we have $b(i')\tilde{P}_{i'} a(i')$. Since $a(i')$ is agent i''s most preferred house at step k' in the TTC algorithm, $i'' := b(i')$ belongs to $S^{k''}$ for some $k'' < k'$. Since $b(T) = e(T)$ and $i' \in T$, we have $i'' \in T$ and thus $b(i'') = a(i'')$. Starting from $i'' \in S^{k''} \cap T$, we can trace the owners of assigned houses along the top trading cycle in $S^{k''}$, where each agent in the cycle belongs to $S^{k''} \cap T$. Hence, there is some $j \in S^{k''} \cap T$ such that $a(j) = i''$. Then $b(j) = a(j)$, and we have $b(i') = i'' = b(j)$ and $i' \neq j$, contradicting the one-to-one property of the allocation b.

Now consider the case where $b(i) = a(i)$ for all $i \in T$. Since $bP_i a$ for some $i \in T$, we have $b \neq a$, and thus there is $i \notin T$ such that $b(i) \neq a(i)$. Choose any $i' \in T^c$ such that $b(i') \neq a(i')$. Note that $b(i') = e(i')$, and thus $a(i') \neq e(i')$. Let k' be the step at which agent i' is removed, i.e., $i' \in S^{k'}$. Since agent i' points to $a(i')$ at step k' while $e(i')$ is available, we have $a(i')\tilde{P}_{i'} e(i')$. Similarly, for any other agent i in the same top trading cycle as i', we have $a(i) \neq e(i)$ and thus $a(i)\tilde{P}_i e(i)$. Since $b(T) = e(T)$ and $b(i) = a(i)$ for all $i \in T$, trade in a occurs among agents in T and among those not in T. Hence, $S^{k'} \subseteq T^c$. Let S be the union of top trading cycles that are outside of T and consist of more than one agent. Then S is nonempty since $i' \in S$, and $a(i)\tilde{P}_i[e(i) = b(i)]$ for all $i \in S$. This implies $aP_i b$ for all $i \in S$, while $b(i) = a(i)$ for all $i \notin S$. Hence, b is γ-dominated by a via $S \subseteq T^c$ in $\langle N, H, R, b\rangle$, a contradiction.

We show that a is in the δ-core of $\langle N, H, R, e\rangle$. Suppose to the contrary that there

[18] With an abuse of notation, for any allocation $a \in \mathcal{A}$, we sometimes use $a(i)$ to refer to the owner of $a(i)$, i.e., $e^{-1}(a(i))$, when there is no ambiguity about the endowment allocation e.

exists an allocation b and coalition T such that b δ-dominates a via T in $\langle N, H, R, e\rangle$. If $b(i) = a(i)$ for all $i \in T$, we have $b = a$, which is a contradiction. Thus, there exists $i \in T$ such that $b(i) \neq a(i)$, and we can apply the same logic as in the first case above to derive a contradiction.

