ECONtribute
MARKETS & PUBLIC POLICY

ECONtribute
Discussion Paper No. 306

Transparent Matching Mechanisms

Markus Möller

May 2024

www.econtribute.de

UNIVERSITÄT BONN
Universität zu Köln

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1-390838866 is gratefully acknowledged.

Transparent Matching Mechanisms*

Markus Möller

January 4, 2024

Abstract

I study a central authority's ability to commit to a publicly announced mechanism in a one-to-one agent-object matching model. The authority announces a strategy-proof mechanism and then privately selects a mechanism to initiate a matching. An agent's observation in form of the final matching has an *innocent explanation* (Akbarpour and Li, 2020), if given the agent's reported preferences, there is a combination with other agents' preferences leading to an identical observation under the announced mechanism. The authority can only commit up to *safe deviations* (Akbarpour and Li, 2020)—mechanisms that produce only observations with innocent explanations. For efficient or stable announcements, I show that no safe deviation exists if and only if the announced mechanism is dictatorial. I establish that the *Deferred Acceptance (DA) Mechanism* (Gale and Shapley, 1962) implies commitment to stability. Finally, I show that group strategy-proof and efficient announcements allow commitment to efficiency only if they are dictatorial.

Keywords: Matching, Transparency, Partial Commitment, Strategy-Proof, Stability, Efficiency, DA, TTC.

JEL Codes: C78, D47, D82.

1 Introduction

In matching theory the central authority usually appears as an honest and faultless operator of the matching mechanism. In practice, however, the authority's conduct

*Thanks to Alexander Westkamp, Yiqiu Chen, Aram Grigoryan, two anonymous referees and the co-editor for providing valuable feedback. I acknowledge financial support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC 2126/1– 390838866. University of Bonn. E-mail: mmoelle2@uni-bonn.de

can be in conflict with announcements made to participants in advance. For instance, as a part of a recent bribery affair at U.S. colleges, some officials have used fake athletic credentials to influence the admissions process in favor of certain applicants.[1] Also in the context of public school assignment, dozens of students were wrongly denied entry to Boston's most prestigious exam schools in 2020. The assignment is conducted by Boston Public Schools (BPS), whose officials declared the instance resulted from failed internal communication. Apparently overseen by BPS's internal audit, the deviation was detected by a student's tutor.[2] In spite of an external audit, a similar case occurred in 2023.[3,4]

This paper asks to what extent participants can be confident that the authority sticks to the announcements made. I employ a one-to-one object allocation model without monetary transfers, where an authority makes a public announcement in form of a strategy-proof direct mechanism.[5] Then, upon receiving agents' preferences, the authority privately selects a mechanism to induce a publicly observable matching. I then adapt the notions of *innocent explanations* and *safe deviations* of Akbarpour and Li (2020). Concretely, an observation in form of the final matching has an innocent explanation for the observing agent if, given her own preferences, there is a possible combination of other agents' preferences that would lead to an identical observation under the announced mechanism. Furthermore, a mechanism is a safe deviation with respect to an announcement, if for each agent, each observation produced by the mechanism has an innocent explanation. An announcement is *transparent* if it has no safe deviations. Neither the agents nor the authority are strategic players in my model. Therefore, different from the commitment criteria studied in Akbarpour and Li (2020), transparency covers unintentional deviations.[6]

[1] https://www.justice.gov/usao-ma/pr/arrests-made-nationwide-college-admissions-scam-alleged-exam-cheating-athletic.

[2] https://www.bostonglobe.com/2020/08/31/metro/boston-public-schools-announces-error-exam-school-admissions-that-kept-dozens-out-recent-years/

[3] https://www.bostonglobe.com/2023/04/12/metro/bps-miscalculated-student-gpas-wrongly-informing-students-they-were-eligible-apply-exam-schools/.

[4] Similarly, in Chicago, various public school officials did not follow announced admissions rules in the admission year 2016-2017, including instances of privileged treatment, documentation errors and screening of applicants (Grigoryan and Möller, 2023; Schuler, 2018). Furthermore, in 1995, the famous National Residency Matching Program (NRMP) failed to follow its promise to use a mechanism that is not manipulable by residents (Williams, 1995; Roth and Peranson, 1997).

[5] Strategy-proof direct mechanisms are still highly prevalent in practice and serve as an important benchmark.

[6] However, there are interesting cases that go beyond the scope this framework. This applies

I consider an informational benchmark where agents' preference rankings over objects are private information while market features that are harder to hide from outsiders are common knowledge. Specifically, the set of agents, the set of objects and all scores are publicly known and each agent observes the final matching. In fact, while a student's reported ranking over schools typically remains confidential even after the matching has been determined, her final assignment and traits such as her walk-zone, or particular abilities can be hard to conceal. Furthermore, whereas students may not know their exact rank number on a school's or college's priority list, they can have a good idea about their relative ranking. Except Proposition 2 all results for priority-based allocation transfer to a setting that is closer to common features of public school assignment (Abdulkadiroğlu and Sönmez, 2003) and college admission (Balinski and Sönmez, 1999).[7]

I show that the unique strategy-proof and stable mechanism, known as the *Deferred Acceptance (DA)* mechanism (Gale and Shapley, 1962), is transparent if and only if it is a *serial dictatorship* (Satterthwaite and Sonnenschein, 1981; Svensson, 1994) (Proposition 1). I then establish that strategy-proof and efficient mechanisms are transparent if and only if they are equivalent to a *sequential dictatorship* (Pápai, 2001; Ehlers and Klaus, 2003; Pápai, 2000) (Theorem 2). It is well-known that sequential dictatorships satisfy various desirable properties. For instance, Pycia and Troyan (2023) recently established that sequential dictatorships are among the few candidates that achieve high standards for strategic simplicity.

I also ask whether weaker forms of transparency can be achieved for non-dictatorial mechanisms. Specifically, I explore whether the authority can commit to desirable properties of her announcement. For example, in applications such as public school assignment or college admission properties such as stability or efficiency are usually perceived as desirable. Yet whether the authority can commit to induce a particular property has not been explicitly studied. I show that a deviation from DA is safe if and only if the deviation is a stable mechanism with respect to the underlying priorities of the market (Theorem 1). Furthermore, a group strategy-proof[8] and

in particular to certain forms of bribery that affect the agents' private information or where the authority can pay bribes to agents.

[7]More concretely, these results apply in a setup, where each agent knows her own scores at each object and only observes her own assignment along with a set of object-specific cutoffs disclosed by the authority. Given a final assignment, the cutoff at an object is the score of the agent with lowest score assigned to this object.

[8]A mechanism is group strategy-proof if there is no group of agents that can generate weakly

efficient mechanism has no inefficient safe deviation if and only if it is a sequential dictatorship (Theorem 3).

Finally, I consider a special case of my model where the authority commits to use a strategy-proof mechanism and agents are strategic. I compare the transparency of DA and the *Top Trading Cycles (TTC)* mechanism (Shapley and Scarf, 1974; Abdulkadiroğlu and Sönmez, 2003) that have both been touted as candidates for assigning students to Boston Public Schools in 2005.[9] I show that while DA is transparent in this setup, TTC is transparent if the priority structure satisfies an acyclicity condition. The condition is weaker than similar conditions that characterize TTC regarding various desirable properties. I also provide necessary conditions for TTC to be transparent in this setup.

Related Work

This paper is among the first to relax the authority's full commitment assumption in the context of matching markets. However, there are some recent studies that offer complementary perspectives on the topic.

Independent from this work, Grigoryan and Möller (2023) follow a non-binary approach to study auditability in allocation problems. They introduce an auditability index based on the minimum-sized group of individuals whose information is sufficient to detect any deviation. While under the *Immediate Acceptance (IA)* mechanism there are two agents whose information is enough to detect any deviation, under the DA and TTC some detections need access to all agents' information. Interestingly, in Grigoryan and Möller (2023) sequential dictatorships can be hard to audit. However, the results on sequential dictatorships are not logically connected.

Hakimov and Raghavan (2023) introduce a form of transparency in allocation problems that is always achievable through sequential public disclosure of interim cutoffs and private feedback.[10] They show that DA, TTC and IA can be induced in a transparent way with a simple sequential protocol that asks each agent to only report one object at a time. A key difference is that in the current paper, mechanisms are static, communication between agents and authority remains private and that

better assignments for all members in the group by misrepresenting their preferences such that at least one agent in the group strictly profits from the misrepresentation.

[9]The committee ultimately chose DA, arguing that *"the behind the scenes mechanized trading [in TTC] makes the student assignment process less transparent."* (Leshno and Lo, 2020).

[10]However, neither the private feedback nor the cutoffs alone are sufficient for transparency.

no information is disclosed by the authority. Thus, transparency is a feature of the mechanism, whereas in Hakimov and Raghavan (2023) transparency is a consequence of designing the general information structure.[11] More generally, different from these concurrent works, I identify necessary and sufficient conditions for entire classes of strategy-proof direct mechanisms and examine commitment to desirable properties. Moreover, I consider a special case of my model where agents' strategic behavior is taken into account.[12]

Akbarpour and Li (2020) and Woodward (2020) study partial commitment in the context of auctions. Akbarpour and Li (2020) develop a general partial commitment framework with sequential private communication between the authority and agents and focus on Bayes-Nash implementation with imperfect information. The key difference to the notions of Akbarpour and Li (2020) and Woodward (2020) is that transparency does not require incentive compatibility for the authority with respect to a known objective function. In fact, in their works, all deviations are intentional by design.[13]

More broadly this paper contributes to our understanding of the structure and verifiability of matching mechanisms (Gonczarowski and Thomas, 2023; Hakimov and Raghavan, 2023; Gangam et al., 2023) and connects to the literature which models limited commitment as measurable with respect to agents' observations on final outcomes (Dequiedt and Martimort, 2015; Baliga et al., 1997; Bester and Strausz, 2000, 2001).

The rest of this paper is organized as follows. Section 2 introduces the basic model along with the partial commitment framework. Section 3 analyzes the transparency characteristics of stable mechanisms. Section 4 contains the analysis of efficient mechanisms. Section 5 studies priority-based allocation with partial commitment and strategic agents.

[11] Note that in both Grigoryan and Möller (2023) and Hakimov and Raghavan (2023) no agent has ex-ante information about the scores of the other agents and only observes her own assignment.

[12] In the context of Arrovian efficiency, Pycia and Ünver (2023) show that for group strategy-proof and Pareto efficient mechanisms any deviation could be unveiled by comparing a single agent's relative ranking of the outcome and a challenger alternative. Yet the identity of the agent and the challenger alternative are usually not known. Thus, their notion is substantially weaker than transparency.

[13] The notions of Akbarpour and Li (2020) and Woodward (2020) are very natural for the context of auctions, since auctioneers are often interested in maximizing revenue. Thus, a sophisticated bidder may have a clear picture of the auctioneer's incentives.

