Essays on Matching Markets

Inauguraldissertation
zur
Erlangung des Doktorgrades
der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der
Universität zu Köln

2022

vorgelegt
von
Markus Möller, M.Sc.
aus
Köln

Referent: Prof. Dr. Alexander Westkamp
Korreferent: Prof. Dr. Christoph Schottmüller

Tag der Promotion: 07.07.2022

This thesis consists of the following works:

Möller, Markus (2021):
Transparent Matching Mechanisms,
Working Paper.

Chen, Yiqiu and Möller, Markus (2021):
Regret-Free Truth-Telling in School Choice with Consent,
Working Paper.

Möller, Markus (2022):
Envy and Strategic Choice in Matching Markets,
Working Paper.

Acknowledgments

I would like to express my sincere gratitude to my thesis advisor Alexander Westkamp for his generous and continuous guidance, valuable advice, patience and support. Many thanks also to Christoph Schottmüller for the co-supervision and many very valuable comments on my projects.

Of course, I also thank Yiqiu Chen, my colleague and great friend, with whom I had the opportunity to experience these exciting years in my PhD studies. These years were full of interesting discussions and were marked by wonderful collaboration and mutual encouragement from the beginning to the end. I also thank Marius Gramb for great suggestions and all the interesting comments and questions.

For financial support, I would like to thank the Cologne Graduate School, which provided me with scholarships for four years, and the German Research Foundation (DFG), which funded me the last year under the German Excellence Strategy - EXC 2126/1- 390838866.

Most of all, however, I thank my parents and family for their unconditional love, support and encouragement in the past years. Last but not least, I thank my companion Nadja, who has always been there for me, giving me the peace and serenity to write this thesis. This thesis is dedicated to all of you.

Contents

1

Introduction

This thesis consists of three essays in matching theory and market design. In particular, the three essays address the effects of incomplete information and behavioral biases on centralized matching markets.

In Chapter 2, I study the transparency of matching mechanisms in an one-to-one object allocation model through the lens of a commitment problem of a central matching authority. Chapter 3 and Chapter 4 are concerned with market participants' incentives of acting truthfully in centralized matching markets if preferences are subject to behavioral biases. Specifically, Chapter 3, which is joint work with Yiqiu Chen, studies regret-based incentives for students in the context of public school assignment. In Chapter 4, I examine an one-to-one object allocation model and ask under what conditions participants' incentives are shaped by aversion to the experience of envy.

Beginning with the seminal work of Gale and Shapley (1962), many applications have benefited from theoretical research on matching markets. This thesis focuses on the assignment of market participants to a limited amount of discrete resources without using transfers, as it is common for applications such as school choice (Abdulkadiroğlu and Sönmez, 2003), campus housing (Hylland and Zeckhauser, 1979; Shapley and Scarf, 1974) or kidney exchange (Roth et al., 2004). These (one-sided) matching markets often operate priority-based in the sense that the limited resources, i.e., objects such as school seats or campus flats, are each equipped with a priority ranking over the participants. However, as an input of the matching mechanism, the priorities are not mere guides to the mechanism; they also fundamentally shape participants' perceptions of how fair the assignment process is. Specifically, under a final matching a participant could find an object she likes more than her own assignment being assigned to another participant. If that participant has lower priority at this more desirable object than she has, then this may be perceived as unfair and is called an instance of justified envy (Abdulkadiroğlu and Sönmez, 2003). Justified envy-freeness can be achieved by adopting stable mechanisms (Gale and Shapley, 1962), a class of mechanisms which has been applied very successfully in practice across various matching applications (Roth, 2008). Now, whereas in some matching models, stability is equivalent to efficiency, this equivalence breaks down in the context of priority-based object allocation. In fact, the equivalence is replaced by an incompatibility of these two desirable criteria, as there is no matching mechanism that always selects an efficient and stable matching (Roth, 1982a; Balinski and Sönmez, 1999).