Suppose that R has aligned interests among externality creators. We show that a is the unique allocation that is stable in $\langle N, H, R\rangle$ and is in the irreversible γ-core of $\langle N, H, R, e\rangle$. Suppose to the contrary that there is an allocation b, different from a, having the two properties. Let $T = \{i \in N : b(i) \neq a(i)\}$. Note that T is nonempty because $b \neq a$. Let k' be the smallest k such that $S^k \cap T \neq \varnothing$. Choose any $i' \in S^{k'} \cap T$. Since $b(i) = a(i)$ for all $i \in S^1 \cup \cdots \cup S^{k'-1}$, we have $a(i')\tilde{R}_{i'}b(i')$. Since $b(i') \neq a(i')$, we have $aP_{i'}b$. Suppose that aP_ib for all $i \in T$. Then T γ-blocks b via a in $\langle N, H, R, b\rangle$, contradicting the stability of b. So there exists $i'' \in T$ such that $bP_{i''}a$. Note that $i'' \in S^k$ for some $k > k'$. By the construction of the TTC algorithm, we have $a(i'')\tilde{R}_{i''}e(i'')$, and thus $b(i'')\tilde{P}_{i''}e(i'')$. Define a function $a' : N \to H$ by $a'(i) = a(i)$ for all $i \in S^{k'}$ and $a'(i) = e(i)$ for all $i \notin S^{k'}$. Since $a(S^{k'}) = e(S^{k'})$, the function a' is an allocation. Since $aP_{i'}b$, we have $a'P_{i'}b$. Since $i'' \notin S^{k'}$, we have $b(i'')\tilde{P}_{i''}[e(i'') = a'(i'')]$ and thus $bP_{i''}a'$. Let $S = \{i \in N : b(i) \neq a'(i)\}$. Then $i', i'' \in S$. Since R has aligned interests among externality creators, we have $a'I_ib$ for all $i \notin S$. Then $a'I_ib$ for all $i \in S^{k'}$ such that $b(i) = a(i)$, while $a'P_ib$ for all $i \in S^{k'}$ such that $b(i) \neq a(i)$. Also, we have $i' \in S^{k'}$ with $b(i') \neq a'(i')$. Thus, $S^{k'}$ irreversibly γ-blocks b via a' in $\langle N, H, R, e\rangle$, which is a contradiction.

Finally, we show that a is the unique allocation that is stable in $\langle N, H, R\rangle$ and is in the δ-core of $\langle N, H, R, e\rangle$. Suppose to the contrary that there is an allocation b, different from a, having the two properties, and define T and k' and choose i' as in the previous paragraph. Define a function $a' : N \to H$ by $a'(i) = a(i)$ for all $i \in S^{k'}$, $a'(i) = e(i)$ for all $i \in (\cup_{j \in S^{k'}} S_j^{b,e}) \setminus S^{k'}$, and $a'(i) = b(i)$ for all $i \notin \cup_{j \in S^{k'}} S_j^{b,e}$. Since $a(S^{k'}) = e(S^{k'})$, the function a' is an allocation, which can be obtained by a deviation of $S^{k'}$ from b in the δ-model. We consider two cases. First, suppose that $(\cup_{j \in S^{k'}} S_j^{b,e}) \setminus S^{k'}$ is nonempty, and choose any i'' in the set. Note that $i'' = b(i)$ for some $i \in S^{k'}$, and thus $b(i'') \neq e(i'')$. Since b is in the δ-core, we have $b(i'')\tilde{P}_{i''}[e(i'') = a'(i'')]$ and thus $bP_{i''}a'$. Since $a'P_{i'}b$ and R has aligned interests among externality creators, we have $a'I_ib$ for all $i \in N$ such that $b(i) = a'(i)$. Then $S^{k'}$ δ-blocks b via a' in $\langle N, H, R, e\rangle$, which is a contradiction. Second, suppose that $(\cup_{j \in S^{k'}} S_j^{b,e}) \setminus S^{k'}$ is empty. Then $\{i \in N : b(i) \neq a'(i)\} = S^{k'} \cap T$. Since $S^{k'} \cap T$ is nonempty and $a'P_ib$ for all $i \in S^{k'} \cap T$, the allocation b is not stable, which is a contradiction. $\square$

Proof of Proposition 1

Proof. The "if" part follows from Theorem 1.

To prove the "only if" part, suppose that an allocation a is stable in $\langle N, H, R\rangle$. By definition, a is in the irreversible γ-core of $\langle N, H, R, a\rangle$. We show that there is no other allocation in the irreversible γ-core of $\langle N, H, R, a\rangle$. Suppose to the contrary that there is an allocation b, different from a, in the irreversible γ-core of $\langle N, H, R, a\rangle$. Let $T = \{i \in N : b(i) \neq a(i)\}$. Note that T is nonempty. Since a is stable, there exists $i' \in T$ such that $a(i')\tilde{P}_{i'}b(i')$. Then the coalition $\{i'\}$ irreversibly γ-blocks b via a in $\langle N, H, R, a\rangle$ with $b(i') \neq a(i')$. This contradicts the assumption that b is in the irreversible γ-core of $\langle N, H, R, a\rangle$. Since the TTC algorithm yields a stable allocation in the irreversible γ-core and a is the only such allocation, a is the TTC allocation of $\langle N, H, R, a\rangle$. $\square$

Proof of Proposition 2

We first strengthen Lemma 3 of Roth (1982) as follows.

