Robust Minimal Instability of the Top Trading Cycles Mechanism[*]

Battal Doğan[†] Lars Ehlers[‡]

March 14, 2020

Abstract

In the context of priority-based resource allocation, we formulate methods to compare assignments in terms of their stability as binary relations (on the set of possible assignments) that depend on the preference and the priority profile. We introduce three basic properties, *stability preferred*, *separability*, and *consistency*, that a reasonable stability comparison should satisfy. We show that, for any stability comparison satisfying the three properties, the top trading cycles (TTC) mechanism is minimally unstable among efficient and strategy-proof mechanisms in one-to-one matching. An important consequence is the robustness of a recent result by Abdulkadiroğlu et al. (2019), which uses a particular stability comparison method where an assignment is more stable than another assignment if the set of blocking pairs in the former assignment is a subset of the set of blocking pairs in the latter assignment. Our unifying approach covers basically all natural comparison methods and it includes many *cardinal* stability comparison methods as special cases.

[*] Battal Doğan acknowledges financial support from the British Academy/Leverhulme Trust (SRG1819\190133). Lars Ehlers acknowledges financial support from the SSHRC (Canada) and the FRQSC (Québec). We thank participants at several seminars and conferences for valuable comments and suggestions.

[†] Department of Economics, University of Bristol, 3B13 The Priory Road Complex, Priory Road, Clifton, BS8 1TU. Email: battal.dogan@bristol.ac.uk.

[‡] Département de Sciences Économiques and CIREQ, Université de Montréal, Montréal, QC H3C 3J7, Canada. Email: lars.ehlers@umontreal.ca.

1 Introduction

In many resource-allocation problems, each resource is endowed with an exogenously given priority ordering over the participants, and a mechanism elicits participants' preferences and allocates resources based on the reported preferences and the exogenous priorities. A well-known example is school choice where students report preference orderings over schools and each school is endowed with a capacity and a priority ordering over students. In a school choice problem, respecting preferences is captured by the efficiency requirement: an assignment is *efficient* if there is no other assignment at which a student is better off while no student is worse off. On the other hand, respecting priorities is captured by the stability requirement: an assignment is *stable* if it does not involve a "blocking pair" of a student and a school such that the student prefers the school to his assigned school and he has a higher priority than another student who is assigned to that school. Unfortunately, there exist school choice problems without an assignment that is both *efficient* and *stable* (Roth, 1982).

The seminal paper (Abdulkadiroğlu and Sönmez, 2003) introduces school choice and proposes to use the *students-proposing deferred-acceptance (DA) mechanism* or the *top trading cycles (TTC) mechanism* for real-life school choice problems. Both DA and TTC are *strategy-proof*: for each student, it is a weakly dominant strategy to report his preferences truthfully (and any (non-)sophisticated student's best interest is to report his true preferences). DA is stable but inefficient whereas TTC is efficient but unstable. However, DA is "constrained efficient" as it chooses the students-optimal stable assignment. The TTC mechanism is based on Gale's TTC algorithm (Shapley and Scarf, 1974) and allows students to trade their priorities among themselves starting with the students with highest priorities. A student and a school may be involved in a blocking pair at a TTC assignment simply because a lower priority student can be assigned a seat at that school by trading his high priority at another school. When considering first stability and second efficiency, DA was suggested, and considering first efficiency and second stability, TTC was suggested (Abdulkadiroğlu and Sönmez, 2003). Although it is easy to see that TTC fails *stability*, intuitively it respects priorities "to some extent" and should be, in some sense, "minimally unstable".

The intuition for TTC has been only recently formalized by Abdulkadiroğlu et al. (2019). They propose to compare assignments in terms of their stability by comparing the sets of blocking pairs at these assignments. An assignment is *more stable* than another assignment if the set of blocking pairs in the former assignment is a subset of the set of blocking pairs in the latter assignment. Using this stability comparison method, they show the following result:

TTC is minimally unstable among efficient and strategy-proof mechanisms in one-to-one matching (when each school has unit capacity), that is, for any other *efficient* and *strategy-proof* mechanism, there exists an instance where the assignment chosen by this mechanism does not produce a subset of blocking pairs of the TTC-assignment.

Although it does not extend to the many-to-one setup,[1] the result of Abdulkadiroğlu et al. (2019) is an important justification for using TTC in priority-based resource allocation especially because their stability comparison method, which relies on comparing the sets of blocking pairs (in the set inclusion sense), is very reasonable. However, there are other plausible ways to compare assignments in terms of stability. A natural alternative is to count the number of blocking pairs, which induces a complete comparison method (as all assignments can be compared by counting blocking pairs) and it is not immediate from Abdulkadiroğlu et al. (2019) whether TTC is minimally unstable among efficient and strategy-proof mechanisms also based on this alternative cardinal comparison method. One may also consider comparison methods that are not based on the set of blocking pairs, but based on alternative sets such as (i) the set of blocking triplets[2] as in Kwon and Shorrer (2019)[3] or (ii) the set of blocking students[4] as in Doğan and Ehlers (2020).[5] Possibly, TTC is minimally unstable among efficient and strategyproof mechanisms for certain comparison methods but not for others.

We show that the result of Abdulkadiroğlu et al. (2019) is robust to the choice of stability comparison method. We formulate stability comparisons as binary relations (on the set of possible assignments) that depend on the preference profile. We introduce three basic properties that any reasonable stability comparison should satisfy. The first property *stability preferred* requires that any stable assignment is strictly more stable than any unstable assignment. The second property *separability* requires that if an assignment is more stable than another assignment, and at the same time some subset of students or their assigned schools are not involved in any blocking pair at the first assignment and the second assignment

[1] For the many-to-one setup, Abdulkadiroğlu et al. (2019) show that TTC outperforms serial dictatorship, an obvious efficient alternative, by admitting fewer blocking pairs in an average sense when every possible priority profile is considered or when participants' priorities are drawn uniform randomly.

[2] A blocking triplet includes, in addition to a blocking pair, a student who violates the priority of the student in the blocking pair.

[3] Kwon and Shorrer (2019) show that TTC mechanism is minimally unstable among efficient and strategy-proof mechanisms in one-to-one matching when stability comparison is based on comparing (in the set-inclusion sense) sets of blocking triplets.

[4] A blocking student is a student who is involved in at least one blocking pair.

[5] In Doğan and Ehlers (2020), we show that there are school choice problems where any Pareto improvement over the deferred acceptance assignment is not minimally unstable among efficient assignments when stability comparison is *cardinal* and compares the *number* of blocking pairs, or when the stability comparison method is based on comparing (in the set-inclusion or cardinal sense) sets of blocking students.

assigns these students to the same schools while some of those students are involved in a blocking pair, then the restriction of the first assignment to the remaining students and schools is strictly more stable than the restriction of the second assignment. The third property *consistency* requires that if two assignments coincide for some subset of students and these students and their assigned schools are involved in blocking pairs only among themselves, then the stability comparison remains identical when considering the assignments restricted to the remaining students and schools. All of the above comparison methods satisfy all three properties. We show that, given any stability comparison satisfying *stability preferred*, *separability* and *consistency*, *TTC* is minimally unstable among efficient and strategy-proof mechanisms when each school has unit capacity. Our main proof arguments are considerably different than the corresponding ones of Abdulkadiroğlu et al. (2019) or for characterizations of TTC by Ma (1994) and Svensson (1999). Loosely speaking, we show that if our theorem is not true, then we can always find a smaller problem (with fewer students) with a contradiction. In particular, our approach handles *cardinal* comparison methods.

The paper is organized as follows. Section 2 introduces school choice problems and mechanisms. Section 3 defines stability comparison methods and basic properties which any reasonable stability comparison method shall satisfy. Section 4 defines TTC and states our main result. Section 5 applies our main result to several natural comparison methods.

2 The Model

Let $\mathcal{N}$ denote an infinite set of potential students and $\mathcal{C}$ denote an infinite set of potential schools.[6] To specify a (school choice) problem, we first draw a finite set of students $N \subset \mathcal{N}$ and a finite set of schools $C \subset \mathcal{C}$.

A problem for (N, C) includes a preference profile $R = (R_i)_{i \in N}$, a capacity profile $q = (q_c)_{c \in C}$, and a priority profile $\succeq = (\succeq_c)_{c \in C}$. For each student $i \in N$, R_i denotes his **preference ordering** over $C \cup \{\emptyset\}$,[7] where $\emptyset$ represents an outside option for the student. The strict part of the preference ordering R_i is denoted by P_i, so if $c_1, c_2 \in C \cup \{\emptyset\}$, $c_1 \neq c_2$, and $c_1 \; R_i \; c_2$, then $c_1 \; P_i \; c_2$. School c is **acceptable** to student i if the student prefers it to the outside option, that is, $c \; P_i \; \emptyset$. For each school $c \in C$, $q_c \in \mathbb{N}$ denotes its **capacity**, which is

[6] We use the "potential students (schools)" terminology since we will be referring to sub-problems (problems restricted to a set of students and schools given an original problem).