The resulting trade-off between efficient mechanisms and stable mechanisms raises two important questions that have been addressed by market design and matching theory over the past two decades and are explored in more detail in this thesis. First, if the market designer only has the choice of selecting a mechanism that belongs to one of the two classes, which of the classes should be selected in favor of the other? Second, does an appropriate relaxation of the stability requirement open up new avenues with alternative matching mechanisms that can also work well in practice? The answers to both questions depend, of course, on the underlying objectives, the particular characteristics of the market, and the intended applications. However, some market characteristics, such as incomplete information and behavioral biases, are present in many real-world applications and are therefore of general interest. From this perspective, this thesis examines participants' incentive problems and central authorities' commitment problems when participants' information about the market is limited and/or when participants have nonstandard preferences. A compelling argument for focusing on incentive and commitment problems is that the benefits of a mechanism are often realized only when they are based on the actual preferences of market participants. Yet, since participants' preferences are private information, their use can only be ensured if participants have an appropriate incentive and trust the responsible authority enough to actually disclose them. Thus, mechanisms that perform well in solving the above problems stand a good chance of proving themselves in practice by effectively steering market participants toward their preferred actions and by improving market transparency.

A theoretical benchmark for good incentive properties of a mechanism in the literature is strategy-proofness which requires that it is a weakly dominant strategy for each student to reveal her true preferences. In Chapter 2, I study the "transparency" of strategy-proof mechanisms using an one-to-one object allocation model by taking the perspective of a central authority's ability to commit to a publicly announced mechanism. Specifically, I develop commitment criteria which describe to which extent agents can be confident that the authority sticks to the announcements made. In the model, a central matching authority publicly announces a strategy-proof mechanism and upon receiving agents' reported preferences the authority privately selects a mechanism to compute a matching that is publicly revealed. Agents' preferences are private information, other market features are common knowledge and agents observe the final matching. In this sense, the authority's commitment to comply with an announcement is linked to the ability of agents to detect deviations from the announced mechanism through their observations and their prior information.

The main result of the chapter is that the authority can commit to stability. By contrast, I show that commitment to efficiency, only works for dictatorships. The key drivers of the main results are intuitive since stability of a matching can be described by passing an independent test for each agent that checks observable conditions on the matching outcome, where the test of an agent is passed or failed without the need for reference to other agents' reported preferences. The efficiency criterion, on the other hand, cannot generally be described in this fashion. Concretely, efficiency requires that agents are aware that there is no remaining mutual interest for exchange, which means that agents need knowledge on other agents' preferences to verify efficiency. The main contribution of this chapter is to identify a transparency advantage of stability over efficiency, which could be an important factor in the choice of mechanisms in practical applications.

Chapter 3 explores students' incentives to report an honest ranking over schools if students wish to avoid regret in the context of public school assignment. It is co-authored by Yiqiu Chen and both authors contributed equally to this project. We adopt a notion of regret proposed by Fernandez (2020): A student regrets her report at her match, if given her feedback from the matching mechanism, she finds another report, which first, does not match her worse for all market unknowns compatible with the feedback and second, matches her strictly better for some of the compatible unknowns. In our framework, students privately submit a report to a matching mechanism and the mechanism gives feedback including the final matching. However, the inputs of the mechanism, such as the reports of other students are not disclosed even ex-post. The incentive concept we study is called regret-free truth-telling (Fernandez, 2020) and is weaker than strategy-proofness. It requires that no student ever regrets reporting her true preferences. We examine whether regret-free truth-telling can be satisfied by mechanisms that aim to address the trade-off between efficiency and stability, none of which are strategy-proof (Abdulkadiroğlu et al., 2009; Erdil and Ergin, 2008; Alva and Manjunath, 2019). Among the candidates one mechanism stands out with a quite elegant approach—the Efficiency Adjusted Deferred Acceptance Mechanism (EDA) (Kesten, 2010) of which the idea is to improve efficiency on stable matchings, however, only if students give their consent for relaxing the stability constraint.

Our main result is to show that EDA is regret-free truth-telling, which can be regarded as a practical contribution highlighting EDA's potential for implementation in applications such as school choice. In addition, we study a class of mechanisms called efficient stable dominating rules. Efficient stable dominating rules are mechanisms which always induce an efficient matching that Pareto dominates a stable matching and is a refinement of stable dominating rules introduced by Alva and Manjunath (2019). In contrast to EDA, efficient stable dominating rules do not consider consent, and the result regarding efficient stable dominating rules is negative, i.e., none of them is regret-free truth-telling.