Lemma 1. *Fix any $\tilde{R} \in \tilde{\mathcal{R}}^N$ and any $i \in N$. Let $a = \varphi^T(\tilde{R})$. Let $\tilde{R}'_i \in \tilde{\mathcal{R}}$ be any preference relation over own-allotments such that $a(i)\tilde{P}'_i h$ for all $h \neq a(i)$. Let $a' = \varphi^T(\tilde{R}'_i, \tilde{R}_{-i})$. Then $a = a'$.*[19]

Proof. For simplicity, let $\tilde{R}' = (\tilde{R}'_i, \tilde{R}_{-i})$. Let k and k' be the steps at which agent i is removed in the TTC algorithm for preference profiles $\tilde{R}$ and $\tilde{R}'$ (written as $T(\tilde{R})$ and $T(\tilde{R}')$), respectively.

Suppose that $k < k'$. Then by Lemma 2 of Roth (1982), $T(\tilde{R})$ and $T(\tilde{R}')$ have the same top trading cycles at any step $l < k$. Let $(j_1, \ldots, j_m, i)$ be the top trading cycle containing agent i formed at step k of $T(\tilde{R})$. Then $(j_1, \ldots, j_m, i)$ is a chain at step k of $T(\tilde{R}')$.[20] Since $a(i) = j_1$, agent i's most preferred house in $\tilde{R}'_i$ is owned by j_1, and thus agent i points to j_1 and is removed at step k of $T(\tilde{R}')$. This is a contradiction. Hence, we have $k \geq k'$.

Again, by Lemma 2 of Roth (1982), $T(\tilde{R})$ and $T(\tilde{R}')$ have the same top trading cycles at any step $l < k'$. Suppose that $k = k'$. Then $T(\tilde{R})$ and $T(\tilde{R}')$ have the same top trading cycles at step k' and at any step $l > k'$. Hence, we have $a = a'$. Now suppose that $k > k'$. Let $(j_1, \ldots, j_m, i)$ be the top trading cycle containing agent i formed at step k' of $T(\tilde{R}')$. Since every agent removed before step k' receives the same allotment at a and a', $a(i)$ is available at step k' of $T(\tilde{R}')$, and thus we have $j_1 = a'(i) = a(i)$. Then $(j_1, \ldots, j_m, i)$ is a chain at step k' of $T(\tilde{R})$, and by Lemma 1 of Roth (1982), it is a chain and thus a cycle at step k of $T(\tilde{R})$.

Consider any top trading cycle formed at step l of $T(\tilde{R})$, where $k' \leq l < k$. Since it does not contain any of the agents in $(j_1, \ldots, j_m, i)$, it is also a cycle at step l of $T(\tilde{R}')$. If there

[19] Lemma 3 of Roth (1982) concludes that $a(i) = a'(i)$.

[20] We say that a sequence $(i_1, \ldots, i_m)$ of agents is a *chain* at a step of the TTC algorithm if agent i_l points to agent i_{l+1} for all $l = 1, \ldots, m-1$ at that step.

is any top trading cycle other than $(j_1, \ldots, j_m, i)$ at step k of $T(\tilde{R})$, then it is also a cycle at step k of $T(\tilde{R}')$, and $T(\tilde{R})$ and $T(\tilde{R}')$ proceed in the same way afterwards. If $(j_1, \ldots, j_m, i)$ is the only top trading cycle at step k of $T(\tilde{R})$, then any cycle at step l of $T(\tilde{R})$ is formed at step $l-1$ of $T(\tilde{R}')$ for any $l > k$. In either case, top trading cycles formed in $T(\tilde{R})$ and $T(\tilde{R}')$ are the same, resulting in $a = a'$. □

Now we prove Proposition 2.