2 Preliminaries

2.1 The Model

Let I be a set of agents and $X \cup \{\emptyset\}$ a set of indivisible objects, where $\emptyset$ denotes the outside option for agents. Throughout the paper, I fix the set of agents and objects and assume $|X| \geq 2$ and $|I| \geq 2$. Let i, j, k denote generic agents in I and let x, y, z refer to generic objects in $X \cup \{\emptyset\}$.

Each object $x \in X$ has a vector of scores $s^x = \{s_i^x\}_{i \in I}$, where $s_i^x \in \mathbf{R}^{++}$ is i's score at object x. We assume that $s_i^x \neq s_j^x$ for any $i, j \in I$ and any $x \in X$, and we say that for each pair of agents $i, j \in I$, i has higher priority at x than j if and only if $s_i^x > s_j^x$. That is, for each object x, the object's scores induce a strict priority ranking over I. For each $i \in I$, let $s_i = \{s_i^x\}_{x \in X}$ be the vector of scores assigned to agent i. Let a score (or priority) structure $s = (s_i)_{i \in I}$ be a collection of scores for each agent and let $s_{-i} = (s_j)_{j \in I \setminus \{i\}}$ be a collection of scores for agents in $I \setminus \{i\}$. Set $\wp$ as the domain of all possible priority structures. For the rest of this paper, fix an arbitrary priority structure $s \in \wp$.

Each agent $i \in I$ has a strict preference relation P_i over $X \cup \{\emptyset\}$, where R_i is the corresponding weak preference relation.[14] For each $x \in X$ and $i \in I$, object x is acceptable if $xP_i\emptyset$ for i and x is unacceptable for i if it is not acceptable. I refer to P_i as agent i's preferences and to $P \equiv (P_i)_{i \in I}$ as a preference profile. For each $i \in I$, let $\mathcal{P}_i$ be the domain of all possible preferences and let $\mathcal{P} = \times_{i \in I} \mathcal{P}_i$ be the domain of all preference profiles. For any $J \subset I$, $P_J = (P_j)_{j \in J}$ is a preference profile for agents J, where $\mathcal{P}_J \equiv \times_{j \in J} \mathcal{P}_j$ is the corresponding domain. Denote with $-i$ the set of all agents except agent i.

A *matching* is a function $\mu : I \to X \cup \{\emptyset\}$ under which each object $x \in X$ ends up with at most one agent and any agent $i \in I$, who is not assigned to some object $x \in X$, is assigned to $\emptyset$. Let $\mathcal{M}$ collect the set of all possible matchings and for each $\mu \in \mathcal{M}$, denote with μ_i the object that is assigned to agent $i \in I$. For any $\mu \in \mathcal{M}$, let μ_X be the set of objects from X assigned to agents under μ and define μ_I symmetrically.

Consider some matching $\mu \in \mathcal{M}$ and some preference profile $P \in \mathcal{P}$. The matching μ is *non-wasteful* if there exists no $i \in I$ and no object $x \in X$ such that $xP_i\mu_i$ and x is unassigned under μ. Call the matching μ *individually rational* if, for each $i \in I$,

[14] Hence R_i is a complete, transitive and anti-symmetric binary relation. For each pair of objects $x, y \in X \cup \{\emptyset\}$, I write xR_iy if either xP_iy or $x = y$.

$\mu_i R_i \emptyset$. The matching μ is *blocked* if there exists a pair of agents $i, j \in I$ and an object $x \in X$ such that $xP_i\mu_i$, $\mu_j = x$ and $s_i^x > s_j^x$. A matching μ is *stable* (with respect to priority structure s) if it is not blocked, individually rational and non-wasteful. Let $\Sigma^s(P)$ be the set of stable matchings for preference profile P (with respect to s). Next, let a matching $\nu \in \mathcal{M}$ *weakly Pareto dominate* matching μ if, for each $i \in I$, $\nu_i R_i \mu_i$, and say that ν *strictly Pareto dominates* μ, if ν weakly Pareto dominates μ and there exists an agent $j \in I$ with $\nu_j P_j \mu_j$. The matching μ is *(Pareto) efficient* if there exists no matching that strictly Pareto dominates it.

A *mechanism* is a function $g : \mathcal{P} \to \mathcal{M}$ from preference profiles into matchings. Throughout, we restrict attention to direct mechanisms. For each $P \in \mathcal{P}$, let $g_i(P)$ denote the assignment of agent $i \in I$ under $g(P)$. Let $\mathcal{D}$ be the set of all direct mechanisms. Consider the following standard properties given any mechanism $g \in \mathcal{D}$. The mechanism g is individually rational, whenever it only leads to individually rational outcomes. If g produces only non-wasteful matchings then g is said to be non-wasteful. The mechanism g is *stable* (with respect to s) if it produces a stable matching with respect to s for each preference profile. A mechanism g is *(Pareto) efficient* if it only induces efficient matchings.

I proceed with two standard incentive notions. Formally, define a mechanism $g \in \mathcal{D}$ as *strategy-proof* if, for all $P \in \mathcal{P}$, there is no $i \in I$ and $P_i' \in \mathcal{P}_i$ such that $g_i(P_i', P_{-i}) P_i g_i(P)$. Denote with $\mathcal{SP} \subset \mathcal{D}$ the set of strategy-proof direct mechanisms. A mechanism $g \in \mathcal{D}$ is *group strategy-proof* if, for all $P \in \mathcal{P}$, there exists no $J \subseteq I$ and $P_J' \in \mathcal{P}_J$ such that $g_i(P_J', P_{-J})\ R_i\ g_i(P)$ for each $i \in J$, and $g_j(P_J', P_{-J})\ P_j\ g_j(P)$ for at least one $j \in J$.

2.2 A Transparency Framework

Consider a central matching authority that makes an *announcement* $g \in \mathcal{SP}$ to agents I. Given announcement g, the authority privately selects a mechanism $\tilde{g}$ from a given set of mechanisms $\mathcal{G} \subseteq \mathcal{D}$. Then, given any preference profile $P \in \mathcal{P}$, the induced final matching $\tilde{g}(P)$ is observable for all agents. Formally, given a preference profile $P \in \mathcal{P}$ and an agent $i \in I$, *observation* $o_i(\tilde{g}(P))$ consists of agent i's preference ranking P_i and the final matching $\tilde{g}(P)$. That is, for each i, preferences P_{-i} are not revealed to agent i. Refer to $\tilde{g}$ as a *deviation* from announcement g, if there exists a preference profile $P' \in \mathcal{P}$ such that $\tilde{g}(P') \neq g(P')$.

Next, I formalize when an agent can infer that the authority has deviated from its announcement. From now on, I assume that agents know I, X, s, and how g maps preference profiles into matching outcomes. Then, given preference profile $P \in \mathcal{P}$, we say that observation o_i has an *innocent explanation* for i, if there exists $P'_{-i} \in \mathcal{P}_{-i}$ such that $o_i = o_i(g(P_i, P'_{-i}))$ (Akbarpour and Li, 2020). Hence agent i's observation has an innocent explanation if i can make the same observation under announcement g. A deviation $\tilde{g} \in \mathcal{G}$ is *safe* if, for each $i \in I$ and each $P \in \mathcal{P}$, observation $o_i(\tilde{g}(P))$ has an innocent explanation for i (Akbarpour and Li, 2020). In words, a deviation is safe if each observation produced by the deviation has an innocent explanation for the agent who makes the observation. The main analysis will be based on the following criterion.

Definition 1. Announcement g is *transparent* if it has no safe deviations.

In other words, transparency requires that any deviation can be detected by at least one agent. Until Section 5, $\mathcal{G}$ is the set of direct mechanisms.

3 Stable Mechanisms

This section explores transparency of stable mechanisms. It is well known that any such mechanisms can be induced with the (agent-proposing) DA. Denote the DA mechanism that operates on s with DA^s.

As a preliminary work, I first show that stability can be verified by agents independently—a feature that does not apply for the efficiency criterion studied in Section 4. Fix a matching $\mu \in \mathcal{M}$ and a preference profile $P \in \mathcal{P}$. Given any agent $i \in I$, let $\Sigma^s(P_i) = \bigcup_{\tilde{P}_{-i} \in \mathcal{P}_{-i}} \Sigma^s(P_i, \tilde{P}_{-i})$. Thus, the following lemma is immediate.

Lemma 1. *For each* $P \in \mathcal{P}$, $\Sigma^s(P) = \bigcap_{i \in I} \Sigma^s(P_i)$.

We are ready for the first main result of this paper.

Theorem 1. *A deviation from* DA^s *is safe if and only if it is stable with respect to* s.

Proof. To prove the only if part of the statement, we rely on Lemma 1. Given announcement DA^s, consider an arbitrary deviation $\tilde{g}$ that is not stable with respect to s. To show that $\tilde{g}$ is not safe, take any $P \in \mathcal{P}$ for which $\tilde{g}(P) \notin \Sigma^s(P)$. By Lemma 1, there exists $i \in I$ such that $\tilde{g}(P)$ is not in $\Sigma^s(P_i)$. Now consider agent

i's observation $o_i(\tilde{g}(P))$. By Lemma 1 and the stability of DA^s with respect to s, for any $P'_{-i} \in \mathcal{P}_{-i}$, we have $DA^s(P_i, P'_{-i}) \in \Sigma^s(P_i)$. Thus, i cannot have an innocent explanation for $o_i(\tilde{g}(P))$. Hence, $\tilde{g}$ is not safe.

Moving to the if part of the statement, consider an arbitrary deviation $\tilde{g}$ from DA^s that is stable with respect to s. One has to show that $\tilde{g}$ is safe by constructing an innocent explanation for each agent and each of her observations under $\tilde{g}$. Take an arbitrary $i \in I$, an arbitrary $P \in \mathcal{P}$ and consider the associated observation $o_i(\tilde{g}(P))$ under the deviation. To show that $o_i(\tilde{g}(P))$ has an innocent explanation, consider the preference profile $P'_{-i} \in \mathcal{P}_{-i}$ where for each $j \neq i$, j's top choice on P'_j is $\tilde{g}_j(P)$. Since $\tilde{g}(P) \in \Sigma^s(P)$ also $\tilde{g}(P) \in \Sigma^s(P_i, P'_{-i})$. For each agent $j \neq i$, $\tilde{g}_j(P)$ is the top-choice and thus, since DA^s implements the agent-optimal stable matching, we must have $\tilde{g}(P) = DA^s(P_i, P'_{-i})$. Thus, $o_i(\tilde{g}(P))$ has an innocent explanation.