Finally, in Chapter 4, I study agents' incentives to be truthful if agents' behavior is influenced by the dislike of envy. I employ an one-to-one agent object allocation model, where an agent can envy another agent if that other agent is assigned to an object that she prefers over her own assignment, i.e., envy is different from justified envy since priorities are not taken into account. I develop a new incentive concept called envy-proofness, which is stronger than strategy-proofness. Envy-proofness requires that for an agent who reports her true preferences, every instance of envy she experiences is inevitable in the sense that no matter how she changes her action, she experiences these instances as well. In other words, given the action of other agents, the instances of envy she experiences under her true preferences is a weak subset of the instances of envy she experiences under any other action she takes. In this sense, envy-proofness ensures that no agent has an incentive to lie about her preferences if she wishes to avoid envy and also wants to be assigned to her most preferred object.

The main results of this chapter are that the efficient Top Trading Cycle (TTC) Mechanism (Shapley and Scarf, 1974) is envy-proof, whereas I show that the unique stable and strategy-proof mechanism, the Deferred Acceptance (DA) Mechanism (Gale and Shapley, 1962) is not envy-proof. The main contribution of Chapter 4 is to provide new insights about the trade-off between efficiency and stability. More concretely, while stable mechanisms by definition minimize certain forms of fairness-related envy (justified envy), stable mechanisms are shown not to account for various forms of envy-driven incentives. On the other hand, some efficient mechanisms including TTC (see also Chapter 4 for more details) seem to be more robust to envy-driven incentives.

2

Transparent Matching Mechanisms*

2.1 Introduction

In matching theory the central authority usually appears as a pure and honest operator of the matching mechanism. In practice, however, the authority's overall objectives could be in conflict with using the mechanism that was promised to participants in advance. For instance, rather than implementing an efficient matching for applicants, a campus housing authority might be interested in fully utilizing

*This chapter is based on Möller (2021a). Thanks to my advisor, Alexander Westkamp. Thanks to Yiqiu Chen, Christoph Schottmüller and Marius Gramb for helpful comments. Remaining errors are of course my own.

the campus housing supply. Another example is a school choice authority that wants to satisfy distributional goals that can not be achieved by using the announced mechanism. As in both examples, the authority's commitment to comply with an announcement is tied to the ability of participants to detect deviations from the announced mechanism. This paper examines to which extent participants can be confident that the authority sticks to the announcements made.

I employ a one-to-one object allocation model, where a central authority publicly announces a *strategy-proof* [1] mechanism. Then, upon receiving agents' reported preferences, the authority privately selects a mechanism to induce a publicly observable matching. I assume that agents' strict preferences over objects are private information, while other market features are common knowledge, including how the announced mechanism works. Agents trust the announcement as long as they have not received any counterfactual evidence in this regard. In this setting, I adopt the notions of *innocent explanations* and *safe deviations* by 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 reported 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 the announced mechanism, if for each agent, each observation produced by the mechanism has an innocent explanation.

I develop a set of "transparency" notions to explore different forms of commitment to an authority's announcement. For the case of full commitment I introduce the notion of *full-transparency* which requires that the authority has no safe deviation from the announced mechanism. It turns out that for the strategy-proof mechanisms studied in this paper, full-transparency can only be achieved for mechanisms which are dictatorial. More precisely, I show that a strategy-proof and efficient mechanism is fully-transparent if and only if it is identical to a *sequential dictatorship* known from (Pápai

[1] Strategy-proofness requires that it is a weakly dominant strategy for agents to report their true preferences over objects.

(2001), Ehlers and Klaus (2003) and Pápai (2000)). Furthermore, the unique strategy-proof and stable[2] mechanism, known as the *Deferred Acceptance (DA) Mechanism* from Gale and Shapley (1962), is fully-transparent if and only if it is equivalent to a *serial dictatorship* from Satterthwaite and Sonnenschein (1981) and Svensson (1994).