Proof. Suppose to the contrary that the TTC mechanism φ^T is not coalitionally strategy-proof. Then there exists $R \in \mathcal{R}$, nonempty $T \subseteq N$, and $R'_T \in \mathcal{R}^T$ such that $\varphi^T(R'_T, R_{-T}) R_i \varphi^T(R)$ for all $i \in T$ and $\varphi^T(R'_T, R_{-T}) P_i \varphi^T(R)$ for some $i \in T$. For simplicity, let $R' = (R'_T, R_{-T})$. Let $\tilde{R}$ and $\tilde{R}'$ be the preference profiles over own-allotments associated with R and R', respectively. Let $a = \varphi^T(\tilde{R})$ and $a' = \varphi^T(\tilde{R}')$. By applying Lemma 1 repeatedly to $\tilde{R}'$, we can assume without loss of generality that agent i's most preferred house in $\tilde{R}'_i$ is $a'(i)$ for all $i \in T$.

First, suppose that $a(i) = a'(i)$ for all $i \in T$. Then starting from $\tilde{R}$, we can obtain $\tilde{R}'$ by changing $\tilde{R}_i$ into $\tilde{R}'_i$ one by one over all $i \in T$. For any $i \in T$, $a(i) = a'(i)$, and thus agent i's most preferred house in $\tilde{R}'_i$ is $a(i)$. Then, by Lemma 1, the TTC allocation does not change as we replace $\tilde{R}_i$ by $\tilde{R}'_i$ for each $i \in T$. By applying this argument repeatedly over all $i \in T$, we obtain $a = a'$, which contradicts $a' P_i a$ for some $i \in T$.

Next, suppose that $a(i) \neq a'(i)$ for some $i \in T$. Let $\{S^1(\tilde{R}), \ldots, S^K(\tilde{R})\}$ and $\{S^1(\tilde{R}'), \ldots, S^{K'}(\tilde{R}')\}$ be the partitions of N obtained from $T(\tilde{R})$ and $T(\tilde{R}')$, respectively. Let $\bar{S}^k(\tilde{R}) = S^1(\tilde{R}) \cup \cdots \cup S^k(\tilde{R})$ for all $k = 1, \ldots, K$, and let $\bar{S}^k(\tilde{R}') = S^1(\tilde{R}') \cup \cdots \cup S^k(\tilde{R}')$ for all $k = 1, \ldots, K'$. Let $T' = \{i \in T : a(i) \neq a'(i)\}$. Then T' is nonempty, and let k' be the smallest k such that $S^k(\tilde{R}) \cap T' \neq \varnothing$. Choose any $i' \in S^{k'}(\tilde{R}) \cap T'$. Since $a(i') \neq a'(i')$ and $a' R_{i'} a$, we have $a'(i') \tilde{P}_{i'} a(i')$. Since $a(i')$ is agent i''s most preferred house at step k' of $T(\tilde{R})$, we have $a'(i') \in \bar{S}^{k'-1}(\tilde{R})$. Then there exists $j \in \bar{S}^{k'-1}(\tilde{R})$ such that $a'(j) \notin \bar{S}^{k'-1}(\tilde{R})$, because otherwise there cannot be a top trading cycle containing the arc $(i', a'(i'))$ in $T(\tilde{R}')$. Below we show that $a(i) = a'(i)$ for all $i \in \bar{S}^{k'-1}(\tilde{R})$. Then, since $j \in \bar{S}^{k'-1}(\tilde{R})$, we have $a(j) = a'(j)$ and $a(j) \in \bar{S}^{k'-1}(\tilde{R})$, which contradicts $a'(j) \notin \bar{S}^{k'-1}(\tilde{R})$, and we are done.