As i and P were chosen arbitrarily, each agent has an innocent explanation for each of her observations under $\tilde{g}$. Thus, $\tilde{g}$ is a safe deviation. Finally, since the choice of $\tilde{g}$ among the stable mechanisms was arbitrary, the proof is complete. □

Interestingly, Theorem 1 is useful to unveil further transparency features of DA^s (see Section 5). For instance, a direct consequence of Theorem 1 is that DA^s is transparent if and only if there exists a unique stable matching in $\Sigma^s(P)$ for each $P \in \mathcal{P}$. However, as shown in the following, this implies that DA^s is a *serial dictatorship* from Satterthwaite and Sonnenschein (1981) and Svensson (1994).[15] Clearly, in case of DA^s, this means that each object's priority scores must induce the same priority ranking over agents.

Proposition 1. *DA^s is transparent if and only if DA^s is a serial dictatorship.*

Proof. See Appendix A. □

However, Theorem 1 also implies that the authority's scope to deviate is substantially limited even if DA^s is no serial dictatorship.

Theorem (Lone Wolf Theorem (McVitie and Wilson, 1970))**.** For any given preference profile P, the set of assigned objects and agents is the same across all matchings in $\Sigma^s(P)$.

[15] A mechanism $g \in \mathcal{D}$ is a *serial dictatorship* if there exists a fixed ordering over agents, such that upon following the ordering, each agent is assigned to the most preferred object that is still available.

Thus, together with Theorem 1 we reach the following corollary.

Corollary 1. *Any safe deviation from DA^s assigns the same agents and the same objects as DA^s.*

In the remainder of this section, I provide a brief intuition on how the results generalize to a setup closer to the features of public school assignment (Balinski and Sönmez, 1999; Abdulkadiroğlu and Sönmez, 2003). Specifically, I assume that each agent's observation consists of her own assignment, there is no ex-ante information on other agents' scores and the objects' capacities are publicly known. In addition, the authority publicly discloses a score that is promised to correspond to the object's cutoff (score).[16,17] First, disclosing scores ensures that agents can verify that the possible underlying final matchings are not stable given the agents' individually known scores. Thus, any such deviation would be detected. Second, an object's disclosed score cannot be higher than its cutoff, since otherwise the agent with the lowest score assigned to this object has no innocent explanation. Finally, if an object's disclosed score is lower than its cutoff, then every agent has an innocent explanation if and only if no agent blocks any matching compatible with the disclosed scores. Thus, the arguments in the proof of Theorem 1 apply and as such the remaining results follow immediately.

4 Efficient Mechanisms

This section studies the transparency features of efficient mechanisms. I first illustrate that there exist inefficient safe deviations for some efficient announcements. As a motivation for such a deviation, consider an authority in the context of public school assignment that wants to satisfy certain distributional constraints (e.g., equal distribution of different genders, meeting some regional quota, or other socioeconomic considerations) that are at odds with efficiency.[18] Thus, a reasonable scenario might

[16]However the authority discloses a score for an object, if and only if, the object has filled its capacity. For a given matching, the cutoff (score) at an object is the lowest object-specific score among all agents assigned to this object.

[17]Alternatively, the results hold, if agents know the set of agents' feasible scores without personal identifiers and the authority promises to disclose cutoffs for all objects. The cutoff at an object is zero, if it does not fill its capacity. A reasonable example is a setting where agents' scores correspond to the agents' ranks in the objects' ranking.

[18]See, for instance, the work on matching under regional constraints (Kamada and Kojima, 2015), affirmative action (Abdulkadiroğlu and Sönmez, 2003); (Abdulkadiroğlu et al., 2005); (Kojima, 2012);

be that the authority initially advertises an efficient mechanism to boost participation, but then deviates in order to comply with its hidden distributional objectives. For illustrative purposes, I keep the size of the following example small.

Example 1. Let $I = \{i, j\}$ and $X = \{x, y\}$ and consider s such that i has highest priority for y and j has highest priority for x. The authority announces the TTC mechanism g that operates on priority structure s. This mechanism is known to be efficient and strategy-proof (Roth, 1982). Consider the following preferences:[19]

P_i	P_i'	P_j	P_j'
x	y	y	x
y	x	x	y
$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$

Given announcement TTC^s, the authority selects mechanism $\tilde{g} \in \mathcal{G}$, where

- $\tilde{g}(P) = \{(i, y), (j, x)\}$ and;
- $\tilde{g}(\tilde{P}) = TTC^s(\tilde{P})$, for all $\tilde{P} \in \mathcal{P} \setminus \{P\}$.

Note that $\tilde{g}$ is a deviation from TTC^s and that $\tilde{g}$ is not efficient, since $TTC^s(P) \neq \tilde{g}(P)$ and agents prefer to swap their assigned objects given P.

To show that $\tilde{g}$ is a safe deviation, consider the following innocent explanations: First, preferences P_j' provide an innocent explanation for $o_i(\tilde{g}(P)) = o_i(TTC^s(P_i, P_j'))$. Symmetrically, P_i' leads to an innocent explanation for $o_j(\tilde{g}(P)) = o_j(TTC^s(P_i', P_j))$. Finally, under any other preference profile the matchings under TTC^s and $\tilde{g}$ coincide. Thus, $\tilde{g}$ is safe and not efficient.[20] □

Next, I introduce *sequential dictatorship mechanisms*—a class of Pareto efficient and strategy-proof mechanisms known from Pápai (2001), Ehlers and Klaus (2003) and Pápai (2000) that is central for the main result of this section. For each $\tilde{X} \subseteq X \cup \{\emptyset\}$ let $\tilde{X}^C$ be the complement of $\tilde{X}$. For any $\tilde{X} \subseteq X \cup \{\emptyset\}$, $i \in I$ and $P_i \in \mathcal{P}_i$, let

Hafalir et al. (2013), matching under complex constraints (Westkamp, 2013), or diversity constraints (Ehlers et al., 2014).

[19]The preference relations in the table are read vertically. Thus, for example, P_i as stated means that i prefers x to y and y to the outside option.

[20]Note that $\tilde{g}$ is not strategy-proof. For instance, if agent i ranks only x acceptable, then i is assigned to y, whenever j reports P_j. Thus, in a setting where the authority deviates intentionally, i has an incentive to misreport her preferences under $\tilde{g}$. I address related questions in Section 5.

$$Top_i(P_i, \tilde{X}) = \{x \in \tilde{X}^C \cup \{\emptyset\} |\ \forall x' \in \tilde{X}^C \cup \{\emptyset\},\ xR_i x'\}$$

be agent $i's$ most preferred object in $\tilde{X}^C \cup \{\emptyset\}$. Let bijection $\pi : \{1, \ldots, |I|\} \to I$ be an ordering over agents I and collect in Π the set of all possible orderings on I. Given any $\pi \in \Pi$, let for each $m \in \{1, \ldots, |I|\}$, be $\pi(m)$ the m_{th}-*dictator* at π.

Definition 2. A mechanism $g \in \mathcal{D}$ is a *sequential dictatorship*, if there is a set of orderings $\Pi^g \subseteq \Pi$ such that the following conditions are satisfied:

(a) For each $P \in \mathcal{P}$, $\pi_P \in \Pi^g$ is an associated ordering such that

$$g_{\pi_P(1)}(P) = Top_{\pi_P(1)}(P_{\pi_P(1)}, \emptyset)$$

and for each $n \in \{2, \ldots, |I|\}$

$$g_{\pi_P(n)}(P) = Top_{\pi_P(n)}(P_{\pi_P(n)}, \cup_{l=1}^{n-1} g_{\pi_P(l)}(P)).$$

(b) Given each pair $P', \tilde{P} \in \mathcal{P}$,

(b_1) we have $\pi_{P'}(1) = \pi_{\tilde{P}}(1)$.

(b_2) if $m' < |I|$ is such that for each $n' \in \{1, \ldots, m'\}$, $\pi_{P'}(n') = \pi_{\tilde{P}}(n')$ and $g_{\pi_{P'}(n')}(P') = g_{\pi_{\tilde{P}}(n')}(\tilde{P})$, then $\pi_{P'}(m'+1) = \pi_{\tilde{P}}(m'+1)$.

In words, condition (a) recursively defines the matchings such that, for each preference profile, the m_{th}-dictator is assigned to her most preferred object still left after all previous dictators have been assigned. Condition (b_1) ensures that the first dictator is the same under each ordering and condition (b_2) requires that the identity of the next dictator only depends on the assignments of previous dictators.

We are ready for the first result of this section.

Theorem 2. *Take any efficient $g \in \mathcal{SP}$. Then, g is transparent if and only if it is a sequential dictatorship.*

Proof. See Appendix C. □

Intuitively, under a sequential dictatorship, at each step, at most one agent has the guarantee to select her favorite object among the remaining ones. Observing the

assignment of the first dictator reveals the identity of the second dictator, whose assignment then reveals the identity of the third dictator and so forth. If the authority deviates from some preference profile, then following the correct ordering of dictators, there must be a first agent who infers the stage she must have been the dictator while she is not assigned to her favorite choice of objects she should have been able to choose from. This agent cannot have an innocent explanation for her observation. Accordingly, the deviation is not safe.

The proof to reach necessity is divided into arguments for those candidates are that group strategy-proof and those that are strategy-proof but not group strategy-proof. Since the arguments for group strategy-proof candidates are central to the next result as well, I briefly explain the basic line of reasoning here. Concretely, consider again how I constructed the safe deviation in Example 1. One can essentially use the general idea of the construction for all efficient mechanisms which are group strategy-proof and not equivalent to a sequential dictatorship. I rely on a characterization by Pycia and Ünver (2017) saying that any efficient and group strategy-proof mechanism is equivalent to a *Trading Cycles (TC) Mechanism*. Each *TC* mechanism can be implemented via the *TC Algorithm*. Under each step of this algorithm, each unmatched object points to an unmatched agent and each unmatched agent points to an unmatched object. Once a cycle forms, agents in the cycle are assigned to the object they point to.[21] If a *TC* mechanism is not a sequential dictatorship, then at some step of the *TC* algorithm, two different agents are pointed by objects. Being pointed by an object is essentially a guarantee to not getting an object that is worse, than the one being pointed by. Thus, once reaching this step, the authority can then exploit agents' guarantees. In particular, consider the case where the two agents prefer each others' guaranteed objects most and the own guarantee is the second choice. Specifically, instead of honestly assigning agents to their top choices the authority assigns them to their second choices, whereas innocent explanations follow from other agents' possible preference for their own guarantees. I refer to Appendix B for the formal statement of the characterization by Pycia and Ünver (2017) along with a description of the *TC* algorithm.

The just outlined arguments provide an intuition for the proof of the following result.

[21]The idea of the *TC* algorithm builds on the idea of the *Top Trading Cycles (TTC) Algorithm*. However, pointing rules are more complex under *TC* compared to *TTC*.