Moving to the case of partial commitment, I define the notions *identity-transparency* and *cardinality-transparency*, which do not permit safe deviations under which the set and number of matched agents change, respectively. A main finding of this paper is that *DA* satisfies both of these criteria, while the same holds for *TTC* if and only if *TTC* is stable. In fact, both *DA* and *TTC* were touted as candidates for assigning students to Boston Public Schools in 2005. The committee ultimately chose *DA*, arguing that *"the behind the scenes mechanized trading [in TTC] makes the student assignment process less transparent."*. More recently, Leshno and Lo (2020) characterized *TTC* outcomes in terms of cutoffs, which take the form of competitive-equilibrium prices assigned to *each pair of objects* and from which the agent's eligibility for an object can be inferred. As Leshno and Lo (2020) argue, the interpretation of *TTC* cutoffs are thus more complicated than those that can be used to explain the results of *DA* (Azevedo and Leshno, 2016), since under *DA* eligibility for an object is described by a single cutoff *for each object*.[3]

Adding another form of partial commitment, I also ask to which extent the authority can commit to a desirable property of the announced mechanism. More concretely, a mechanism with property p is *p-transparent*, if each safe deviation from the mechanism must still have property p. A main result of this paper is that *DA* is *stable-transparent*. However, *DA* allows the authority to safely deviate to any

[2]Stability of a matching is defined given a set of object specific priorities over agents and a set of agents' preferences: A matching is *blocked*, if there exists an agent and an object she prefers to her own match and another agent with lower priority is assigned to this preferred object. A matching is *non-wasteful* if there is no object that is unassigned although there is an agent that prefers the object over her own assignment. A matching is *individually rational* if no agent prefers the outside option over her final assignment. A matching is *stable* if it individually rational, non-wasteful and not blocked. A mechanism is stable if it only induces stable matchings for each reported preference profile.

[3]In case of *DA*, the cutoff entails information about the lowest priority an agent must have at an object to be eligible for it.

other stable mechanism. I also show that an efficient and group strategy-proof mechanism is *efficient-transparent* if and only if it is equivalent to a sequential dictatorship mechanism.[4]

Related literature

This paper is among the first to relax the authority's full commitment assumption in the context of matching markets. Closely related in this regard is Hakimov and Raghavan (2020) who propose a transparency notion for general centralized allocation settings. Similar to the notion of full transparency considered in the current paper, their transparency notion requires that agents detect every deviation and thus only applies to the case of full commitment. They show that sequential versions of *DA* and serial dictatorships can be implemented in a transparent way. However, a crucial difference to the current paper is that transparency is mainly a consequence of designing suitable sequential public communication between authority and agents. By contrast, in this paper, the communication between each agent and the authority *remains* private. Thus, the transparency notions I consider are features of the mechanism, whereas in Hakimov and Raghavan (2020) transparency is a feature of the general information structure.[5]

The second paper that is closely connected is Akbarpour and Li (2020) from which I adopt the notions of innocent explanations and safe deviations. Akbarpour and Li (2020) study a partial commitment framework with sequential private communication between the authority and agents and examine the authority's commitment to various standard auction formats. They focus on *credible* implementations of Bayes-Nash Mechanisms with imperfect information, which requires that the au-

[4]A mechanism is group strategy-proof if there is no group of agents that can generate weakly better assignments by misrepresenting their preferences and at least one agent in the group profits from the misrepresentation.

[5]In this sense related is also Woodward (2020), who analyzes a setting where an auctioneer maximizes her utility conditional on information that is released to bidders after the auction is run. For different auction formats and incentives of the auctioneer, Woodward (2020) asks what information bidders need to have to conclude that the auction was run as claimed.

thority has no incentive to safely deviate to another mechanism. Akbarpour and Li (2020) show that the first-price auction is the unique static credible and optimal auction, which drives a wedge between incentive compatibility for the authority and strategy-proofness for agents. The key differences to my analysis are the following. First, whereas credibility requires incentive compatibility for the authority, transparency remains silent on the authority's incentives and only asks for potential safe deviations. Second, Akbarpour and Li (2020) analyze also extensive-form mechanisms. Third, Akbarpour and Li (2020) study credibility given an equilibrium of the announced mechanism and additionally require that agents' equilibrium strategies remain optimal conditional on the authority's set of safe deviations. By contrast, in the current paper agents do not take the authority's strategic behavior into account. To motivate this last departure, note that in many matching applications (e.g. school choice, college admission or house allocation) participants are often inexperienced and typically take part only once. It thus seems natural that participants will first not distrust the authority in these settings. In the context of auctions, however, bidders may be more sophisticated and can be repeatedly exposed to the same auctioneer. In this sense, a bidder may have a clearer picture of the auctioneer's incentives and thus takes the auctioneer's strategic behavior into account when deciding how to bid.