[7] Formally, a preference ordering over $C \cup \{\emptyset\}$ is a complete, transitive, and anti-symmetric binary relation over $C \cup \{\emptyset\}$. Binary relation R_i over $C \cup \{\emptyset\}$ is *complete* if, for every $c_1, c_2 \in C \cup \{\emptyset\}$, $c_1 R_i c_2$ or $c_2 R_i c_1$. It is *transitive* if, for every $c_1, c_2, c_3 \in C \cup \{\emptyset\}$, $c_1 R_i c_2$ and $c_2 R_i c_3$ imply $c_1 R_i c_3$. It is *anti-symmetric* if, for every $c_1, c_2 \in C \cup \{\emptyset\}$, $c_1 R_i c_2$ and $c_2 R_i c_1$ imply $c_1 = c_2$.

the maximum number of students that the school can admit, and $\succeq_c$ is a **priority ordering** over the set of students N.[8] The strict part of the priority ordering $\succeq_c$ is denoted by $\succ_c$. We call the quintuple $(N, C, R, q, \succeq)$ a **problem**. Let $\mathcal{P}^{(N,C)}$ denote the set of all problems for (N, C), and let $\mathcal{P}$ denote the set of all problems for any finite sets of students and schools.

Given a problem $P = (N, C, R, q, \succeq) \in \mathcal{P}$, a set of students $N' \subseteq N$ and schools $C' \subseteq C$, we call $P|_{(N\backslash N', C\backslash C')}$ as the **restriction** of P to $(N\backslash N', C\backslash C')$, where $P|_{(N\backslash N', C\backslash C')}$ is obtained from P by simply removing N' and C', and also removing them from q, R, and $\succeq$ while keeping relative orderings of the remaining students and the relative orderings and capacities of the remaining schools the same.

Given a problem $P = (N, C, R, q, \succeq) \in \mathcal{P}$, an assignment is a mapping $\mu : N \cup C \rightarrow N \cup C \cup \{\emptyset\}$ such that

(i) for each $i \in N$, $\mu(i) \in C \cup \{\emptyset\}$,

(ii) for each $c \in C$, $\mu(c) \subseteq N$ such that $|\mu(c)| \leq q_c$, and

(iii) for each $i \in N$ and each $c \in C$, $i \in \mu(c)$ if and only if $c = \mu(i)$.

Let $\mathcal{A}(P)$ denote the set of all possible assignments at the problem P. Note that, essentially, $\mathcal{A}(P)$ is determined by (N, C, q).

For any set of students $N' \subseteq N$, we denote the **aggregate assignment** of N' at μ by $\mu(N') = \{c \in C | \exists i \in N' : \mu(i) = c\}$.[9] We say that a set of students is isolated at μ if no other student is assigned to a school in their aggregate assignment, i.e., N' is **isolated** at μ if there is no $i \in N \setminus N'$ and $c \in \mu(N')$ such that $\mu(i) = c$. Note that if N' is isolated at μ, then $N \setminus N'$ is isolated as well. Note also that if μ is a one-to-one assignment (i.e., if no school is assigned more than one student), then any set of students is isolated at μ.

Given an isolated set of students N' for μ with $\mu(N') = C'$, we denote by $\mu|_{N\backslash N'}$ the **restriction** of μ to $N\backslash N'$ and $C\backslash C'$, where $\mu|_{N\backslash N'}$ is obtained from μ by simply removing N' and C' while keeping the assignments of $N\backslash N'$ the same as in μ. Note that $\mu|_{N\backslash N'} \in \mathcal{A}(P|_{(N\backslash N', C\backslash C')})$ (as N' is isolated in μ and $\mu(N') = C'$).

An assignment μ is **individually rational** if for each $i \in N$, $\mu(i)\ R_i\ \emptyset$. An assignment μ **Pareto dominates** another assignment μ' if for each $i \in N$, $\mu(i)\ R_i\ \mu'(i)$ and there exists $i \in N$ such that $\mu(i)\ P_i\ \mu'(i)$. An assignment μ is **efficient** if it is not Pareto dominated.[10]

[8] The priority ordering $\succeq_c$ is a complete, transitive, and anti-symmetric binary relation over N. Our results extend to the more general setup where some students may be unacceptable for some schools.

[9] Note that $\mu(N') = \emptyset$ if and only if all students in N' are assigned their outside options.

[10] Note that efficiency implies individual rationality.

A pair $(i,c) \in N \times C$ **blocks** μ if $c\ P_i\ \mu(i)$ and $[|\mu(c)| < q_c$ or there exists $j \in \mu(c)$ such that $i \succ_c j]$. Let

$$B(\mu) = \{(i,c) \in N \times C : (i,c) \text{ blocks } \mu\}$$

denote the set of blocking pairs at μ. In addition, for each $i \in N$, let $B_i(\mu) = \{c \in C : (i,c) \in B(\mu)\}$ denote the set of schools together with which student i constitute a blocking pair, and for each $c \in C$, $B_c(\mu) = \{i \in N : (i,c) \in B(\mu)\}$ denote the set of students together with whom school c constitute a blocking pair.

An assignment μ is **stable** if it is individually rational and includes no blocking pair. Unfortunately, there exist school choice problems without an assignment that is both *efficient* and *stable* (Roth, 1982).

Given $N' \subseteq N$, we say that μ is **N'-stable** if no student in N' is involved in a blocking pair and also no school in $\mu(N')$ is involved in a blocking pair, i.e., for each $i \in N'$, $B_i(\mu) = \emptyset$ and for each $c \in \mu(N')$, $B_c(\mu) = \emptyset$. We say that μ is N'-unstable if μ is not N'-stable.

A **mechanism** associates each problem with an assignment. When we say that a mechanism satisfies a certain assignment property, such as *efficiency*, we mean that at each problem, the assignment prescribed by the mechanism satisfies the property.

A mechanism φ is **strategy-proof** if reporting true preferences is a weakly dominant strategy for each student in the preference revelation game induced by φ, that is, for each problem $(N, C, R, q, \succeq)$, each $i \in N$ and each preference ordering R'_i,

$$\varphi_i(N, C, R, q, \succeq)\ R_i\ \varphi_i(N, C, (R'_i, R_{-i}), q, \succeq).$$

When $(N, C, q, \succeq)$ is clear, we often denote a problem simply by its preference profile R. Now, using our convention, the above simply says $\varphi_i(R)\ R_i\ \varphi_i(R'_i, R_{-i})$.

3 A Unifying Approach to Stability Comparisons

Given a problem $P \in \mathcal{P}$, a binary relation over assignments is a subset $\gtrsim \subseteq \mathcal{A}(P) \times \mathcal{A}(P)$. We use the convention and write $\mu \gtrsim \nu$ instead of $(\mu, \nu) \in \gtrsim$, and $[\mu \gnsim \nu \Leftrightarrow \mu \gtrsim \nu \&$ not $\nu \gtrsim \mu]$. Let $\mathcal{L}(P)$ denote the set of all binary relations at P. Given $\gtrsim \in \mathcal{L}(P)$, (i) $\gtrsim$ is complete if for all $\mu, \nu \in \mathcal{A}(P)$ we have $\mu \gtrsim \nu$ or $\nu \gtrsim \mu$ and (ii) $\gtrsim$ is transitive if $\mu \gtrsim \nu$ and $\nu \gtrsim \eta$ imply $\mu \gtrsim \eta$. Furthermore, given $\gtrsim, \gtrsim' \in \mathcal{L}(P)$ such that $\gtrsim \subseteq \gtrsim'$ we say that $\gtrsim$ is coarser than $\gtrsim'$ and $\gtrsim'$ is finer than $\gtrsim$.

A **stability comparison** is a function f associating with each problem $P \in \mathcal{P}$ a binary relation $f(P) \in \mathcal{L}(P)$. Instead of $f(P)$, we write $\gtrsim_f^P$ (where $\mu \gtrsim_f^P \nu$ means that μ is f-more stable than ν at P). Note that, at this point, we do not impose any structure on a stability comparison (such as neither completeness nor transitivity). Later we will describe several examples of stability comparison methods. Also note that, when (N, C, q) is fixed, while the set of assignments does not vary with the preference or the priority profile, the stability comparison may vary with the preference and the priority profile, that is, stability comparisons depend on the preference and the priority profile.