To show that $a(i) = a'(i)$ for all $i \in \bar{S}^{k'-1}(\tilde{R})$, we prove by induction that every top trading cycle in $\bar{S}^k(\tilde{R})$ is a top trading cycle in $\bar{S}^k(\tilde{R}')$ for all $k = 1, \ldots, k'-1$. Note that $a(i) = a'(i)$ for all $i \in \bar{S}^{k'-1}(\tilde{R}) \cap T$, and thus agent i's most preferred house in $\tilde{R}'_i$ is $a(i)$ for all $i \in \bar{S}^{k'-1}(\tilde{R}) \cap T$. First, consider any top trading cycle $(j_1, \ldots, j_m)$ in $\bar{S}^1(\tilde{R})$. In $\tilde{R}'$, agent j_l's most preferred house is owned by $j_{l+1} = a(j_l)$, regardless of whether $j_l \in T$ or not, for all $l = 1, \ldots, m$, where we take $j_{m+1} = j_1$. Hence, $(j_1, \ldots, j_m)$ is a top trading cycle in $\bar{S}^1(\tilde{R}')$. Next, suppose that every top trading cycle in $\bar{S}^k(\tilde{R})$ is a top trading cycle in $\bar{S}^k(\tilde{R}')$, and consider any top trading cycle $(j_1, \ldots, j_m)$ in $S^{k+1}(\tilde{R})$, where $k+1 \leq k'-1$. Let agent j_l

be an agent who is removed first in $T(\tilde{R}')$ among those in $(j_1, \dots, j_m)$. If $j_l \in T$, agent j_l is assigned the house owned by $j_{l+1} = a(j_l)$ in $T(\tilde{R}')$ since $a(j_l) = a'(j_l)$. Suppose that $j_l \notin T$. If $a'(j_l)\tilde{P}_{j_l}a(j_l)$, then $a'(j_l) \in \bar{S}^k(\tilde{R})$, but this is not possible because $a(i) = a'(i)$ for all $i \in \bar{S}^k(\tilde{R})$ by the induction hypothesis. Thus, if $j_l \notin T$, then $a(j_l)\tilde{R}_{j_l}a'(j_l)$. Since agent j_l is an agent who is removed first in $T(\tilde{R}')$ among those in $(j_1, \dots, j_m)$, $j_{l+1} = a(j_l)$ is available at the step where agent j_l is removed in $T(\tilde{R}')$, and thus $a'(j_l) = a(j_l)$. Hence, agent j_{l+1} is removed at the same step as agent j_l in $T(\tilde{R}')$, and applying the same argument to j_{l+1} and so on, we can see that all the agents in $(j_1, \dots, j_m)$ are removed simultaneously forming a top trading cycle in $T(\tilde{R}')$. Suppose that the agents in $(j_1, \dots, j_m)$ are available at step $k+1$ of $T(\tilde{R}')$. If $j_l \in T$, agent j_l's most preferred house in $\tilde{R}'_{j_l}$ is owned by $j_{l+1} = a(j_l)$. If $j_l \notin T$, agent j_l's most preferred house in $\tilde{R}_{j_l}$ among those owned by agents in $N \setminus \bar{S}^k(\tilde{R})$ is owned by $j_{l+1} = a(j_l)$. Since no agent in $\bar{S}^k(\tilde{R})$ is available at step $k+1$ of $T(\tilde{R}')$ by the induction hypothesis, $(j_1, \dots, j_m)$ is a top trading cycle in $S^{k+1}(\tilde{R}')$. This shows that $(j_1, \dots, j_m)$ is a top trading cycle at step $k+1$ or earlier of $T(\tilde{R}')$, i.e., it is in $\bar{S}^{k+1}(\tilde{R}')$. $\square$

Proof of Proposition 3

Proof. Theorem 1 and Proposition 2 imply that the TTC mechanism is individually rational, stable, and strategy-proof, which proves the "if" part.

To prove the "only if" part, we can apply the proof of Theorem 2 in Svensson (1999) using stability instead of Pareto efficiency. $\square$

References

[1] Abdulkadiroğlu, A. and T. Sönmez (1998), "Random Serial Dictatorship and the Core from Random Endowments in House Allocation Problems," *Econometrica* 66, 689–701.

[2] Abdulkadiroğlu, A. and T. Sönmez (2013), "Matching Markets: Theory and Practice," *Advances in Economics and Econometrics* 1, Tenth World Congress, 3–47.