Also related is the line of research that follows Li (2017)'s work on *obvious-strategy-proofness (OSP)*.[6] He characterizes OSP mechanisms as those that can be supported by bilateral commitments between agents and authority. Li (2017) applies his framework to matching contexts and obtains that *TTC* is not OSP-implementable in some settings. Following the work of Li (2017), Troyan (2019) shows that *TTC* is OSP-implementable if and only if the priority structure satisfies an acyclicity condition that is weaker than the one characterizing the priority structures for which *TTC* is identity-transparent.[7] Ashlagi and Gonczarowski (2018) establish that stable mechanisms are not always OSP-implementable,

[6] A mechanism is OSP if each agent has an obviously dominant strategy which requires that the worst possible outcome from following the truth is better than the best case outcome from any possible untruthful report.

[7] Specifically, the acyclicity condition used in the current work is due to Ergin (2002) and characterizes the priority structures for which *TTC* is stable and thus those for which *TTC* is identity-transparent.

whereas Thomas (2020) characterizes the priority structures under which DA is OSP-implementable. Note that apart from the fact that Li (2017) limits attention to strategy-proof mechanisms and that the authority's incentives do not play a role, the commitment framework is similar to the framework in Akbarpour and Li (2020). In particular, the key differences to the current work and the work that follows Li (2017) coincide with those outlined for Akbarpour and Li (2020) in the previous paragraph.

More broadly, this paper relates to the literature which models limited commitment as measurable with respect to agents' observations on final outcomes, as for example in Dequiedt and Martimort (2015), Baliga et al. (1997), Bester and Strausz (2000) and Bester and Strausz (2001).

The rest of this paper is organized as follows. Section 2.2 introduces the basic model along with the transparency framework. Section 2.3 analyzes the transparency characteristics of DA. Section 2.4 contains the analysis of efficient mechanisms. Section 2.5 concludes.

2.2 Preliminaries

2.2.1 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. To avoid trivial cases I 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 $X \cup \{\emptyset\}$. Equip each object $x \in X$ with a strict priority ranking $\rhd_x$ over agents I. A priority structure $\rhd \equiv (\rhd_x)_{x \in X}$ is a profile of priority rankings and the domain of all priority structures is denoted with $\bar{P}r$. For the rest of this paper, fix an arbitrary priority structure $\rhd \in \bar{P}r$.

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.[8] For each $x \in X$ and $i \in I$, object x is acceptable if $x \; P_i \; \emptyset$ for i and x is

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

unacceptable for i if it is not acceptable. I refer to P_i as agent i's type and to $P \equiv (P_i)_{i \in I}$ as a type profile. For each $i \in I$, let $\mathcal{P}_i$ be the domain of all possible types and let $\mathcal{P} = \times_{i \in I} \mathcal{P}_i$ be the domain of all type profiles. For any $J \subset I$, $P_J = (P_j)_{j \in J}$ is a type 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 type 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 $x \; P_i \; \mu_i$ and x is unassigned under μ. Call the matching μ *individually rational* if, for each $i \in I$, $\mu_i \; R_i \; \emptyset$. The matching μ is *blocked* if there exists a pair of agents $i, j \in I$ and an object $x \in X$ such that $x \; P_i \; \mu_i$, $\mu_j = x$ and $i \rhd_x j$. Then, one refers to the matching μ as $\rhd$*-stable* with respect to P if it is not blocked, individually rational and non-wasteful. Let $\Sigma^{\rhd}(P)$ be the set of $\rhd$-stable matchings with respect to P. 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 type profiles into matchings. For each $P \in \mathcal{P}$, let $g_i(P)$ denote the assignment of agent $i \in I$ under $g(P)$. Let $\mathcal{G}$ be the set of all mechanisms. Consider the following standard properties given any mechanism $g \in \mathcal{G}$. 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 $\rhd$*-stable* if it produces a $\rhd$-stable matching for each type profile. Moreover, mechanism g is *(Pareto) efficient* if it only induces efficient matchings.

I proceed with two standard incentive notions, where the first one requires that it is a weakly dominant strategy for agents to report their true type. Formally, define a mechanism $g \in \mathcal{G}$ 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 $\mathcal{SP} \subset \mathcal{G}$ the set of strategy-proof mechanisms. The second notion requires that no group of agents can jointly misrepresent their types, such that each agent in the group is weakly better off and at least one agent in the group is strictly better off. A mechanism $g \in \mathcal{G}$ 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 $j \in J$.