We introduce the following basic properties for a stability comparison f.

The first property *stability preferred* requires that any stable assignment is strictly f-more stable than any unstable assignment. Formally, f satisfies **stability-preferred** if for each $P \in \mathcal{P}$ and $\mu, \nu \in \mathcal{A}(P)$, if $B(\mu) = \emptyset \neq B(\nu)$, then $\mu \gnsim_f^P \nu$.

The second property *separability* requires the following. Suppose that an assignment is f-more stable than another assignment, and a set of students is isolated and have the same aggregate assignment at both assignments. If this isolated set of students or their assigned schools are not involved in any blocking pair at the first assignment while some of these students are involved in a blocking pair at the second assignment, then the restriction of the first assignment to the remaining students and schools is strictly f-more stable than the restriction of the second assignment.

Formally, f satisfies **separability** if for each $P \in \mathcal{P}$ and $\mu, \nu \in \mathcal{A}(P)$ such that $\nu \gtrsim_f^P \mu$, if μ is N'-stable for some nonempty isolated $N' \subseteq N$, N' is also isolated at ν and $\mu(N') = \nu(N') = C'$, and $B_i(\nu) \neq \emptyset$ for some $i \in N'$, then $\nu|_{N \setminus N'} \gnsim_f^{P'} \mu|_{N \setminus N'}$, where $P' = P|_{(N \setminus N', C \setminus C')}$.

The third property *consistency* requires the following. Suppose that two assignments coincide for some subset of students which is isolated at both assignments. If these students and their assigned schools are involved in blocking pairs only among themselves, then the f-stability comparison remains unchanged when considering the assignments restricted to the remaining students and schools.

Formally, f satisfies **consistency** if for each $P \in \mathcal{P}$ and $\mu, \nu \in \mathcal{A}(P)$ such that $\nu \gtrsim_f^P \mu$, if for some $\emptyset \neq N' \subseteq N$ that is isolated at both μ and ν, $\nu(i) = \mu(i)$ for all $i \in N'$, $B_i(\mu) = B_i(\nu) \subseteq \mu(N') = \nu(N') = C'$ for all $i \in N'$, and $B_c(\mu) = B_c(\nu) \subseteq N'$ for all $c \in \mu(N')$, then $\nu|_{N \setminus N'} \gtrsim_f^{P'} \mu|_{N \setminus N'}$, where $P' = P|_{(N \setminus N', C \setminus C')}$.

Now given a stability comparison f and problem P, we say that μ is **f-minimally unstable at P among efficient assignments** if there exists no efficient assignment ν such that $\nu \gnsim_f^P \mu$.

Given a mechanism ψ, we say that ψ is **f-minimally unstable among efficient and strategyproof mechanisms** if for any efficient and strategyproof mechanism φ, $[\varphi(P) \gtrsim_f^P \psi(P)$ for all $P \in \mathcal{P}]$ implies $\varphi = \psi$.

Given a mechanism ψ, we say that ϕ is **weakly f-minimally unstable among efficient and strategyproof mechanisms** if there does not exist any efficient and strategyproof mechanism φ with (i) $\varphi(P) \gtrsim_f^P \psi(P)$ for all $P \in \mathcal{P}$ and (ii) $\varphi(P') \gnsim_f^{P'} \psi(P')$ for some $P' \in \mathcal{P}$.

Note that the second definition is slightly different than the first one. Furthermore, if ψ is f-minimally unstable among efficient and strategyproof mechanisms, then ψ is weakly f-minimally unstable among efficiency and strategyproof mechanisms (but the converse does not hold as there might exist a mechanism different from ψ with the identical f-stability measure). Abdulkadiroğlu et al. (2019) and Kwon and Shorrer (2019) use the weaker second definition and impose in addition comparability (among mechanisms) for their partial relations.

4 Robust Minimal Instability of Top Trading Cycles

The **top trading cycles (TTC) mechanism** (Abdulkadiroğlu and Sönmez, 2003) is based on Gale's TTC algorithm (Shapley and Scarf, 1974) which runs, given a problem, as follows.

Top Trading Cycles (TTC) Algorithm:[11]

Step 1. Assign a counter for each school which keeps track of how many seats are still available at the school. Initially set the counters equal to the capacities of the schools. Each student points to her top-ranked school. Each school points to the student who has the highest priority for the school. Since the number of students and schools are finite, there is at least one cycle. (A cycle is an ordered list of distinct students and distinct schools $(k, c_k)_{k\in\{1,\ldots,K\}}$ such that for each $k \in \{1, \ldots, K\}$, student k points to school c_k and school c_k points to student $k + 1$ with the convention that $K + 1 = 1$. Moreover, each school can be part of at most one cycle. Similarly, each student can be part of at most one cycle. Every student in a cycle is assigned a seat at the school she points to and is removed. The counter of each school in a cycle is reduced by one and if it reduces to zero, the school is also removed. Counters of all other schools stay put.

Step $t \geq 2$. Each remaining student points to her top-ranked school among the remaining schools and each remaining school points to the student with highest

[11] Morrill (2015a), Morrill (2015b), and Hakimov and Kesten (2018) propose variants of TTC for the many-o-one setup. For one-to-one problems, all variants coincide.

priority among the remaining students. There is at least one cycle. Every student in a cycle is assigned a seat at the school that she points to and is removed. The counter of each school in a cycle is reduced by one and if it reduces to zero the school is also removed. Counters of all other schools stay put.

Our main result is the following.

Theorem 1 *Let f be a stability comparison satisfying stability preferred, separability and consistency. Then TTC is f-minimally unstable among efficient and strategy-proof mechanisms when each school has unit capacity.*

The proof of Theorem 1 builds on the observation that if TTC is not f-minimally unstable among *efficient* and *strategy-proof* mechanisms, then there must be a *smallest* number of students, say n, such that there exists a mechanism φ that is defined on the domain of problems including n students and *strategy-proof*, *efficient*, and *f-more stable* than *TTC*. The heart of the proof contains two key steps: first, if some students who are assigned seats in Step 1 of the TTC algorithm receive different schools at φ, then we construct a new domain of problems including fewer students where TTC is not f-minimally unstable among *efficient* and *strategy-proof* mechanisms; and otherwise, we construct a smaller problem (using our basic properties of the stability comparison method) where the restricted TTC assignment is more stable than the restricted φ assignment.

Proof. Suppose not. Let n be the smallest number of students such that there exists a mechanism defined on the domain of problems with n students and a set of schools C that is *strategy-proof*, *efficient*, and *f-more stable* than TTC. Note that $n \geq 3$ since at any problem including 1 or 2 students, the TTC assignment includes no blocking pair and f satisfies stability preferred.

Suppose that φ is *strategy-proof*, *efficient*, and *f-more stable than TTC* on the domain of problems with n students and a set of schools C.

Unless otherwise noted, all parameters of a problem except for the preference profile will be fixed and we will denote a problem simply by its preference profile.

Lemma 1 *Let P be an arbitrary problem including n students with the preference profile $R = (R_1, \ldots, R_n)$. Let $i \in N$ be a student who is assigned a seat at Step 1 of $TTC(R)$. Let $c \in C$ be the school that points to i at Step 1 of $TTC(R)$. Let R'_i be a preference relation for student i at which c is the only acceptable school. Then, $\varphi_i(R'_i, R_{-i}) = c$.*

Proof. Suppose not, i.e., suppose that $\varphi_i(R'_i, R_{-i}) \neq c$ (note that $\varphi_i(R'_i, R_{-i}) = \emptyset$). By *efficiency*, there exists $j_1 \neq i$ such that $\varphi_{j_1}(R'_i, R_{-i}) = c$. Let R'_{j_1} be a preference relation for student j_1 at which c is the only acceptable school. By *strategy-proofness*, $\varphi_{j_1}(R'_i, R'_{j_1}, R_{-\{i,j_1\}}) = c$.

Now, suppose that there exists a preference profile $\overline{R}_{-\{i,j_1\}}$ of students $N \setminus \{i, j_1\}$ such that $\varphi_c(R'_i, R'_{j_1}, \overline{R}_{-\{i,j_1\}}) \in N \setminus \{i, j_1\}$, i.e., c is assigned to a student different from i or j_1. Let $j_2 \in N \setminus \{i, j_1\}$ such that $\varphi_{j_2}(R'_i, R'_{j_1}, \overline{R}_{-\{i,j_1\}}) = c$. Let R'_{j_2} be a preference relation for student j_2 at which c is the only acceptable school. By *strategy-proofness*, $\varphi_{j_2}(R'_i, R'_{j_1}, R'_{j_2}, \overline{R}_{-\{i,j_1,j_2\}}) = c$.