[3] Baccara, M., A. İmrohoroğlu, A. J. Wilson, and L. Yariv (2012), "A Field Study on Matching with Network Externalities," *American Economic Review* 102, 1773–1804.

[4] Bando, K. (2012), "Many-to-One Matching Markets with Externalities among Firms," *Journal of Mathematical Economics* 48, 14–20.

[5] Bando, K. (2014), "A Modified Deferred Acceptance Algorithm for Many-to-One Matching Markets with Externalities among Firms," *Journal of Mathematical Economics* 52, 173–181.

[6] Bando, K., R. Kawasaki, and S. Muto (2016), "Two-Sided Matching with Externalities: A Survey," *Journal of the Operations Research Society of Japan* 59, 35–71.

[7] Bird, C. G. (1984), "Group Incentive Compatibility in a Market with Indivisible Goods," *Economics Letters* 14, 309–313.

[8] Dutta, B. and J. Massó (1997), "Stability of Matchings When Individuals Have Preferences over Colleagues," *Journal of Economic Theory* 75, 464–475.

[9] Ehlers, L. (2014), "Strategy-Proofness and Essentially Single-Valued Cores Revisited," working paper.

[10] Fisher, J. and I. Hafalir (2016), "Matching with Aggregate Externalities," *Mathematical Social Sciences* 81, 1–7.

[11] Hart, S. and M. Kurz (1983), "Endogenous Formation of Coalitions," *Econometrica* 51, 1047–1064.

[12] Hylland, A. and R. Zeckhauser (1979), "The Efficient Allocation of Individuals to Positions," *Journal of Political Economy* 87, 293–314.

[13] Kawasaki, R. (2015), "Roth–Postlewaite Stability and von Neumann–Morgenstern Stability," *Journal of Mathematical Economics* 58, 1–6.

[14] Kóczy, L. Á. (2007), "A Recursive Core for Partition Function Form Games," *Theory and Decision* 63, 41–51.

[15] Li, S. (1993), "Competitive Matching Equilibrium and Multiple Principal-Agent Models," Discussion Paper No. 267, University of Minnesota.

[16] Ma, J. (1994), "Strategy-Proofness and the Strict Core in a Market with Indivisibilities," *International Journal of Game Theory* 23, 75–83.

[17] Miyagawa, E. (2002), "Strategy-Proofness and the Core in House Allocation Problems," *Games and Economic Behavior* 38, 347–361.

[18] Moulin, H. (1995), *Cooperative Microeconomics: A Game-Theoretic Introduction*, Princeton Univ. Press, Princeton, NJ.

[19] Mumcu, A, and I. Saglam (2007), "The Core of a Housing Market with Externalities," *Economics Bulletin* 3, 1–5.

[20] Mumcu, A. and I. Saglam (2010), "Stable One-to-One Matchings with Externalities," *Mathematical Social Sciences* 60, 154–159.

[21] Roth, A. E. (1982), "Incentive Compatibility in a Market with Indivisible Goods," *Economics Letters* 9, 127–132.

[22] Roth, A. E. and A. Postlewaite (1977), "Weak Versus Strong Domination in a Market with Indivisible Goods," *Journal of Mathematical Economics* 4, 131–137.

[23] Sasaki, H. and M. Toda (1996), "Two-Sided Matching Problems with Externalities," *Journal of Economic Theory* 70, 93–108.

[24] Satterthwaite, M. A. and H. Sonnenschein (1981), "Strategy-Proof Allocation Mechanisms at Differentiable Points," *Review of Economic Studies* 48, 587–597.

[25] Shapley, L. S. and H. E. Scarf (1974), "On Cores and Indivisibility," *Journal of Mathematical Economics* 1, 23–37.

[26] Sönmez, T. (1999), "Strategy-Proofness and Essentially Single-Valued Cores," *Econometrica* 67, 677–689.

[27] Svensson, L.-G. (1999), "Strategy-Proof Allocation of Indivisible Goods," *Social Choice and Welfare* 16, 557–567.