2.2.2 A Transparency Framework

Let a central matching authority make a public announcement in form of a strategy-proof mechanism $g \in \mathcal{SP}$. The announcement is treated as a promise that g will be used to assign objects $X \cup \{\emptyset\}$ to agents I. Agents know $I, X, \triangleright$, their own type and how g maps type profiles into matching outcomes. Furthermore, assume that agents do not receive any information on other agents' types and actions.

Once the announcement has been made the authority privately selects a mechanism $\tilde{g} \in \mathcal{G}$ that will be used to produce a final matching. Refer to $\tilde{g}$ as a *deviation* from g, if there exists a type profile $P' \in \mathcal{P}$ such that $\tilde{g}(P') \neq g(P')$. Given any reported type profile, the induced final matching is publicly revealed and observable for all agents.

For expositional simplicity, it will be useful to treat an agent's type as a part of her individual observation. Formally, given a mechanism $g' \in \mathcal{G}$, a type profile $P \in \mathcal{P}$ and an agent $i \in I$, *observation* $o_i(g'(P))$ is a pair which consists of agent i's type P_i and the final matching $g'(P)$. In this setting, I adopt criteria developed by Akbarpour and Li (2020) which specify the conditions under which agents' individual observations conceal or reveal a deviation:

Definition 2.1 (Akbarpour and Li (2020))**.** Given announcement $g \in \mathcal{SP}$, deviation $\tilde{g} \in \mathcal{G}$ and a type profile $P \in \mathcal{P}$, agent $i \in I$ has an *innocent explanation* for observation $o_i(\tilde{g}(P))$ if there exists $P'_{-i} \in \mathcal{P}_{-i}$ such that $o_i(\tilde{g}(P)) = o_i(g(P_i, P'_{-i}))$.

That is, agent i's observation under the deviation has an innocent explanation if one of i's observations that could follow from the announcement is identical to the one she made under the deviation.

Definition 2.2 (Akbarpour and Li (2020))**.** Given announcement $g \in \mathcal{SP}$, 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.

In words, a deviation is safe if each observation produced by the deviation has an innocent explanation for the agent who makes the observation. In what follows, I assume that agents tell the truth about their preferences if the designer sticks to selecting only safe deviations. That is, given any type profile $P \in \mathcal{P}$, each agent $i \in I$ submits her true type P_i to the authority.

I now define a set of notions to study the authority's commitment to its announcement. The first and straightforward notion requires full commitment power.

Definition 2.3. An announcement $g \in \mathcal{SP}$ satisfies *full-transparency* if there exists no safe deviation $\tilde{g} \in \mathcal{G}$ from g.

In other words, a mechanism is fully-transparent if any deviation will be detected by at least one agent. However, as will become clear in the course of this paper, the more interesting findings lie beyond the case of full commitment, since of the mechanisms examined, only dictatorships guarantee full-transparency.

Specifically, I define a set of weaker transparency notions that address different forms of partial commitment. For example, take a school choice setting in which properties such as stability or efficiency are usually perceived as desirable by participants. In the case of an announced mechanism which satisfies such a desirable property, that property may even be publicly advertised to convince agents of the mechanisms' benefits. However, it is unclear to what extent participants may view the authority's claim to implement a particular property as valid. In light of these considerations, the next transparency notion sets up the conditions under which an authority may commit to a desirable property of the announced mechanism. Concretely, let p be an arbitrary property or condition that can be satisfied by a mechanism and collect in $\mathcal{G}^p \subseteq \mathcal{G}$ all mechanisms sharing the property p.

Definition 2.4. An announcement $g \in \mathcal{SP}$ is *p-transparent* if there exists no safe deviation $\tilde{g} \notin \mathcal{G}^p$ from g.

In other words, given announcement $g \in \mathcal{G}^p$ the transparency notion ensures that the authority has full commitment power to a mechanism with property p. As a concrete example, if the authority announces a mechanism with property $\triangleright$-stable and the mechanism is $\triangleright$-stable-transparent, then the authority has full commitment to induce only $\triangleright$-stable outcomes.