Successive applications of the above argument imply that there exist $\{j_1, \ldots, j_m\}$ and a preference profile $R^*_{-\{i,j_1,j_2,\ldots,j_m\}}$ for students $N \setminus \{i, j_1, j_2, \ldots, j_m\}$ such that

- for each $t \in \{1, \ldots, m\}$, R'_{j_t} is a preference relation for student j_t at which c is the only acceptable school,
- $\varphi_{j_m}(R'_i, R'_{j_1}, R'_{j_2}, \ldots, R'_{j_m}, R^*_{-\{i,j_1,j_2,\ldots,j_m\}}) = c$, and
- for any preference profile $R^{**}_{-\{i,j_1,j_2,\ldots,j_m\}}$ for students $N \setminus \{i, j_1, j_2, \ldots, j_m\}$, we have $\varphi_c(R'_i, R'_{j_1}, R'_{j_2}, \ldots, R'_{j_m}, R^{**}_{-\{i,j_1,j_2,\ldots,j_m\}}) \in \{i, j_1, \ldots, j_m\}$.

Let $P' = (R'_i, R'_{j_1}, R'_{j_2}, \ldots, R'_{j_m}, R^*_{-\{i,j_1,j_2,\ldots,j_m\}})$. Note that if $m = n$, then $|B(TTC(P'))| = 0 < |B(\varphi(P'))|$ since i has the highest priority among all students at c. Moreover, this contradicts that φ is *f-more stable* than TTC as f satisfies stability preferred and $TTC(P') \gtrsim_f^{P'} \varphi(P')$. Thus, $m < n$.

Now, we will construct a mechanism φ' defined on the domain of problems with students $N' = N \setminus \{i, j_1, j_2, \ldots, j_m\}$ and schools $C' = C \setminus \{c\}$ that is *strategy-proof*, *efficient*, and *f-more stable* than TTC, which will contradict that n is the smallest number of students such a domain entails.

Let φ' be defined as follows. For each preference profile $\overline{R}_{N'}$ of N',

- If for each $j \in N'$, $\overline{R}_j$ agrees with R^*_j on the relative orderings of C', then $\varphi'(\overline{R}_{N'}) = \varphi(R'_i, R'_{j_1}, R'_{j_2}, \ldots, R'_{j_m}, R^*_{N'})|_{N'}$
- If for each $j \in N'' \subsetneq N'$, $\overline{R}_j$ agrees with R^*_j on the relative orderings of C', and for each $j' \in N' \setminus N''$, $\overline{R}_j$ does not agree with $R^*_{j'}$ on the relative orderings of C', then let $\varphi'(\overline{R}_{N'}) = \varphi(R'_i, R'_{j_1}, R'_{j_2}, \ldots, R'_{j_m}, R'_{N'})|_{N'}$ where for each $j \in N''$, $R'_j = R^*_j$, and for

each $j \in N' \setminus N''$, R'_j is a preference ordering which bottom-ranks c and agrees with $\overline{R}_{j'}$ on the relative orderings of C'.

Note that φ' is well-defined, in particular when $\{i, j_1, \ldots, j_m\}$ report $(R'_i, R'_{j_1}, R'_{j_2}, \ldots, R'_{j_m})$, no student in N' can receive school c under any preference profile of N' (and $\{i, j_1, \ldots, j_m\}$ is always isolated). To see that φ' is *strategy-proof*, observe that manipulability of φ' would immediately imply the manipulability of φ. *Efficiency* of φ' also follows directly from the *efficiency* of φ. We will next show that φ' is *f-more stable* than TTC.

Note that at any problem R (in the domain where there are n students) such that $(i, j_1, \ldots, j_m)$ report $(R'_i, R'_{j_1}, R'_{j_2}, \ldots, R'_{j_m})$, no student in $(i, j_1, \ldots, j_m)$ is involved in a blocking pair at the TTC assignment; moreover, no student in N' is included in a blocking pair together with c at the TTC assignment. Thus, $TTC(R)$ is $N \backslash N'$-stable. On the other hand, consider the problem $P' = (R'_i, R'_{j_1}, R'_{j_2}, \ldots, R'_{j_m}, R^*_{-\{i,j_1,j_2,\ldots,j_m\}})$. Note that (i, c) is a blocking pair at $\varphi(P')$. Since φ is *f-more stable* than TTC, we have $\varphi(P') \gtrsim_f^{P'} TTC(P')$. Furthermore, by efficiency of φ and construction, we have $\cup_{h \in N \backslash N'} \{\varphi_h(P')\} = \{c\} = \cup_{h \in N \backslash N'} \{TTC_h(P')\}$ and $N \backslash N'$ is isolated under both $\varphi(P')$ and $TTC(P')$. Consequently, by separability of f, at the problem $\overline{R}_{N'}$ (in the domain where there are $n - m$ students) where for each $j \in N'$, $\overline{R}_j$ agrees with R^*_j on the relative orderings of C', $\varphi'(\overline{R}_{N'}) = \varphi(P')|_{N'} \gnsim_f^{\overline{R}_{N'}} TTC(P')|_{N'} = TTC(\overline{R}_{N'})$ (where the equalities follow from the definition of φ' and TTC).

Now consider any problem R (in the domain where there are n students) such that $(i, j_1, \ldots, j_m)$ report $(R'_i, R'_{j_1}, R'_{j_2}, \ldots, R'_{j_m})$. Then for R, $TTC(R)$ is $N \backslash N'$-stable. If for some $i \in N \backslash N'$, $\varphi_i(R) \neq \emptyset$, then by efficiency $\varphi_i(R) = c$ and we use the same arguments as above (as $N \backslash N'$ is isolated under both $TTC(R)$ and $\varphi(R)$). If $TTC_i(R) = \varphi_i(R)$ for all $i \in N \backslash N'$, then by construction, $\varphi(R')$ is $N \backslash N'$-stable and $N \backslash N'$ is isolated under both $TTC(R)$ and $\varphi(R)$. Hence, by consistency of f and $\varphi(R) \gtrsim_f^R TTC(R)$, we obtain $\varphi(R)|_{N'} \gtrsim_f^{R_{N'}} TTC(R)|_{N'}$. Thus (as R was arbitrary), for any profile $\overline{R}_{N'}$ of N' we have $\varphi'(\overline{R}_{N'}) \gtrsim_f^{R_{N'}} TTC(\overline{R}_{N'})$ (from the definition of φ' and TTC). Hence, φ' is f-more stable than TTC, contradicting that n is the smallest number of students such a domain entails. ■

Lemma 2 *Let P be an arbitrary problem including n students with the preference profile $R = (R_1, \ldots, R_n)$. Let $i \in N$ be a student who is assigned a seat at Step 1 of $TTC(R)$. Then, $\varphi_i(R) = TTC_i(R)$.*

Proof. Let I_1 denote the set of students who are assigned a seat at Step 1 of $TTC(R)$ and C_1 denote the set of schools that are allocated at Step 1 of $TTC(R)$. Note that if for each $i \in I_1$, $\varphi_i(R) \in C_1$, then by *efficiency*, $\varphi_i(R) = TTC_i(R)$ for each $i \in I_1$.

Suppose that there exists $i_1 \in I_1$ such that $\varphi_{i_1}(R) \notin C_1$. Let $c_1 \in C_1$ be the school that points to i_1 in Step 1 of $TTC(R)$. Let R'_{i_1} be a preference ordering for i_1 at which $TTC_{i_1}(R)$ is top-ranked and c_1 is second-ranked.[12] By strategy-proofness, $\varphi_{i_1}(R'_{i_1}, R_{-i_1}) \neq TTC_{i_1}(R)$. By Lemma 1 and *strategy-proofness*, $\varphi_{i_1}(R'_{i_1}, R_{-i_1}) = c_1$.