Theorem 3. *If $g \in \mathcal{SP}$ is efficient and group strategy-proof, then the following three statements are equivalent:*

1. *g is transparent.*
2. *g is a sequential dictatorship.*
3. *g has only efficient safe deviations.*

Proof. See Appendix C. □

With very similar arguments, a characterization akin to Theorem 2 and Theorem 3 holds for the entire class of TC mechanisms in the many-to-one framework. Also, any efficient mechanism that is not group-strategy-proof is not transparent.[22] The example below illustrates that group strategy-proofness cannot be relaxed to strategy-proofness in the statement of Theorem 3.

Example 2. Let $I = \{i, j, k\}$ and $X = \{x, y\}$. Denote $\hat{\mathcal{P}} = \{\hat{P} \in \mathcal{P} \mid \hat{P}_i = P_i\}$, where $P_i : x, y, \emptyset$ and consider g such that:

1. Given any $P' \in \mathcal{P} \setminus \hat{\mathcal{P}}$, agents select their favorite objects among the remaining ones according to ordering i,j,k.
2. Given any $P \in \hat{\mathcal{P}}$, agents select their favorite objects among the remaining ones according to ordering i,k,j.

Clearly, g strategy-proof and efficient. To see that g is not group-strategy proof, consider a preference profile $\hat{P} \in \hat{\mathcal{P}}$ such that $\hat{P}_{i'} : x, y, \emptyset$ for all $i' \in I$, and a preference profile $P' \notin \hat{\mathcal{P}}$ such that $P'_i = x, \emptyset, y$ and where j and k have the same preferences as under P. Then, $g_i(\hat{P}) = g_i(P')$ with $\hat{P}_i \neq P'_i$, while $g_k(\hat{P}) = y$, $g_j(\hat{P}) = \emptyset$ and $g_j(P') = y$, $g_k(P') = \emptyset$. Thus, i is assigned to x for both preference profiles and j is strictly better of under preference profile P'. Since only the preferences for i have changed across $\hat{P}$ and P', g is not group strategy-proof.

Next, I show that there is no safe deviation from g that is inefficient. First, it is clear that no deviation $\tilde{g}$ is safe, if there exists $P \in \mathcal{P}$ such that $g_i(P) \neq \tilde{g}_i(P)$, since i must always get her top-choice. Second, for every profile $P \in \mathcal{P}$, under which

[22] TC mechanisms remain efficient and group strategy-proof in the many-one environment (Pycia and Ünver, 2011; Abdulkadiroğlu and Sönmez, 2003).

i does not rank x as her top-choice, similar arguments imply that j cannot have an innocent explanation for $o_j(\tilde{g}(P))$, whenever $g_j(P) \neq \tilde{g}_j(P)$. Therefore, $\tilde{g}$ is safe only if $\tilde{g}(P) = g(P)$, for all $P \in \mathcal{P} \setminus \hat{\mathcal{P}}$.

Hence, if $\tilde{g}$ is not efficient, then there must exist $P' \in \hat{\mathcal{P}}$ such that $\tilde{g}(P')$ is not efficient: Now, recall that $\tilde{g}$ can only be safe, if $\tilde{g}$ is non-wasteful and individually rational and if j and k would agree on their relative ranking of y and $\emptyset$, then any non-wasteful $\tilde{g}(P')$ is efficient. It is also clear that $\tilde{g}(P')$ cannot be individually rational if j and k both rank y unacceptable. Thus, in the remaining case exactly one of j and k must rank y unacceptable. However, this also means that for j or k, the matching $\tilde{g}(P')$ is not individually rational. □

As a remark, one can show that if $|X| < |I|$, similar arguments work for markets of any size and with mechanisms that are not dictatorial. By contrast, if $|X| \geq |I|$, then for different variants of TC mechanisms that are not group strategy-proof, inefficient safe deviations can be easily found by extending the key ideas sketched for the proof of the main results.

5 Strategic Agents

In this final section, the authority ex-ante commits to use a strategy-proof mechanism. This means that given announcement g, any deviation $\tilde{g}$ from g is strategy-proof itself (i.e the set of feasible deviations $\mathcal{G}$ is the set of strategy-proof mechanisms). From now on, all deviations by the authority are intentional and agents are strategic. The latter assumption is motivated by the idea that if all deviations are intentional, sophisticated agents might take the authority's incentives into account. Within this section, I focus on the transparency features of DA^s and TTC^s. Starting with DA^s, the following result is immediate.

Corollary 2. *DA^s is transparent.*

The result follows from Theorem 1 and the fact that DA^s is outcome equivalent to any mechanism that is strategy-proof and stable with respect to s.[23]

[23]Corollary 2 describes sufficient conditions for TTC^s to be transparent given the well-known equivalence between DA^s and TTC^s for Kesten-acyclic priority structures (Kesten, 2006). A priority structure is *Kesten-acyclic* (Kesten, 2006), if there exist no three agents $i, j, k \in I$ and no two objects $x, y \in X$ all distinct, such that: $s_i^x > s_k^x > s_j^x$ and $s_j^y > s_i^y, s_k^y$. This also holds in the many-to-one framework of Abdulkadiroğlu and Sönmez (2003) by adding the scarcity condition of Kesten (2006).

Next, I provide necessary conditions for TTC^s to be transparent. We need the following definition.

Definition 3. A *full replacement cycle* consists of four agents $i, j, k, l \in I$ and two objects $x, y \in X$ all distinct, such that:

(1) $s_i^x > s_k^x > s_j^x, s_l^x$ and $s_j^y > s_l^y > s_i^y, s_k^y$, or

(2) $s_i^x > s_k^x, s_l^x > s_j^x$ and $s_j^y > s_l^y, s_k^y > s_i^y$.

As stated below, a full replacement cycle in s means that TTC^s has a strategy-proof safe deviation.

Proposition 2. *If s has a full replacement cycle, then TTC^s is not transparent.*

Proof. See Appendix D. □

Basically, if s contains a full replacement cycle as described in Definition 3, then there are situations under the TTC algorithm, where a trading cycle forms between i and j for the objects x,y and the authority can deviate to a strategy-proof $\tilde{g}$ as follows: Under deviation $\tilde{g}$ from TTC^s there are preference profiles, where i is assigned to x and j to y although they would prefer to swap their assignments. The full replacement cycle ensures that k and l have scores at x and y high enough to replace i and j at x and y, in case that i ranks only y and j ranks only x acceptable. By contrast, if no such full replacement cycle would exists, i and j could force the authority to assign i to y and j to x as done under TTC^s. Otherwise, $\tilde{g}$ would be wasteful and hence not safe. Accordingly, deviation $\tilde{g}$ is not strategy-proof for i or j.

Next, the condition introduced below ensures that there are no full replacement cycles in s.

Definition 4. Priority structure s has the *imperfect replacement property* if, for any three agents $i, j, k \in I$, there exist no two objects $x, y \in X$, all distinct such that:

(1) $s_i^x > s_k^x, s_l^x > s_j^x$,

(2) $s_l^y > s_i^y, s_j^y, s_k^y$ or $s_i^y, s_j^y, s_k^y > s_l^y$.

The condition is weaker than some acyclicity notions that characterize TTC^s with regard to various desirable properties (Kesten, 2006; Ergin, 2002). Furthermore, if the imperfect replacement property is satisfied, no safe deviation from TTC^s is strategy-proof.

Proposition 3. *If s has the imperfect replacement property, then TTC^s is transparent.*

Proof. See Appendix D. □

Interestingly, the domain of priority structures with the imperfect replacement property is richer than the domains for other acyclicity notions on a natural dimension. In particular, the acyclicity conditions of Kesten (2006) and Mandal and Roy (2022) are not satisfied, if TTC^s allows for top-trading-cycles that contain strictly more than two agents.[24] This is not true for the imperfect replacement property, since Proposition 3 implies that it is satisfied for all markets with three agents and three objects.

Appendix A Proof of Proposition 1

($\Leftarrow$) If DA^s is a serial dictatorship then for given any pair of agents i, j and objects x,y, we have $s_i^x > s_j^x$ if and only if $s_i^y > s_j^y$. Given any $P \in \mathcal{P}$, following the ordering of the induced score ranking for some $x \in X$, for each $n \in \{1, \ldots, |I|\}$, the n_{th}-ranked agent is guaranteed her top choice among the remaining objects after all previous agents in line have left.[25] Hence for each $P \in \mathcal{P}$ it is clear that $\Sigma^s(P)$ is a singleton. Therefore, Theorem 1 implies that there exists no safe deviation from DA^s and thus DA^s is transparent.

($\Rightarrow$) Suppose that DA^s is not a serial dictatorship. By definition, this means that there exist two agents $i, j \in I$ and two objects x,y such that $s_i^x > s_j^x$ and $s_j^y > s_i^y$.

Denote $I' = I \setminus \{i, j\}$ and let preference profile $P_{I'} \in \mathcal{P}_{I'}$ be such that for each $k \in I'$, P_k ranks $\emptyset$ as the top choice and the ranking below $\emptyset$ is specified arbitrarily. Consider the following preferences for agents i and j.

Let $P_i, P_i' \in \mathcal{P}_i$ be described by

- xP_iy and for all $x' \in X \cup \{\emptyset\} \setminus \{x, y\}$: yP_ix' and

[24] A priority structure is *strongly-acyclic* (Mandal and Roy, 2022), if there exist no three agents $i, j, k \in I$ and three objects $x, y, z \in X$ all distinct, such that $s_i^x > s_k^x, s_j^x$ as well as $s_j^y > s_i^y, s_k^y$ and $s_k^z > s_i^z, s_j^y$. Strong acyclicty characterizes priority structures for which TTC^s is *obviously strategy-proof* (Li, 2017; Mandal and Roy, 2022; Troyan, 2019). Yet a violation of the imperfect replacement property is neither weaker nor stronger than the acyclicity conditions from Mandal and Roy (2022).

[25] The first ranked agent must receive her top choice under any stable matching in $\Sigma^s(P)$. Next, the second ranked agent receives, under any stable matching in $\Sigma^s(P)$, her top choice among objects once the first agent left, and so forth.

- $yP_i'x$ and for all $x' \in X \cup \{\emptyset\} \setminus \{x, y\}$: $xP_i'x'$.

Similarly, for agent j let the preferences $P_j, P_j' \in \mathcal{P}_j$ be

- yP_jx and for all $x' \in X \cup \{\emptyset\} \setminus \{x, y\}$: xP_jx' and
- $xP_j'y$ and for all $x' \in X \cup \{\emptyset\} \setminus \{x, y\}$: $yP_j'x'$.

Next, I construct a safe deviation $\tilde{g} \in \mathcal{G}$ from DA^s. For profile $P = (P_i, P_j, P_{I'})$ suppose that $\tilde{g}(P)$ yields $\tilde{g}_i(P) = y$, $\tilde{g}_j(P) = x$, and for all $k \in I'$, $\tilde{g}_k(P) = \emptyset$. Now consider $\tilde{g}$ such that

$$\forall P' \in \mathcal{P} \setminus \{P\} : \tilde{g}(P') = DA^s(P').$$

It is simple to check that the DA algorithm yields $DA_i^s(P) = x$, $DA_j^s(P) = y$, and for all $k \in I'$, $DA_k^s(P) = \emptyset$. Thus, $\tilde{g}$ is a deviation.