Note that I_1 is still the set of students who are assigned a seat at Step 1 of $TTC(R'_{i_1}, R_{-i_1})$ and C_1 is still the set of schools that are allocated at Step 1 of $TTC(R'_{i_1}, R_{-i_1})$. Now, if for each $i \in I_1 \setminus \{i_1\}$, $\varphi_i(R) \in C_1$, then *efficiency* would imply that $\varphi_i(R'_{i_1}, R_{-i_1}) = TTC_i(R'_{i_1}, R_{-i_1})$ for each $i \in I_1$, which would contradict $\varphi_{i_1}(R'_{i_1}, R_{-i_1}) \neq TTC_{i_1}(R) = TTC_{i_1}(R'_{i_1}, R_{-i_1})$. Thus, there exists $i_2 \in I_1 \setminus \{i_1\}$ such that $\varphi_{i_2}(R) \notin C_1$. Let $c_2 \in C_1$ be the school that points to i_2 in Step 1 of $TTC(R)$. Let R'_{i_2} be a preference ordering for i_2 at which $TTC_{i_2}(R)$ is top-ranked and c_2 is second-ranked. By strategy-proofness, $\varphi_{i_2}(R'_{i_1}, R'_{i_2}, R_{-\{i_1,i_2\}}) \neq TTC_{i_2}(R)$. By Lemma 1 and *strategy-proofness*, $\varphi_{i_2}(R'_{i_1}, R'_{i_2}, R_{-\{i_1,i_2\}}) = c_2$.

Continuing in a similar fashion, we identify a list of students $(i_1, \dots, i_m)$ and a preference profile $R' = (R'_1, \dots, R'_m)$ such that $\{i_1, \dots, i_m\} = I_1$, $\varphi_i(R') \in C_1$ for each $i \in I_1$, and $\varphi_{i_m}(R') \neq TTC_{i_m}(R')$, which contradicts *efficiency* of φ. ■

Lemma 3 *Let k be a number. Suppose that at any problem $\overline{P}$ including n students and a preference profile $\overline{R}$, if a student i is assigned a seat at an earlier step than Step k at $TTC(\overline{R})$, then $\varphi_i(\overline{R}) = TTC_i(\overline{R})$. Let P be an arbitrary problem including n students with the preference profile $R = (R_1, \dots, R_n)$. Let $i \in N$ be a student who is assigned a seat at Step k of $TTC(R)$. Let $c \in C$ be the school that points to i at Step k of $TTC(R)$. Let R'_i be a preference relation for student i at which c is the only acceptable school. Then, $\varphi_i(R'_i, R_{-i}) = c$.*

Proof. The proof shares some arguments with the proof of Lemma 1. We repeat these arguments for the sake of completeness.

Suppose not, i.e., suppose that $\varphi_i(R'_i, R_{-i}) \neq c$ (note that $\varphi_i(R'_i, R_{-i}) = \emptyset$). By *efficiency*, there exists $j_1 \neq i$ such that $\varphi_{j_1}(R'_i, R_{-i}) = c$. Let $I_{<k}$ denote the set of students who are assigned seats at an earlier step than Step k at $TTC(R)$. Note that any student $j \in I_{<k}$ is still assigned the same seat at an earlier step than Step k at $TTC(R'_i, R_{-i})$. Then, by our supposition, for any student $j \in I_{<k}$, $\varphi_j(R) = TTC_j(R)$. But then, $j_1 \notin I_{<k}$. Hence, by the definition of TTC, i has higher priority than j_1 at c since c points to i at Step k of $TTC(R)$. Let R'_{j_1} be a preference relation for student j_1 at which c is the only acceptable school. By *strategy-proofness*, $\varphi_{j_1}(R'_i, R'_{j_1}, R_{-\{i,j_1\}}) = c$.

[12] We will use the convention and write sometimes for short $R'_{i_1} : TTC_{i_1}(R)c_1$.

Now, suppose that there exists a preference profile $\overline{R}_{-\{i,j_1,I_{<k}\}}$ of students $N\setminus(\{i,j_1\}\cup I_{<k})$ such that $\varphi_c(R'_i, R'_{j_1}, R_{I_{<k}}, \overline{R}_{-\{i,j_1,I_{<k}\}}) \in N\setminus(\{i,j_1\}\cup I_{<k})$. Let $j_2 \in N\setminus(\{i,j_1\}\cup I_{<k})$ be such that $\varphi_{j_2}(R'_i, R'_{j_1}, R_{I_{<k}}, \overline{R}_{-\{i,j_1,I_{<k}\}}) = c$. Let R'_{j_2} be a preference relation for student j_2 at which c is the only acceptable school. By *strategy-proofness*, $\varphi_{j_2}(R'_i, R'_{j_1}, R'_{j_2}, R_{I_{<k}}, \overline{R}_{-\{i,j_1,j_2,I_{<k}\}}) = c$.

Successive applications of the above argument imply that there exist $\{j_1, \ldots, j_m\}$ and a preference profile $R^*_{-\{i,j_1,j_2,\ldots,j_m,I_{<k}\}}$ for students $N \setminus (\{i, j_1, j_2, \ldots, j_m\} \cup I_{<k})$ such that

- for each $t \in \{1, \ldots, m\}$, R'_{j_t} is a preference relation for student j_t at which c is the only acceptable school,
- $\varphi_{j_m}(R'_i, R'_{j_1}, R'_{j_2}, \ldots, R'_{j_m}, R_{I_{<k}}, R^*_{-\{i,j_1,j_2,\ldots,j_m,I_{<k}\}}) = c$, and
- for any preference profile $R^{**}_{-\{i,j_1,j_2,\ldots,j_m,I_{<k}\}}$ for students $N\setminus(\{i, j_1, j_2, \ldots, j_m\}\cup I_{<k})$, we have $\varphi_c(R'_i, R'_{j_1}, R'_{j_2}, \ldots, R'_{j_m}, R_{I_{<k}}, R^{**}_{-\{i,j_1,j_2,\ldots,j_m,I_{<k}\}}) \notin N \setminus (\{i, j_1, j_2, \ldots, j_m\} \cup I_{<k})$.

Let $P' = (R'_i, R'_{j_1}, R'_{j_2}, \ldots, R'_{j_m}, R_{I_{<k}}, R^*_{-\{i,j_1,j_2,\ldots,j_m,I_{<k}\}})$. First note that, each $j \in I_{<k}$ is still assigned the same seat as in $TTC(R)$ at an earlier step than Step k at $TTC(P')$. Hence, by our supposition, for any $j \in I_{<k}$, $\varphi_j(P') = TTC_j(P') = TTC_j(R)$. For later purposes, let $c = c^1$ and $J_{c^1} = \{i, j_1, j_2, \ldots, j_m\}$.

Now, note that if $m = n - |I_{<k}|$, then $N = J_{c^1} \cup I_{<k}$ and $TTC_i(P') = \varphi_i(P')$ for all $i \in I_{<k}$. Hence, for all $i \in I_{<k}$ we have $B_i(TTC(P')) = B_i(\varphi(P'))$ and for all $c \in \cup_{h\in I_{<k}} TTC_h(P')$, $B_c(TTC(P')) = B_c(\varphi(P')) \subseteq I_{<k}$. Thus, $I_{<k}$ is isolated under both $\varphi(P')$ and $TTC(P')$. But then $\varphi(P') \succsim_f^{P'} TTC(P')$ and consistency of f imply $\varphi(P')|_{J_{c^1}} \succsim_f^{P'_{J_{c^1}}} TTC(P')|_{J_{c^1}}$. But this is a contradiction to stability preferred of f as under $P'_{J_{c^1}}$ we have $B(TTC(P')|_{J_{c^1}}) = \emptyset \neq B(\varphi(P')|_{J_{c^1}})$ since i has the highest priority among students J_{c^1} at c^1. This contradicts that φ is *f-more stable* than TTC. Thus, $m < n - |I_{<k}|$.

Next we show that for all $i \in I_k\setminus J_{c^1}$ we have $TTC_i(P') = \varphi_i(P')$. If $\cup_{i\in I_k} TTC_i(P') = \cup_{i\in I_k} \varphi_i(P')$, then this follows from efficiency of $\varphi(P')$ and $TTC(P')$. Thus, for some $i \in I_k$, $\varphi_i(P') \notin \cup_{h\in I_k}\{TTC_h(P')\}$. Thus, by construction of J_{c^1} and the induction hypothesis, $\varphi_i(P') \notin \{c\} \cup [\cup_{h\in I_{\leq k}}\{TTC_h(P')\}]$ (where $I_{\leq k} = I_{<k} \cup I_k$). Let $i = h_l$ belong in $TTC(P')$ to a cycle $c_1 \to h_1 \to \cdots \to c_l \to h_l \to c_{l+1} \to h_{l+1} \to \cdots \to c_1$ but $\varphi_{h_l}(R') \neq c_{l+1}$, i.e. $TTC_{h_l}(P') = c_{l+1}$, $TTC_{h_{l-1}}(P') = c_l$ and c_l points to h_l in the TTC-algorithm. Let $\hat{P}_{h_l} : c_{l+1}c_l$ and $\hat{P} = (\hat{P}_{h_l}, P'_{-h_l})$. By strategyproofness and efficiency, $\varphi_{h_l}(\hat{P}) = \emptyset$ or $\varphi_{h_l}(\hat{P}) = c_l$.