It remains to show that $\tilde{g}$ is safe. Thus, each observation possibly made under the deviation $\tilde{g}$ must have an innocent explanation for the observing agent. Except for preference profile P, any observation has an innocent explanation for the respective agent, since observations produced by the deviation are identical to those under the announcement DA^s.

To complete the proof, we need for each $i' \in I$ an innocent explanation for her observation $o_{i'}(\tilde{g}(P))$. It is easily checked that one reaches

$$DA^s(P_i', P_j, P_{I'}) = DA^s(P_i, P_j', P_{I'}) = \tilde{g}(P).$$

from which one can see that for each agent $i' \in I$, the observation $o_{i'}(\tilde{g}(P))$ has an innocent explanation. Hence $\tilde{g}$ is a safe deviation and DA^s is not transparent.

Appendix B Trading Cycles and Characterizations of Group Strategy-Proofness

In this section, I introduce *Trading Cycles (TC) Mechanisms* (Pycia and Ünver, 2017) and Pycia and Ünver (2014) together with the main characterization of group strategy-proof and Pareto efficient mechanisms.[26] I provide an additional characterization

[26] I augment the description of Pycia and Ünver (2017) to the setting with outside options as described in (Pycia and Ünver, 2017, Supplement, p.5) and Pycia and Ünver (2014). The characterization of group strategy-proof and Pareto efficient mechanisms presented at the end of this

of group strategy-proof mechanisms by Pápai (2000) that has been extended to the setting with outside options by Pycia and Ünver (2014) that will be useful for the proofs of Theorem 2 and Theorem 3.

Starting with some necessary terminology, a submatching for $J \subseteq I$ is a matching $\sigma : J \to X \cup \{\emptyset\}$ restricted to agents J. The set of possible submatchings is $\mathcal{S}$ and let $\hat{\mathcal{M}} \equiv \mathcal{S} \setminus \mathcal{M}$. Denote with σ_I the set of agents assigned under submatching $\sigma \in \mathcal{S}$ and with σ_X the set of objects from X that are matched under submatching σ. Moreover, let $\hat{I}_\sigma \equiv I \setminus \sigma_I$ and let $\hat{X}_\sigma \equiv X \setminus \sigma_X$ be the set of unmatched agents and objects from X under σ, respectively. Note that an agent does not belong to the set of unmatched agents if she is assigned to the outside option. Denote the empty submatching with $\sigma_\emptyset$. The set of submatchings is ordered if one associates each submatching with its graph: for any $\sigma, \sigma' \in \mathcal{S}$, $\sigma \subset \sigma'$ if and only if each agent-object pair matched under σ is also matched under σ'.

The TC algorithm operates on a well-defined control right structure on the set of submatchings, which is defined as follows.

Definition 5. A *structure of control rights* is a collection of mappings

$$(c, b) \equiv \{(c_\sigma, b_\sigma) : \hat{X}_\sigma \to \hat{I}_\sigma \times \{owner, broker\}\}_{\sigma \in \hat{\mathcal{M}}}$$

That is, for a given submatching σ and an unmatched object x, the mapping c_σ appoints the unmatched agent $c_\sigma(x)$ as the unique controller of x. The type of control is determined by b_σ. The agent $c_\sigma(x)$ owns x at σ if $b_\sigma(x) = owner$ and $c_\sigma(x)$ brokers x at σ if $b_\sigma(x) = broker$. In the former case, call an agent an *owner* of x and in the latter case call an agent a *broker* of x. Refer to x as the *owned object* or *brokered object*, respectively. Note that the outside option is neither owned nor brokered.

The control right structure has to satisfy several consistency conditions to ensure that the induced mechanism is group strategy-proof and efficient. I will discuss some of these conditions when explicitly needed in the proof of Theorem 3. The interested reader is kindly referred to an excellent discussion and interpretation of these conditions in Pycia and Ünver (2017) and Pycia and Ünver (2014). The version depicted below is from Pycia and Ünver (2014).

section extends to the setting with outside options according to (Pycia and Ünver, 2017, Supplement, p.6) and Pycia and Ünver (2014).

Definition 6. A control right structure (c, b) is *consistent* if each of the following conditions is satisfied. For any $\sigma \in \hat{\mathcal{M}}$

$(C1)$ there is at most one brokered object at σ.

$(C2)$ if i is the only unmatched agent at σ, then i owns all unmatched objects at σ.

$(C3)$ if agent i brokers an object at σ, then i does not control any other object at σ.

For any two submatchings $\sigma, \sigma' \in \hat{\mathcal{M}}$ such that $|\sigma'| = |\sigma| + 1$ and $\sigma \subset \sigma'$ with an agent $i \in \hat{I}_{\sigma'}$ who controls an object $x \in \hat{X}_{\sigma'}$ at σ it holds that

$(C4)$ If i owns x at σ, then i owns x at σ'.

$(C5)$ Assume that at least two agents from $\hat{I}_{\sigma'}$ own objects at σ. If i' brokers object x' at σ, then i brokers x' at σ'

$(C6)$ If agent $i' \in \hat{I}_{\sigma'}$ controls $x' \in \hat{X}_{\sigma'}$ at σ, then i' owns x at $\sigma \cup \{(i, x')\}$ and if i' brokers x' at σ but not at σ', then i owns x' at σ'.

Let the domain of consistent control right structures be $\mathcal{C}$ and in the following take any $(c, b) \in \mathcal{C}$. I now describe the TC algorithm operating on (c, b), where $TC^{(c,b)}$ denotes the induced TC mechanism.[27]

The TC algorithm For any $P \in \mathcal{P}$, one calculates $TC^{(c,b)}(P)$ as follows: There is a finite sequence of steps $t = 1, 2, ...$ Denote with σ^{t-1} the submatching of agents and objects matched before step t. Prior to the first step, the submatching is empty, i.e. $\sigma^0 = \emptyset$. The algorithm terminates with σ^{t-1} if each agent is matched to an object, that is, if $\sigma^{t-1} \in \mathcal{M}$. If $\sigma^{t-1} \in \hat{\mathcal{M}}$, then the algorithm proceeds with the following substeps in Step t:

Step t(a): Pointing Let each object $x \in \hat{X}_{\sigma^{t-1}}$ point to its controller $c_{\sigma^{t-1}}(x)$. If there is a broker in $\hat{I}_{\sigma^{t-1}}$ for whom the brokered object is the only acceptable object, let the broker point to the outside option. Otherwise, let the broker point to her most preferred object among all objects that are owned. Each agent $i \in \hat{I}_{\sigma^{t-1}}$ that is not a broker points to her top choice x among objects $\hat{X}_{\sigma^{t-1}} \cup \{\emptyset\}$.

[27] To keep the notation of a common outside option throughout the paper, the description of the algorithm is slightly modified compared to Pycia and Ünver (2014). However, a quick glance reveals that the descriptions are equivalent.

Step t(b): Trading Cycles Given $n \in \mathbb{N}$, there is a *cycle* at σ^{t-1}

$$x^1 \to i^1 \to \dots x^n \to i^n \to x^1$$

in which agents $i^l \in \hat{I}_{\sigma^{t-1}}$ point to $x^{l+1} \in \hat{X}_{\sigma^{t-1}}$, and objects x^l point to agents i^l (here $l = 1, ..., n$ and superscripts are added modulo n).

Step t(c): Matching Collect all cycles which do not contain a broker and match each agent in a cycle to the object she points to. Match agents in a cycle with a broker if and only if there is at least one owner who points to the brokered object. Assign each owner who points to the outside option to the outside option. Let σ^t be the union of σ^{t-1}, the set of just assigned agent-object pairs and assigned owner-outside option pairs.

No pair of cycles intersect, there is at least one pair matched at each step and the number of steps is thus finite. Given a consistent control-right structure (c, b), we define a submatching $\sigma \in \mathcal{S}$ as *on-path* on $TC^{(c,b)}$, if there exists a preference profile $P \in \mathcal{P}$ such that there exists some step $t \in \mathbb{N}$, where $\sigma = \sigma_{t-1}$ while running the TC algorithm with input P on control-rights structure (c, b).

The proof of Theorem 3 presented in Appendix C builds on the following result.

Theorem (Pycia and Ünver (2017, 2014))**.** A mechanism $g \in \mathcal{D}$ is group strategy-proof and Pareto efficient if and only if it is equivalent to a TC mechanism $TC^{(c,b)}$ with some consistent control right structure $(c, b) \in \mathcal{C}$.

Finally, $TC^{(c,b)}$ satisfies the following property of *non-bossiness* as defined by Satterthwaite and Sonnenschein (1981).A mechanism $g \in \mathcal{D}$ is *non-bossy* if for all $P \in \mathcal{P}$, there is no $i \in I$, and $P_i' \in \mathcal{P}_i$, such that $g_i(P) = g_i(P_i', P_{-i})$, but $g(P) \neq g(P_i', P_{-i})$. More specifically, as known from Pápai (2000), the domain of group strategy-proof mechanisms is characterized through the collection of strategy-proof and *non-bossy* mechanisms.

Lemma 2 (Pápai (2000); Pycia and Ünver (2014))**.** *A mechanism is group strategy-proof if and only if it is strategy-proof and non-bossy.*

Appendix C Proofs of Theorem 2 and Theorem 3

This section contains all results needed to obtain Theorem 2 and Theorem 3. Specifically, Lemma 3 presented first, implies the sufficiency parts of the statements. Necessity for Theorem 2 follows from applying Lemma 4 and Lemma 5, whereas the necessity parts of Theorem 3 only require Lemma 4.

Lemma 3. *If announcement $g \in \mathcal{SP}$ is a sequential dictatorship, then g is transparent.*

Proof. Suppose that announcement g is a sequential dictatorship and let $\tilde{g} \in \mathcal{G}$ be an arbitrary deviation from g. I aim to show that there exists at least one agent who has no innocent explanation for one of her observations she makes under deviation $\tilde{g}$.

To start, by definition of a deviation, there must exist a preference profile $P \in \mathcal{P}$ such that $\tilde{g}(P) \neq g(P)$. Let $I' = \{i' \in I |\ g_{i'}(P) \neq \tilde{g}_{i'}(P)\}$. Next, select $i \in I'$ such that, for all $i' \in I' \setminus \{i\}$, we have $\pi_P^{-1}(i) < \pi_P^{-1}(i')$. Thus, since for all $k \in I$ with $\pi_P^{-1}(k) < \pi_P^{-1}(i)$, $\tilde{g}_k(P) = g_k(P)$ and since Definition 2 (a) implies that $g_i(P) = Top_i(P_i, \cup_{l=1}^{\pi_P^{-1}(i)-1} g_{\pi_P(l)}(P))$, we have $g_i(P) P_i \tilde{g}_i(P)$.

I now show that agent i has no innocent explanation for her observation $o_i(\tilde{g}(P))$. Note that Definition 2 implies that $\pi_P^{-1}(i) = \pi^{-1}_{(P_i, \tilde{P}_{-i})}(i)$, if $\tilde{P}_{-i} \in \mathcal{P}_{-i}$ is such that $g_k(P) = g_k(P_i, \tilde{P}_{-i})$ for all $k \in I$ with $\pi_P^{-1}(k) < \pi_P^{-1}(i)$. Thus, since for all $k \in I$ with $\pi_P^{-1}(k) < \pi_P^{-1}(i)$, we have $\tilde{g}_k(P) = g_k(P)$, we obtain $g_i(P) = g_i(P_i, \tilde{P}_{-i})$. However, since $g_i(P) \neq \tilde{g}_i(P)$, the previous arguments then imply that agent i cannot have an innocent explanation for $o_i(\tilde{g}(P))$. We thus conclude that $\tilde{g}$ is not safe and therefore g is transparent. □

I now turn to two additional lemmas for the necessity parts of Theorem 2 and Theorem 3. The next lemma shows that a non-efficient safe deviation exists for an efficient and group strategy-proof mechanism which is no sequential dictatorship.

Lemma 4. *Let announcement $g \in \mathcal{SP}$ be efficient and group-strategy-proof. If g is not a sequential dictatorship, then there exists a safe deviation from g, which is not efficient.*

Proof. Let the authority announce a group strategy-proof mechanism g which is no sequential dictatorship. Thus, g is equivalent to a TC mechanism with some consistent control right structure.

We first derive an equivalent definition of sequential dictatorships (Definition 2 in terms of TC mechanisms. First, as has been shown by Pycia and Ünver (2017) (Theorem 6) and Pycia and Ünver (2014) (Proposition 3), given any consistent (c', b') and any submatching $\sigma \in \hat{\mathcal{M}}$, if there is a single agent who owns all objects in $\hat{X}_\sigma$, then there is no broker at σ. Second, given any consistent control right structure (c', b'), if there is a single owner at each on-path submatching on $TC^{(c',b')}$, then as can be easily shown $TC^{(c',b')}$ is equivalent to a sequential dictatorship according to Definition 2.

Conversely, since g is not a sequential dictatorship, there is a submatching $\sigma^* \in \hat{\mathcal{M}}$ such that given any consistent (c, b) for which $TC^{(c,b)}$ is equivalent to g, σ^* is on-path on $TC^{(c,b)}$. Moreover, there exist two agents i and j such that given any such (c, b), both agents i, j each own at least one object at σ^*.

Now fix an arbitrary consistent (c, b) such that $TC^{(c,b)}$ is equivalent to g. To prove the result, it is sufficient to show that there exists a non-efficient safe deviation from $TC^{(c,b)}$. We construct the deviation as follows: Since σ^* is on-path and $TC^{(c,b)}$ non-bossy, we can select $P_{\sigma_I^*}$ such that for each $k \in \sigma_I^*$, $\sigma^*(k)$ is k's top choice under P_k, so that under any profile $(P_{\sigma_I^*}, \tilde{P}_{\hat{I}_{\sigma^*}}) \in \mathcal{P}$, where $\tilde{P}_{\hat{I}_{\sigma^*}}$ is chosen arbitrarily, the TC algorithm arrives at submatching $\sigma^{t^*-1} = \sigma^*$ in some Step t^*.

Now consider Step t^* and the two agents $i, j \in \hat{I}_{\sigma^*}$ who we know are both owners at σ^*. Let i own object $x \in \hat{X}_{\sigma^*}$ and let j own object $y \in \hat{X}_{\sigma^*}$. The following preferences of agent i and j will be central. Let $P_i, P_i' \in \mathcal{P}_i$ be described by

- yP_ix and for all $x' \in X \cup \{\emptyset\} \setminus \{x, y\}$: xP_ix' and
- $xP_i'y$ and for all $x' \in X \cup \{\emptyset\} \setminus \{x, y\}$: $yP_i'x'$.

Similarly, for agent j let the preferences $P_j, P_j' \in \mathcal{P}_j$ be

- xP_jy and for all $x' \in X \cup \{\emptyset\} \setminus \{x, y\}$: yP_jx' and
- $yP_j'x$ and for all $x' \in X \cup \{\emptyset\} \setminus \{x, y\}$: $xP_j'x'$.

Denote $K = I \setminus \{\sigma_I^* \cup \{i, j\}\}$ and let $P_K \in \mathcal{P}_K$ be specified arbitrarily.

I now construct the candidate deviation $\tilde{g} \in \mathcal{G}$ as follows:

- For all $\tilde{P} \in \mathcal{P} \setminus \{(P_{\sigma_I^*}, P_i, P_j, P_K)\}$, suppose that $\tilde{g}(\tilde{P}) = TC^{(c,b)}(\tilde{P})$ and,
- let $\tilde{g}(P_{\sigma_I^*}, P_i, P_j, P_K) = TC^{(c,b)}(P_{\sigma_I^*}, P_i', P_j', P_K)$.

I first establish that $\tilde{g}$ is indeed a non-efficient deviation from $TC^{(c,b)}$. As argued before, under any profile where each $k \in \sigma_I^*$ reports P_k, we eventually arrive at submatching $\sigma^{t^*-1} = \sigma^*$ in Step t^*. Hence, for all $k \in \sigma_I^*$,

$$\tilde{g}_k(P_{\sigma_I^*}, P_i, P_j, P_K) = TC_k^{(c,b)}(P_{\sigma_I^*}, P_i', P_j', P_K) = TC_k^{(c,b)}(P_{\sigma_I^*}, P_i, P_j, P_K).$$

Next, under preference profile $(P_{\sigma_I^*}, P_i, P_j, P_K)$ at Step t^*, there is a cycle consisting only of owners, namely

$$x \to i \to y \to j \to x,$$

and as such, we must have that

$$TC_i^{(c,b)}(P_{\sigma_I^*}, P_i, P_j, P_K) = y,$$
$$TC_j^{(c,b)}(P_{\sigma_I^*}, P_i, P_j, P_K) = x.$$

However, if agents report $(P_{\sigma_I^*}, P_i', P_j', P_K)$, then there are two cycles only of owners at Step t^*, namely:

$$x \to i \to x, \quad y \to j \to y,$$

and thus,

$$TC_i^{(c,b)}(P_{\sigma_I^*}, P_i', P_j', P_K) = x,$$
$$TC_j^{(c,b)}(P_{\sigma_I^*}, P_i', P_j', P_K) = y,$$

which implies that

$$TC^{(c,b)}(P_{\sigma_I^*}, P_i, P_j, P_K) \neq TC^{(c,b)}(P_{\sigma_I^*}, P_i', P_j', P_K).$$

Hence, $\tilde{g}$ is a deviation from $TC^{(c,b)}$ and $\tilde{g}$ is not efficient since agent i and j would both prefer to swap their assignments.

It remains to be shown that $\tilde{g}$ is safe. First, it is clear that for all preference profiles $\tilde{P} \in \mathcal{P} \setminus (P_{\sigma_I^*}, P_i, P_j, P_K)$, since $\tilde{g}(\tilde{P}) = TC^{(c,b)}(\tilde{P})$, innocent explanations for observations are immediate. Second, for each $i' \in I$, we need an innocent explanation for observation $o_{i'}(\tilde{g}(P_{\sigma_I^*}, P_i', P_j', P_K))$. Again, innocent explanations are immediate

for each $k \in \sigma_I^* \cup K$, since

$$\tilde{g}_k(P_{\sigma_I^*}, P_i, P_j, P_K) = TC_k^{(c,b)}(P_{\sigma_I^*}, P'_i, P'_j, P_K).$$

Note that this holds irrespective of whether agents in K have been affected by the deviation or not.

I proceed with considering agents i and j and the pair of candidate profiles $(P_{\sigma_I^*}, P'_i, P_j, P_K)$ and $(P_{\sigma_I^*}, P_i, P'_j, P_K)$. I aim to show that

$$o_i(TC^{(c,b)}(P_{\sigma_I^*}, P_i, P'_j, P_K)) = o_i(\tilde{g}(P_{\sigma_I^*}, P_i, P_j, P_K)), \tag{1}$$

$$o_j(TC^{(c,b)}(P_{\sigma_I^*}, P'_i, P_j, P_K)) = o_j(\tilde{g}(P_{\sigma_I^*}, P_i, P_j, P_K)). \tag{2}$$

We already know that for each $k \in \sigma_I^*$ the assignment is identical to the one under the deviation and that under both candidate profiles above we have to arrive at submatching σ^* at Step t^*. Now consider Step t^* under candidate profile $(P_{\sigma_I^*}, P_i, P'_j, P_K)$, where cycle

$$y \to j \to y$$

exists and hence j must be assigned to y. This implies that i is assigned to x, since i owns x at $\sigma^* = \sigma^{t^*-1}$ and it is her favorite choice among the remaining objects according to P_i.[28] Symmetrically, at Step t^* with candidate profile $(P_{\sigma_I^*}, P'_i, P_j, P_K)$, there is a cycle

$$x \to i \to x.$$

This implies that j is assigned to y, since j owns y at $\sigma^* = \sigma^{t^*-1}$ and it is her favorite remaining choice according to P_j. Thus, for both $i' \in \{i, j\}$, we have

$$TC_{i'}^{(c,b)}(P_{\sigma_I^*}, P'_i, P'_j, P_K) = TC_{i'}^{(c,b)}(P_{\sigma_I^*}, P'_i, P_j, P_K) = TC_{i'}^{(c,b)}(P_{\sigma_I^*}, P_i, P'_j, P_K)$$

Using non-bossiness of $TC^{(c,b)}$ (See Lemma 2), it must be true that, for all $k \in K$, we have that

$$TC_k^{(c,b)}(P_{\sigma_I^*}, P'_i, P'_j, P_K) = TC_k^{(c,b)}(P_{\sigma_I^*}, P'_i, P_j, P_K) = TC_k^{(c,b)}(P_{\sigma_I^*}, P_i, P'_j, P_K)$$

[28] Note that ownership rights persist according to Condition $(C4)$ of a consistent control rights structure, as long as the owner is not yet assigned to a different object.

and as such conditions (1) and (2) are satisfied and each agent has an innocent explanation for any of her observations under $\tilde{g}$. The choice of $TC^{(c,b)}$ was arbitrary among TC mechanisms that are equivalent to g. Finally, since g is equivalent to $TC^{(c,b)}$, the same conclusion holds for announcement g. We conclude that $\tilde{g}$ is a safe deviation from g and hence that g does not allow to commit to efficiency. This completes the proof of the lemma. □

To complete the argument, I next establish that each efficient mechanism allows safe deviations if it is strategy-proof but not group strategy-proof.