If $\varphi_{h_l}(\hat{P}) = c_l$, then $\varphi_{h_{l-1}}(\hat{P}) \neq c_l$. Then let $\hat{P}'_{h_{l-1}} : c_lc_{l-1}$ and $\hat{P}' = (\hat{P}'_{h_{l-1}}, \hat{P}_{-h_l})$. By strategyproofness and efficiency, $\varphi_{h_{l-1}}(\hat{P}') = \emptyset$ or $\varphi_{h_{l-1}}(\hat{P}') = c_{l-1}$. In the latter case, again we have $\varphi_{h_{l-2}}(\hat{P}') \neq c_{l-1}$, and so on until each agent h_l receives c_l and we find a contradiction to efficiency. Thus, at some point for $h_t \in I_k \backslash J_{c^1}$ and $P_{h_t} : c_{t+1} c_t$ we have for the constructed profile R, $\varphi_{h_t}(R) = \emptyset$. Then let $R''_{h_t} : c_t$ and $R'' = (R''_{h_t}, R_{-h_t})$.

But then set $c^2 \equiv c_t$. Analogous successive applications of the above arguments show that there exists J_{c^2} and a preference profile $R''_{J_{c_2}}$ such that for all $i \in J_{c^2}$, $R''_i : c^2$, and a preference profile $R^*_{-J_{c^1} \cup J_{c^2} \cup I_{<k}}$ for students $N \setminus (J_{c^1} \cup J_{c^2} \cup I_{<k})$ such that

- for each $i \in J_{c^2}$, R''_i is a preference relation for student i at which c^2 is the only acceptable school,
- $\varphi_h(R'_{J_{c^1}}, R''_{J_{c^2}}, R_{I_{<k}}, R^*_{-J_{c^1} \cup J_{c^2} \cup I_{<k}}) = c^1$ for some $h \in J_{c^1}$,
- $\varphi_h(R'_{J_{c^1}}, R''_{J_{c^2}}, R_{I_{<k}}, R^*_{-J_{c^1} \cup J_{c^2} \cup I_{<k}}) = c^2 \neq TTC_h(R'_{J_{c^1}}, R''_{J_{c^2}}, R_{I_{<k}}, R^*_{-J_{c^1} \cup J_{c^2} \cup I_{<k}})$ for some $h \in J_{c^2}$, and
- for any preference profile $R^{**}_{-J_{c^1} \cup J_{c^2} \cup I_{<k}}$ for students $N \setminus (J_{c^1} \cup J_{c^2} \cup I_{<k})$, we have $\varphi_{c^2}(R'_{J_{c^1}}, R''_{J_{c^2}}, R_{I_{<k}}, R^{**}_{-J_{c^1} \cup J_{c^2} \cup I_{<k}}) \notin N \setminus (J_{c^1} \cup J_{c^2} \cup I_{<k})$.

Let $P'' = (R'_{J_{c^1}}, R''_{J_{c^2}}, R_{I_{<k}}, R^*_{-J_{c^1} \cup J_{c^2} \cup I_{<k}})$. If for some profile $R = (R'_{J_{c^1}}, R''_{J_{c^2}}, R_{I_{<k}}, R^{**}_{-J_{c^1} \cup J_{c^2} \cup I_{<k}})$ and some $i \in I_k \backslash (J_{c^1} \cup J_{c^2})$ we have $\varphi_i(R) \neq TTC_i(R)$, then we do the same as above and find c^3 and J_{c^3} together with a profile $R'''_{J_{c^3}}$ (and continue).

Otherwise we have for any profile $R = (R'_{J_{c^1}}, R''_{J_{c^2}}, R_{I_{<k}}, R^{**}_{-J_{c^1} \cup J_{c^2} \cup I_{<k}})$ and all $i \in I_k \backslash (J_{c^1} \cup J_{c^2})$, $\varphi_i(R) = TTC_i(R)$.

Now consider P'' and I_{k+1}. If for some $i \in I_{k+1} \backslash (J_{c^1} \cup J_{c^2})$, $\varphi_i(R) \neq TTC_i(R)$, then we find as above c^3 and J_{c^3}, and so on.

Thus, we find $\{c^1, \ldots, c^q\}$ and mutually disjunct sets $J_{c^1}, \ldots, J_{c^q}$ and $I_{<k}$ such that for $P^{(q)} = (R'_{J_{c^1}}, R''_{J_{c^2}}, \ldots, R^{(q)}_{J_{c^q}}, R_{I_{<k}}, R^*_{-J_{c^1} \cup J_{c^2} \cup \cdots \cup J_{c^q} \cup I_{<k}})$ we have

- for each $i \in J_{c^p}$ (with $p \in \{1, \ldots, q\}$), $R^{(p)}_i$ is a preference relation for student i at which c^p is the only acceptable school,
- for each $p \in \{1, \ldots, q-1\}$, $\varphi_h(P^{(p)}) = c^p$ for some $h \in J_{c^p}$,
- $\varphi_h(P^{(q)}) = c^q \neq TTC_h(P^{(q)})$ for some $h \in J_{c^q}$, and
- $\varphi_i(P^{(q)}) = TTC_i(P^{(q)})$ for all $i \in N \backslash (J_{c^1} \cup \cdots \cup J_{c^q})$.

Let $\mu = \varphi(P^{(q)})$ and $\nu = TTC(P^{(q)})$. Because φ is f-more stable than TTC, we have $\mu \succsim_f^{P^{(q)}} \nu$. Now we will successively remove in the order $J_{c^1}, \ldots, J_{c^q}$.

If $\mu(i) = \nu(i)$ for all $i \in J_{c^1}$, then we have $B_i(\mu) = B_i(\nu) = \emptyset$ for all $i \in J_{c^1}$ and $B_{c^1}(\mu) = \emptyset = B_{c^1}(\nu)$. Thus, by the fact that J_{c^1} is isolated under both μ and ν and consistency of f we obtain $\mu|_{N\setminus J_{c^1}} \gtrsim_f^{P^{(q)}_{N\setminus J_{c^1}}} \nu|_{N\setminus J_{c^1}}$. Otherwise, $B_i(\nu) = \emptyset$ for all $i \in J_{c^1}$ and $B_{c^1}(\nu) = \emptyset$ but for some $j \in J_{c^1}$, $B_j(\mu) \neq \emptyset$. Furthermore, $\mu(J_{c^1}) = \{c^1\} = \nu(J_{c^1})$ and J_{c^1} is isolated under both μ and ν. But then by separability of f we obtain $\mu|_{N\setminus J_{c^1}} \gnsim_f^{P^{(q)}_{N\setminus J_{c^1}}} \nu|_{N\setminus J_{c^1}}$. Note that in both cases we obtain $\mu|_{N\setminus J_{c^1}} \gtrsim_f^{P^{(q)}_{N\setminus J_{c^1}}} \nu|_{N\setminus J_{c^1}}$.

Then we continue with J_{c^2} and obtain $\mu|_{N\setminus(J_{c^1}\cup J_{c^2})} \gtrsim_f^{P^{(q)}_{N\setminus(J_{c^1}\cup J_{c^2})}} \nu|_{N\setminus(J_{c^1}\cup J_{c_2})}$, and so on until we obtain $\mu|_{N\setminus(J_{c^1}\cup\cdots\cup J_{c^{q-1}})} \gtrsim_f^{P^{(q)}_{N\setminus(J_{c^1}\cup\cdots\cup J_{c^{q-1}})}} \nu|_{N\setminus(J_{c^1}\cup\cdots\cup J_{c_{q-1}})}$. Note that by construction under $P_{N\setminus(J_{c^1}\cup\cdots\cup J_{c^{q-1}})}$, for J_{c^q} we have $B_i(\nu|_{N\setminus(J_{c^1}\cup\cdots\cup J_{c_{q-1}})}) = \emptyset$ for all $i \in J_{c^q}$ and $B_{c^q}(\nu|_{N\setminus(J_{c^1}\cup\cdots\cup J_{c_{q-1}})}) = \emptyset$ but for some $j \in J_{c^q}$, $B_j(\mu|_{N\setminus(J_{c^1}\cup\cdots\cup J_{c_{q-1}})}) \neq \emptyset$. Furthermore, $\mu(J_{c^q}) = \{c^q\} = \nu(J_{c^q})$ and J_{c^q} is isolated under both μ and ν (and their restrictions). But then by separability of f we obtain

$$\mu|_{N\setminus(J_{c^1}\cup\cdots\cup J_{c^q})} \gnsim_f^{P^{(q)}_{N\setminus(J_{c^1}\cup\cdots\cup J_{c^q})}} \nu|_{N\setminus(J_{c^1}\cup\cdots\cup J_{c_q})}.$$

This is a contradiction as for all $i \in N\setminus(J_{c^1} \cup \cdots \cup J_{c^q})$, $\mu(i) = \nu(i)$.[13] ■

Concluding the proof: We show, by induction, that φ and TTC coincide on the domain of problems with n students, which contradicts that φ is *more stable* than TTC, and concludes the proof.