Lemma 5. *Let announcement $g \in \mathcal{SP}$ be efficient and not group strategy-proof. Then, there exists a safe deviation from g.*

Proof. Since g is not group strategy-proof but strategy-proof, g is bossy by Lemma 2. If g is bossy then, by definition, there exists an agent $i \in I$ with $P_i, P_i' \in \mathcal{P}_i$ and $P_{-i} \in \mathcal{P}_{-i}$ such that $g(P_i, P_{-i}) \neq g(P_i', P_{-i})$ and $g_i(P_i, P_{-i}) = g_i(P_i', P_{-i})$. Second, since g is strategy-proof, for any $P_i^* \in \mathcal{P}_i$, where $g_i(P_i, P_{-i})$ is ranked as $i's$ top choice, it must hold

$$g_i(P_i^*, P_{-i}) = g_i(P_i, P_{-i}) = g_i(P_i', P_{-i}).$$

Thus, since $g(P_i, P_{-i}) \neq g(P_i', P_{-i})$, it is either true that $g(P_i^*, P_{-i}) \neq g(P_i, P_{-i})$ or $g(P_i^*, P_{-i}) \neq g(P_i', P_{-i})$ or both. Assume in the following that $g(P_i^*, P_{-i}) \neq g(P_i, P_{-i})$ (a symmetric argument will apply for the case, where $g(P_i^*, P_{-i}) \neq g(P_i', P_{-i})$ and not $g(P_i^*, P_{-i}) = g(P_i, P_{-i})$).

Next, consider a deviation $\tilde{g}$ with $\tilde{g}(P_i^*, P_{-i}) = g(P_i, P_{-i})$ and $\tilde{g}(\tilde{P}) = g(\tilde{P})$ for all $\tilde{P} \in \mathcal{P} \setminus \{(P_i^*, P_{-i})\}$. Since all observations under g and $\tilde{g}$ coincide except under preference profile (P_i^*, P_{-i}), in order to obtain that $\tilde{g}$ is safe, it remains to show that each agent $k \in I$ has an innocent explanation for her observation $o_k(\tilde{g}(P_i^*, P_{-i}))$.

First, note that for each $j \neq i$, the preference profile of other agents P_{-j} provides an innocent explanation for observation $o_j(\tilde{g}(P_i^*, P_{-i}))$. Second, for agent i consider preference profile P_{-i}^* such that for each agent $j \neq i$, P_j^* ranks $g_j(P_i, P_{-i})$ as the top choice. Now note that under preference profile (P_i^*, P_{-i}^*), the unique Pareto efficient matching is $g(P_i, P_{-i})$ and since g is Pareto efficient, we thus must have $g(P_i^*, P_{-i}^*) = g(P_i, P_{-i})$. Thus, P_{-i}^* provides an innocent explanation for $o_i(\tilde{g}(P_i^*, P_{-i}))$. Hence $\tilde{g}$ is a safe deviation from g. □

Appendix D Proof of Proposition 2 and Proposition 3

Proof of Proposition 2. In the following, denote $g = TTC^s$. Suppose that s has a full replacement cycle. That is, s either satisfies condition (1) or condition (2) of the definition of a full replacement cycle. Let s satisfy condition (1). The same arguments will apply if s satisfies condition (2). That is, there exist four agents $i, j, k, l \in I$ and two objects $x, y \in X$ all distinct, such that:

(1') $s_i^x > s_k^x > s_j^x, s_l^x$,

(2') $s_j^y > s_l^y > s_i^y, s_k^y$.

Now consider the following preferences that will be central for the construction of the strategy-proof safe deviation $\tilde{g}$ from g. Denote $I' = I \setminus \{i, j, k, l\}$ and let preference profile $P_{I'} \in \mathcal{P}_{I'}$ be such that for each $m \in I'$, $P_m : \emptyset, \dots$ where the ranking below $\emptyset$ is specified arbitrarily. Consider the following preferences for agents in I':

- Consider the set of preferences $\tilde{\mathcal{P}}_i \subseteq \mathcal{P}_i$ such that, for each $P_i \in \tilde{\mathcal{P}}_i$, we have yP_ix' for all $x' \in X \cup \{\emptyset\} \setminus \{y\}$. Also, for each $P_i \in \tilde{\mathcal{P}}_i$, let $\bar{P}_i$ rank y last and all other objects in the same order as under P_i.
- Similarly, for j, let the set of preferences $\tilde{\mathcal{P}}_j \subseteq \mathcal{P}_j$ be such that, for each $P_j \in \tilde{\mathcal{P}}_j$, we have xP_jx' for all $x' \in X \cup \{\emptyset\} \setminus \{x\}$. For each $P_j \in \tilde{\mathcal{P}}_j$, the preferences $\bar{P}_j$ rank x last and all other objects in the same order as under P_j.
- For agent k, consider the set of preferences $\tilde{\mathcal{P}}_k \subseteq \mathcal{P}_k$ such that, for each $P_k \in \tilde{\mathcal{P}}_k$, we have yP_kx' and xP_kx' for all $x' \in X \cup \{\emptyset\} \setminus \{x, y\}$.
- For agent l, consider the set of preferences $\tilde{\mathcal{P}}_l \subseteq \mathcal{P}_l$ such that, for each $P_l \in \tilde{\mathcal{P}}_l$, we have yP_lx' and xP_lx' for all $x' \in X \cup \{\emptyset\} \setminus \{x, y\}$.

Next, I construct a safe deviation $\tilde{g} \in \mathcal{SP}$ from g. For any profile $(P_i, P_j, P_k, P_l, P_{I'})$ with $P_i \in \tilde{\mathcal{P}}_i$, $P_j \in \tilde{\mathcal{P}}_j$ $P_k \in \tilde{\mathcal{P}}_k$ and $P_l \in \tilde{\mathcal{P}}_l$, let $\tilde{g}(P_i, P_j, P_k, P_l, P_{I'}) = g(\bar{P}_i, \bar{P}_j, P_k, P_l, P_{I'})$. For all other preference profiles $\hat{P} \in \mathcal{P}$, let $\tilde{g}(\hat{P}) = g(\hat{P})$ and hence for these preference profiles each agent has an innocent explanation for her respective observation under $\tilde{g}$.

To see that $\tilde{g}$ is a safe deviation, take any four preference rankings for agents i, j, k, l, where $P_i \in \tilde{\mathcal{P}}_i$, $P_j \in \tilde{\mathcal{P}}_j$, $P_k \in \tilde{\mathcal{P}}_k$, $P_l \in \tilde{\mathcal{P}}_l$ and note that $g(P_i, P_j, P_k, P_l, P_{I'}) \neq g(\bar{P}_i, \bar{P}_j, P_k, P_l, P_{I'})$.

Thus, $\tilde{g}$ is a deviation. To verify that $\tilde{g}$ is safe, note that for all $i' \notin I \setminus \{i, j\}$, $(\bar{P}_i, \bar{P}_j, P_{-\{i,j\}})$ is an innocent explanation for $o_{i'}(\tilde{g}(\bar{P}_i, \bar{P}_j, P_k, P_l, P_{I'}))$. Moreover,

- $(\bar{P}_j, P_k, P_l, P_{I'})$ is an innocent explanation for $o_i(\tilde{g}(P_i, P_j, P_k, P_l, P_{I'}))$, and
- $(\bar{P}_i, P_k, P_l, P_{I'})$ is an innocent explanation for $o_j(\tilde{g}(P_i, P_j, P_k, P_l, P_{I'}))$.

It remains to show that $\tilde{g}$ is strategy-proof. To start, note that for all agents in I', $\tilde{g}$ produces exactly the same assignments as g. Since g is strategy-proof, no such agent can get a strictly better assignment by misreporting her preferences. Consider the following arguments for agents not in I':

- For agent l and each $P' \in \mathcal{P}$, either $g_l(P') = \tilde{g}_l(P')$ or $\tilde{g}_l(P') \; P_l \; g_l(P')$. First, for each P' such that $g(P') = \tilde{g}(P')$, we have $g_l(\hat{P}_l, P'_{-l}) \; R'_l \; \tilde{g}_l(\hat{P}_l, P'_{-l})$ for all $\hat{P}_l \in \mathcal{P}_l$. Second, for each P' such that $g(P') \neq \tilde{g}(P')$, we have $\tilde{g}_l(P') \; R'_l \; \tilde{g}_l(\hat{P}_l, P'_{-l}) \; R'_l \; g_l(\hat{P}_l, P'_{-l})$ for all $\hat{P}_l \in \mathcal{P}_l$. Together this implies that, l has no incentive to deviate under $\tilde{g}$. Similar arguments apply to k.
- For agent i and each $P' \in \mathcal{P}$, either $g_i(P') = \tilde{g}_i(P')$ or $g_i(P') \; P_i \; \tilde{g}_i(P')$. First, for each P' such that $g(P') = \tilde{g}(P')$, we have $g_i(\hat{P}_i, P'_{-i}) \; R'_i \; \tilde{g}_i(\hat{P}_i, P'_{-i})$ for all $\hat{P}_i \in \mathcal{P}_i$. Also, for each P' such that $g_i(P') \neq \tilde{g}_i(P')$, we have $\tilde{g}_i(P') \; R'_i \; \tilde{g}_i(\hat{P}_i, P'_{-i})$ for all $\hat{P}_i \in \mathcal{P}_i$. Similar arguments apply to j.

Thus, $\tilde{g}$ is strategy-proof and safe. This completes the proof.

□

Proof of Proposition 3. In the following, denote $g = TTC^s$ and suppose that s satisfies the imperfect replacement property. One needs to show that there is no safe deviation that is strategy-proof. Let $\tilde{g}$ be an arbitrary deviation from g. First, if $\tilde{g}$ would be wasteful, then it is not a safe deviation by efficiency of g. Thus, $\tilde{g}$ must be non-wasteful. Second, regarding all preference profiles in $\mathcal{P}$, find the smallest step $t \in \mathbb{N}$ and preference profile $P \in \mathcal{P}$, such that $\sigma^t(P)$ implies that $g(P) \neq \tilde{g}(P)$ in the TTC algorithm. From now on, we call the pair (t, P) the *earliest branch-off* of g and

$\tilde{g}$ and given any $P' \in \mathcal{P}$, we say that we are at (t, P), when we reach submatching $\sigma^{t-1}(P)$ while we run the TTC algorithm with input P'.

Next, consider input P and suppose that we are at (t, P). Then, let I^t be the set of agents that are pointed by an object from $\hat{X}_{\sigma^{t-1}(P)}$. Let $I'^t =: \{i' \in I^t | g_{i'}(P) \neq \tilde{g}_{i'}(P)\}$ and be $\hat{I}^t \subseteq I'^t$ such that $i' \in \hat{I}^t$ if and only if i is assigned at step t with input P. Since we are at (t, P), this implies that for each $i' \in \hat{I}^t$, $g_{i'}(P) P_i \tilde{g}_{i'}(P)$.