Base case: For any problem with preference profile R including n students, for each student i who is assigned a seat at Step 1 of the TTC algorithm, $\varphi_i(R) = TTC_i(R)$. This follows from Lemma 2.

Inductive step: Assume that for any problem with preference profile R including n students, for each student i who is assigned a seat at an earlier step than Step k of the TTC algorithm, $\varphi_i(R) = TTC_i(R)$. We will show that for each student j who is assigned a seat at Step k of the TTC algorithm, $\varphi_j(R) = TTC_j(R)$.

Let I_k denote the set of students who are assigned a seat at Step k of $TTC(R)$ and C_k denote the set of schools that are allocated at Step k of $TTC(R)$. Note that if for each $i \in I_k$, $\varphi_k(R) \in C_k$, then by *efficiency*, $\varphi_i(R) = TTC_i(R)$ for each $i \in I_k$.

Suppose that there exists $i_1 \in I_k$ such that $\varphi_{i_1}(R) \notin C_k$. Let $c_1 \in C_k$ be the school that points to i_1 in Step k of $TTC(R)$. Let R'_{i_1} be a preference ordering for i_1 at which $TTC_{i_1}(R)$

[13] Note that $k > 1$ and $I_{<k} \neq \emptyset$.

is top-ranked and c_1 is second-ranked. By strategy-proofness, $\varphi_{i_1}(R'_{i_1}, R_{-i_1}) \neq TTC_{i_1}(R)$. By Lemma 3 and *strategy-proofness*, $\varphi_{i_1}(R'_{i_1}, R_{-i_1}) = c_1$.

Note that I_k is still the set of students who are assigned a seat at Step k of $TTC(R'_{i_1}, R_{-i_1})$ and C_k is still the set of schools that are allocated at Step k of $TTC(R'_{i_1}, R_{-i_1})$. Now, if for each $i \in I_k \setminus \{i_1\}$, $\varphi_i(R) \in C_k$, then *efficiency* would imply that $\varphi_i(R'_{i_1}, R_{-i_1}) = TTC_i(R'_{i_1}, R_{-i_1})$ for each $i \in I_k$, which would contradict $\varphi_{i_1}(R'_{i_1}, R_{-i_1}) \neq TTC_{i_1}(R) = TTC_{i_1}(R'_{i_1}, R_{-i_1})$. Thus, there exists $i_2 \in I_k \setminus \{i_1\}$ such that $\varphi_{i_2}(R) \notin C_k$. Let $c_2 \in C_k$ be the school that points to i_2 in Step k of $TTC(R)$. Let R'_{i_2} be a preference ordering for i_2 at which $TTC_{i_2}(R)$ is top-ranked and c_2 is second-ranked. By strategy-proofness, $\varphi_{i_2}(R'_{i_1}, R'_{i_2}, R_{-\{i_1,i_2\}}) \neq TTC_{i_2}(R)$. By Lemma 3 and *strategy-proofness*, $\varphi_{i_2}(R'_{i_1}, R'_{i_2}, R_{-\{i_1,i_2\}}) = c_2$.

Continuing in a similar fashion, we identify a list of students $(i_1, \ldots, i_m)$ and a preference profile $R' = (R'_1, \ldots, R'_m)$ such that $\{i_1, \ldots, i_m\} = I_k$, $\varphi_i(R') \in C_k$ for each $i \in I_k$, and $\varphi_{i_m}(R') \neq TTC_{i_m}(R')$, which contradicts *efficiency* of φ. ■

Remark 1 *When schools may have multiple available seats, there exists a strategy-proof and efficient mechanism φ such that there exists a problem where the φ assignment is stable while the TTC assignment is unstable, and at any other problem where the φ assignment is different from the TTC assignment, the φ assignment is stable (Abdulkadiroğlu et al., 2019). Therefore, as long as the stability measure f satisfies stability preferred, TTC is not f-minimally unstable among efficient and strategyproof mechanisms when schools may have multiple seats.*

5 Applications

Below, we apply our main result to different natural stability comparison methods. Some of them are inclusion methods whereas others are the (corresponding) cardinal methods. Furthermore, it is easy to verify that any of the comparison methods below satisfies stability preferred, separability and consistency, and hence, Theorem 1 holds.

This shows the robust minimal instability of TTC among efficient and stretegy-proof mechanisms (with unit capacities).

5.1 Blocking Pairs

The blocking pairs inclusion comparison (*pincl*) is defined as follows. For each problem $P \in \mathcal{P}$ and $\mu, \nu \in \mathcal{A}(P)$,

$$\mu \succsim^P_{pincl} \nu \Leftrightarrow B(\mu) \subseteq B(\nu).$$

Among others, Abdulkadiroğlu et al. (2019) and Tang and Zhang (2017) study this stability comparison.

Corollary 1 *[Theorem 1 of Abdulkadiroğlu, Che, Pathak, Roth and Tercieux, 2019] TTC is weakly pincl-minimally unstable among efficient and strategyproof mechanisms when each school has unit capacity.*

The blocking pairs cardinality comparison (*pcard*) is defined as follows. For each problem $P \in \mathcal{P}$ and $\mu, \nu \in \mathcal{A}(P)$,

$$\mu \succsim^P_{pcard} \nu \Leftrightarrow |B(\mu)| \leq |B(\nu)|.$$

Note that $\succsim^P_{pincl} \subseteq \succsim^P_{pcard}$.[14]

Corollary 2 *TTC is pcard-minimally unstable among efficient and strategyproof mechanisms when each school has unit capacity.*

Obviously, for any problem P: (i) $\succsim^P_{pincl} \subseteq \succsim^P_{pcard}$, (ii) $\succsim^P_{pincl}$ is transitive but not complete, and (iii) $\succsim^P_{pcard}$ is complete (as any two assignments can be compared) and transitive. Hence, Corollary 2 implies Corollary 1.

5.2 Blocking Triplets

The blocking triplets inclusion comparison (*tincl*) is defined as follows. Let $(i, j, c) \in T(\mu)$ if and only if $i \succ_c j$, $\mu(j) = c$, and $cP_i\mu(i)$. For each $P \in \mathcal{P}$ and $\mu, \nu \in \mathcal{A}(P)$,

$$\mu \succsim^P_{tincl} \nu \Leftrightarrow T(\mu) \subseteq T(\nu).$$

Kwon and Shorrer (2019) study this stability comparison.

Corollary 3 *[Proposition 7 of Kwon and Shorrer, 2019] TTC is weakly tincl-minimally unstable among efficient and strategyproof mechanisms when each school has unit capacity.*

[14] Doğan and Ehlers (2020) study this stability comparison method for efficient assignments.

The blocking triplets cardinality comparison *tcard* is defined as follows. For each $P \in \mathcal{P}$ and $\mu, \nu \in \mathcal{A}(P)$,

$$\mu \succsim^P_{tcard} \nu \Leftrightarrow |T(\mu)| \leq |T(\nu)|.$$

Note that $\succsim^P_{tincl} \subseteq \succsim^P_{tcard}$.

Corollary 4 *TTC is tcard-minimally unstable among efficient and strategyproof mechanisms when each school has unit capacity.*

Obviously, for any problem P: (i) $\succsim^P_{tincl} \subseteq \succsim^P_{tcard}$, (ii) $\succsim^P_{tincl}$ is transitive but not complete, and (iii) $\succsim^P_{tcard}$ is complete and transitive. Hence, Corollary 4 implies Corollary 3.

5.3 Blocking Students

The blocking students inclusion comparison (*sincl*) is defined as follows. Let $BS(\mu) = \{i \in N : B_i(\mu) \neq \emptyset\}$. For each problem $P \in \mathcal{P}$ and $\mu, \nu \in \mathcal{A}(P)$,

$$\mu \succsim^P_{sincl} \nu \Leftrightarrow BS(\mu) \subseteq BS(\nu).$$

Corollary 5 *TTC is sincl-minimally unstable among efficient and strategyproof mechanisms when each school has unit capacity.*

The blocking students cardinality comparison (*scard*) is defined as follows. For each $P \in \mathcal{P}$ and $\mu, \nu \in \mathcal{A}(P)$,

$$\mu \succsim^P_{scard} \nu \Leftrightarrow |BS(\mu)| \leq |BS(\nu)|.$$

Corollary 6 *TTC is scard-minimally unstable among efficient and strategyproof mechanisms when each school has unit capacity.*

Obviously, (i) $\succsim^P_{sincl} \subseteq \succsim^P_{scard}$, (ii) $\succsim^P_{sincl}$ is transitive but not complete and (iii) $\succsim^P_{scard}$ is complete and transitive.[15]

6 Appendix

The examples below show that the three properties, *stability preferred*, *separability*, and *consistency*, are independent for stability comparison methods.

[15] Doğan and Ehlers (2020) also study the stability comparison methods based on blocking students.

Example 1 (Only stability preferred violated) *Consider the following stability comparison $\gtrsim = \emptyset$, that is, for any problem P and any $\mu, \nu \in \mathcal{A}(P)$, μ and ν are incomparable in terms of $\gtrsim^P$, i.e $\gtrsim^P = \emptyset$. Note that separability and consistency are vacuously satisfied, while stability preferred is clearly violated.*

Example 2 (Only separability violated) *Consider the following stability comparison $\gtrsim$. For any $P \in \mathcal{P}$ and $\mu, \nu \in \mathcal{A}(P)$, let $\mu \gtrsim_f^P \nu$ if and only if $B(\nu)$ is not a proper subset of $B(\mu)$, i.e., $B(\nu) \not\subsetneq B(\mu)$. Note that $\mu \gnsim_f^P \nu$ if and only if $B(\mu) \subsetneq B(\nu)$.*

Clearly, stability preferred is satisfied. To see that consistency is satisfied, take any μ and ν such that $\nu \gtrsim^P \mu$ and for some $\emptyset \neq N' \subseteq N$, $\nu(i) = \mu(i)$ for all $i \in N'$, $B_i(\mu) = B_i(\nu) \subseteq \mu(N')$ for all $i \in N'$, and $B_c(\mu) = B_c(\nu) \subsetneq N'$ for all $c \in \mu(N') = \nu(N') = C'$ and N' is isolated under both μ and ν. But then, $B(\mu|_{N \setminus N'}) \not\subsetneq B(\nu|_{N \setminus N'})$ and therefore $\nu|_{N \setminus N'} \gtrsim_f^{P_{N \setminus N'}} \mu|_{N \setminus N'}$.

To see that separability is violated, consider the following problem P where $N = \{1, 2, 3, 4\}$ and $C = \{c_1, c_2, c_3, c_4\}$. Only the relevant top parts of the preference and priority profiles are depicted.

R_1	R_2	R_3	R_4	$\succeq_{c_1}$	$\succeq_{c_2}$	$\succeq_{c_3}$	$\succeq_{c_4}$
c_2	c_3	c_3	c_4	1	3	2	3
c_1	c_2	c_4			1	3	4
		c_2			2		

Consider

$$\mu = \begin{pmatrix} 1 & 2 & 3 & 4 \\ c_1 & c_3 & c_2 & c_4 \end{pmatrix}$$

$$\nu = \begin{pmatrix} 1 & 2 & 3 & 4 \\ c_1 & c_2 & c_3 & c_4 \end{pmatrix}$$

where $B(\mu) = \{(3, c_4)\}$ and $B(\nu) = \{(1, c_2), (2, c_3)\}$. Let $N' = \{1\}$. Note that $\nu \gtrsim_f^P \mu$, $\mu_{N'}$ is N'-stable, $\mu(N') = \nu(N')$, and $B_1(\nu) \neq \emptyset$ where $1 \in N'$. Moreover, $\mu|_{N \setminus N'} \gtrsim_f^{P_{N \setminus N'}} \nu|_{N \setminus N'}$, implying that separability is violated.

Example 3 (Only consistency violated) *Consider the following stability comparison $\gtrsim$. For any problem $P \in \mathcal{P}$ and $\mu, \nu \in \mathcal{A}(P)$, let $\mu \gtrsim_f^P \nu$ if and only if $B(\mu) = \emptyset$ or ($|B(\nu)| \geq 2$ and $|B(\mu)| \leq |B(\nu)|$). Note that $\mu \gnsim_f^P \nu$ if and only if $B(\mu) = \emptyset \neq B(\nu)$ or ($|B(\nu)| \geq 2$ and $B(\mu) < |B(\nu)|$). Also note that when $|B(\mu)| = |B(\nu)| = 1$, μ and ν are incomparable in terms of $\gtrsim^P$.*

By definition, stability preferred is satisfied. To see that separability is satisfied, take any μ and ν such that $\nu \gtrsim_f^P \mu$, and take any $\emptyset \neq N' \subseteq N$ such that μ is N'-stable,

$\mu(N') = \nu(N') = C'$ and $B_i(\nu) \neq \emptyset$ for some $i \in N'$. Note that $|B(\mu|_{N\setminus N'})| \geq 2$ and $|B(\nu|_{N\setminus N'})| < |B(\mu|_{N\setminus N'})|$. Hence, $\nu|_{N\setminus N'} \gtrsim_f^{P_{N\setminus N'}} \mu|_{N\setminus N'}$.

To see that consistency is violated, consider the following problem P where $N = \{1,2,3,4,5\}$ and $C = \{c_1, c_2, c_3, c_4, c_5\}$. Only the relevant top parts of the preference and priority profiles are depicted.

R_1	R_2	R_3	R_4	R_5	$\succeq_{c_1}$	$\succeq_{c_2}$	$\succeq_{c_3}$	$\succeq_{c_4}$	$\succeq_{c_5}$
c_2	c_2	c_4	c_4	c_5	1	1	3	3	4
c_1		c_3	c_5			2		4	5
			c_3						

Consider

$$\mu = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ c_1 & c_2 & c_3 & c_4 & c_5 \end{pmatrix}$$

$$\nu = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ c_1 & c_2 & c_4 & c_3 & c_5 \end{pmatrix}$$

where $B(\mu) = \{(1, c_2), (3, c_1)\}$ and $B(\nu) = \{(1, c_2), (4, c_5)\}$. Let $N' = \{1, 2\}$.

Note that $\nu \gtrsim_f^P \mu$, $\nu(i) = \mu(i)$ for all $i \in N'$, $B_i(\mu) = B_i(\nu) \subseteq \mu(N')$ for all $i \in N'$, and $B_c(\mu) = B_c(\nu) \subseteq N'$ for all $c \in \mu(N') = \nu(N') = C'$. Yet, $\nu|_{N\setminus N'}$ and $\mu|_{N\setminus N'}$ are incomparable, implying that consistency is violated.

References

Abdulkadiroğlu, Atila and Tayfun Sönmez, "School choice: A mechanism design approach," *American Economic Review*, June 2003, *93* (3), 729–747.

_ , Yeon-Koo Che, Parag A. Pathak, Alvin E. Roth, and Olivier Tercieux, "Efficiency, Justified Envy, and Incentives in Priority-Based Matching," Technical Report 2019.

Doğan, Battal and Lars Ehlers, "Efficient and Minimally Unstable Pareto Improvement over the Deferred Acceptance Mechanism," *Working paper*, 2020.

Hakimov, Rustamdjan and Onur Kesten, "The Equitable Top Trading Cycles Mechanism for School Choice," *International Economic Review*, 2018, *59*, 2219 – 2258.

Kwon, Hyukjun and Ran I. Shorrer, "Justified-Envy Minimal Mechanisms in School Choice," *Working Paper Available at SSRN: https://ssrn.com/abstract=3495266*, 2019.

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

Morrill, Thayer, "Making Just School Assignments," *Games and Economic Behavior*, 2015, *92*, 18 – 27.

_ , "Two Simple Variations of Top Trading Cycles," *Economic Theory*, 2015, *60*, 123 – 140.

Roth, Alvin E., "The Economics of Matching: Stability and Incentives," *Mathematics of Operations Research*, 1982, *7*, 617–628.

Shapley, Lloyd and Herbert Scarf, "On cores and indivisibility," *Journal of Mathematical Economics*, 1974, *1* (1), 23 – 37.

Svensson, Lars-Gunnar, "Strategy-proof allocation of indivisible goods," *Social Choice and Welfare*, 1999, *16*, 557–567.

Tang, Qianfeng and Yongchao Zhang, "Weak Stability and Pareto Efficiency in School Choice," *Working Paper Available at SSRN: https://ssrn.com/abstract=2972611*, 2017.