It is clear that $|I^t| \leq 3$, if s satisfies the imperfect replacement property. Furthermore, if $g_{i'}(P) \neq \tilde{g}_{i'}(P)$ and $i' \in \hat{I}^t$ is pointed by and points to $g_{i'}(P)$ at (t, P) at step t, then similar arguments as in the proof of Lemma 3 imply that $\tilde{g}$ is not safe. This means that $2 \leq |I^t| \leq 3$. The same arguments imply that we must have $|\hat{I}^t| > 1$. In the following, given any $i' \in I$, consider a preference ranking $P^*_{i'} : g_{i'}(P), \emptyset, \ldots,$ In the following, take input P as given in the TTC algorithm.

Case 1 Let $|\hat{I}^t| = 2$. Thus, there are two agents i and j both in $\hat{I}^t$ and a top-trading cycle $g_j(P) \to i \to g_i(P) \to j$ at (t, P). W.l.o.g. let $\tilde{g}_i(P) \neq g_i(P)$. First, it must be $\tilde{g}_i(P^*_i, P_{-i}) \neq g_i(P)$, as otherwise $\tilde{g}$ would not be strategy-proof for i. Thus, $\tilde{g}_i(P^*_i, P_{-i}) = \emptyset$. Next, assume that $g_j(P) = \tilde{g}_j(P^*_i, P_{-i})$ and recall that $g_j(P)$ must point to i at (t, P). However, this implies that there cannot be $l \in \hat{I}_{\sigma^{t-1}(P)}$, with $l \neq i$ and $g_i(P) = \tilde{g}_l(P^*_i, P_{-i})$, while $g_j(P^*_i, P_{-i}) = \tilde{g}_j(P) = g_j(P)$. Thus, $g_j(P) \neq \tilde{g}_j(P)$. Also, $\tilde{g}_j(P_i, P^*_j, P_{-\{i,j\}}) = \emptyset$, as otherwise $\tilde{g}$ cannot be strategy-proof again. Using a symmetric argument as above, it is also not possible that $\tilde{g}_i(P_i, P^*_j, P_{-\{i,j\}}) = g_i(P)$. Then, strategy-proofness for $\tilde{g}$ would require that $\tilde{g}_i(P^*_i, P^*_j, P_{-\{i,j\}}) = \tilde{g}_j(P^*_i, P^*_j, P_{-\{i,j\}}) = \emptyset$. Thus, by non-wastefulness of $\tilde{g}$ and since (t, P) is the earliest branch-off, there must exist l,$l' \in \hat{I}_{\sigma^{t-1}(P)} \setminus \{i, j\}$ such that $\tilde{g}_l(P^*_i, P^*_j, P_{-\{i,j\}}) = g_i(P)$ and $\tilde{g}_{l'}(P^*_i, P^*_j, P_{-\{i,j\}}) = g_j(P)$. However, this means that $l \neq k$ or $l' \neq k$. W.l.o.g let $l \neq k$ and since s satisfies the imperfect replacement property, there is no $x \in X$ such that $s^x_l > s^x_i$. Hence, i cannot have an innocent explanation for $o_i(\tilde{g}(P^*_i, P^*_j, P_{-\{i,j\}}))$ since l can never be assigned to $g_i(P)$, as long as $g_i(P)$ is i's top choice.

Case 2 Let $|\hat{I}^t| = 3$. First, recall that there cannot be an agent in $\hat{I}^t$ that is assigned to an object she points to at (t, P). Thus, w.l.o.g, we have a cycle with three agents at the earliest branch-off (t, P), $i \to g_i(P) \to j \to g_j(P) \to k \to g_k(P) \to i$ at (t, P). Let $\tilde{g}_i(P) \neq g_i(P)$. Strategy-proofness requires again that $\tilde{g}_i(P^*_i, P_{-i}) \neq g_i(P)$. Hence

$\tilde{g}_i(P_i^*, P_{-i}) = \emptyset$. However, since $|\hat{I}^t| = |I^t|$ and $\tilde{g}$ is non-wasteful, there must exist $l \in \hat{I}_{\sigma^{t-1}(P)} \setminus I^t$ such that $\tilde{g}_l(P_i^*, P_{-i}) = g_{i'}(P)$ for at least one $i' \in \hat{I}^t$. However, this means that $g_{i'}(P)P_{i'}\tilde{g}_{i'}(P)$ and because s satisfies the imperfect replacement property, there is no $x \in X$ such that $s_l^x > s_i^x$. Therefore, i' cannot have an innocent explanation for observation $o_{i'}(\tilde{g}(P_i^*, P_{-i}))$, since l can never be assigned to $g_{i'}(P)$ under g, as long as $g_{i'}(P)$ prefers it to all objects in $\hat{X}_{\sigma^{t-1}(P)}$.

Thus, if $\tilde{g}$ is a safe deviation, then it cannot be strategy-proof. As we selected $\tilde{g}$ arbitrarily, this complete the proof. □

References

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

Atila Abdulkadiroğlu, Parag A Pathak, Alvin E Roth, and Tayfun Sönmez. The boston public school match. *American Economic Review*, 95(2):368–371, 2005.

Mohammad Akbarpour and Shengwu Li. Credible auctions: A trilemma. *Econometrica*, 88(2):425–467, 2020.

Sandeep Baliga, Luis C Corchon, and Tomas Sjöström. The theory of implementation when the planner is a player. *Journal of Economic Theory*, 77(1):15–33, 1997.

Michel Balinski and Tayfun Sönmez. A tale of two mechanisms: student placement. *Journal of Economic Theory*, 84(1):73–94, 1999.

Helmut Bester and Roland Strausz. Imperfect commitment and the revelation principle: the multi-agent case. *Economics Letters*, 69(2):165–171, 2000.

Helmut Bester and Roland Strausz. Contracting with imperfect commitment and the revelation principle: the single agent case. *Econometrica*, 69(4):1077–1098, 2001.

Vianney Dequiedt and David Martimort. Vertical contracting with informational opportunism. *American Economic Review*, 105(7):2141–82, 2015.

Lars Ehlers and Bettina Klaus. Coalitional strategy-proof and resource-monotonic solutions for multiple assignment problems. *Social Choice and Welfare*, 21:265–280, 02 2003.

Lars Ehlers, Isa E Hafalir, M Bumin Yenmez, and Muhammed A Yildirim. School choice with controlled choice constraints: Hard bounds versus soft bounds. *Journal of Economic theory*, 153:648–683, 2014.

Haluk I Ergin. Efficient resource allocation on the basis of priorities. *Econometrica*, 70(6):2489–2497, 2002.

David Gale and Lloyd S Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.

Rohith R. Gangam, Tung Mai, Nitya Raju, and Vijay V. Vazirani. A Structural and Algorithmic Study of Stable Matching Lattices of Multiple Instances. arXiv preprint:2304.02590, 2023.

Yannai A. Gonczarowski and Clayton Thomas. Structural Complexities of Matching Mechanisms. arXiv prepint: 2212.08709, 2023.

Aram Grigoryan and Markus Möller. A theory of auditability for allocation and social choice problems. Technical report, Working paper, 2023.

Isa E Hafalir, M Bumin Yenmez, and Muhammed A Yildirim. Effective affirmative action in school choice. *Theoretical Economics*, 8(2):325–363, 2013.

Rustamdjan Hakimov and Madhav Raghavan. Improving transparency and verifiability in school admissions: Theory and experiment. Technical report, Working paper, 2023.

Yuichiro Kamada and Fuhito Kojima. Efficient matching under distributional constraints: Theory and applications. *American Economic Review*, 105(1):67–99, 2015.

Onur Kesten. On two competing mechanisms for priority-based allocation problems. *Journal of Economic Theory*, 127(1):155–171, 2006.

Fuhito Kojima. School choice: Impossibilities for affirmative action. *Games and Economic Behavior*, 75(2):685–693, 2012.

Jacob D Leshno and Irene Lo. The Cutoff Structure of Top Trading Cycles in School Choice. *The Review of Economic Studies*, 88(4):1582–1623, 11 2020.

Shengwu Li. Obviously strategy-proof mechanisms. *American Economic Review*, 107 (11):3257–87, 2017.

Pinaki Mandal and Souvik Roy. On obviously strategy-proof implementation of fixed priority top trading cycles with outside options. *Economics Letters*, 211:110239, 2022.

David G McVitie and Leslie B Wilson. Stable marriage assignment for unequal sets. *BIT Numerical Mathematics*, 10(3):295–309, 1970.

Szilvia Pápai. Strategyproof assignment by hierarchical exchange. *Econometrica*, 68 (6):1403–1433, 2000.

Szilvia Pápai. Strategyproof and nonbossy multiple assignments. *Journal of Public Economic Theory*, 3(3):257–271, 2001.

Marek Pycia and Peter Troyan. A theory of simplicity in games and mechanism design. *Econometrica*, 91(4):1495–1526, 2023.

Marek Pycia and M Utku Ünver. Trading cycles for school choice. Technical report, Working paper, 2011.

Marek Pycia and M Utku Ünver. Incentive compatible allocation and exchange of discrete resources. Technical report, Working paper, UCLA and Boston College, 2014.

Marek Pycia and M Utku Ünver. Incentive compatible allocation and exchange of discrete resources. *Theoretical Economics*, 12(1):287–329, 2017.

Marek Pycia and M Utku Ünver. Ordinal simplicity and auditability in discrete mechanism design. *Available at SSRN*, 2023.

Alvin E Roth. Incentive compatibility in a market with indivisible goods. *Economics letters*, 9(2):127–132, 1982.

Alvin E Roth and Elliott Peranson. The effects of the change in the nrmp matching algorithm. *JAMA*, 278(9):729–732, 1997.

Mark A Satterthwaite and Hugo Sonnenschein. Strategy-proof allocation mechanisms at differentiable points. *The Review of Economic Studies*, 48(4):587–597, 1981.

Nicholas Schuler. CPS OIG Uncovers Widespread Admissions Irregularities in K-8 Options for Knowledge Program. Office of Inspector General, Chicago Board of Education. Press Release, February 21, 2018.

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

Lars-Gunnar Svensson. Queue allocation of indivisible goods. *Social Choice and Welfare*, 11(4):323–330, 1994.

Peter Troyan. Obviously strategy-proof implementation of top trading cycles. *International Economic Review*, 60(3):1249–1261, 2019.

Alexander Westkamp. An analysis of the german university admissions system. *Economic Theory*, 53(3):561–589, 2013.

Kevin Jon Williams. A reexamination of the nrmp matching algorithm. national resident matching program. *Academic medicine: journal of the Association of American Medical Colleges*, 70(6):470–6, 1995.

Kyle Woodward. Self-auditable auctions. Technical report, Working paper, 2020.