Next, I introduce two transparency notions which are not property-specific. Under these notions the authority can not change either the sets or the numbers of matched agents without being detected. In this way, one can already substantially limit the authority's scope even if one does not have detailed information what deviations the authority would perceive as desirable. For any mechanism $g \in \mathcal{G}$ and any type profile P, be $g_I(P)$ the set of agents assigned to an object in X under $g(P)$. Then, given a mechanism $g \in \mathcal{G}$, another mechanism $\tilde{g} \in \mathcal{G}$ is *identity-equivalent* to g if for each $P \in \mathcal{P}$, we have $\tilde{g}_I(P) = g_I(P)$. In a similar vein, mechanism $\tilde{g} \in \mathcal{G}$ is *cardinality-equivalent* to g, if for each $P \in \mathcal{P}$, we have $|\tilde{g}_I(P)| = |g_I(P)|$. Now consider the following transparency notions.

Definition 2.5. An announcement $g \in \mathcal{SP}$ is

1. *identity-transparent* if any deviation which is not identity-equivalent to g is not safe.

2. *cardinality-transparent* if any deviation which is not cardinality-equivalent to g is not safe.

In words, whereas identity-transparency requires that under any safe deviation and given any profile of types, the set of matched agents is the same as under the announcement, under cardinality-transparency it is only required that the number of matched agents is the same. Hence identity-transparency implies cardinality-transparency, whereas the converse is not true.

2.3 Stable Mechanisms

This section deals with the transparency of $\rhd$-stable mechanisms. It should be noted that the results in this section generalize to the setting of many-to-one matching. As a preliminary work for the derivation of the main results, I establish that stability can be verified by agents independently—a feature that does not apply for the efficiency criterion studied in Section 2.4. To make this feature apparent, I first decompose the conditions of $\rhd$-stability agent by agent. Fix a matching $\mu \in \mathcal{M}$ and a type profile $P \in \mathcal{P}$. Given any agent $i \in I$, say that μ is *i-blocked* at $x \in X$, if there exists $j \in I \setminus \{i\}$ such that $x \; P_i \; \mu_i$, $\mu_j = x$ and $i \rhd_x j$. Matching μ is *i-non-wasteful* if there is no $x \in X$ unassigned under μ for which $x \; P_i \; \mu_i$. Say that μ is *i-individually rational* if $\mu_i \; R_i \; \emptyset$. Then, matching μ is *i-$\rhd$-stable* with respect to P_i, if μ is not i-blocked, i-individually rational and i-non-wasteful. Let $\Sigma^{\rhd}(P_i)$ be the set of matchings which are i-$\rhd$-stable with respect to P_i. Note that it does not depend on the preferences of any agent $j \neq i$. The following result is immediate from the two stability definitions.

Lemma 2.1. *For each $P \in \mathcal{P}$, $\Sigma^{\rhd}(P) = \bigcap_{i \in I} \Sigma^{\rhd}(P_i)$.*

I now move to the unique strategy-proof and $\rhd$-stable mechanism which is known to be equivalent to the *agent-proposing Deferred Acceptance (DA) Mechanism* (Gale and Shapley, 1962). Denote the *DA* mechanism that is $\rhd$-stable with $DA^{\rhd}$. The *DA Algorithm* (Gale and Shapley, 1962) which induces $DA^{\rhd}$ can be found in Appendix 2.D.

We are ready for the main result of this section.

Theorem 2.1. *$DA^{\rhd}$ is $\rhd$-stable-transparent.*

To prove this statement, I mostly rely on Lemma 2.1.

Proof. Given announcement $DA^{\triangleright}$ consider an arbitrary deviation $\tilde{g}$ that is not $\triangleright$-stable. One has to show that deviation $\tilde{g}$ is not safe. Take any $P \in \mathcal{P}$ for which $\tilde{g}(P) \notin \Sigma^{\triangleright}(P)$. By Lemma 2.1, there exists $i \in I$ such that $\tilde{g}(P)$ is not in $\Sigma^{\triangleright}(P_i)$. Now consider agent i's observation $o_i(\tilde{g}(P))$. By Lemma 2.1, for any $P'_{-i} \in \mathcal{P}_{-i}$, we have $DA(P_i, P'_{-i}) \in \Sigma^{\triangleright}(P_i)$. Thus, i cannot have an innocent explanation for $o_i(\tilde{g}(P))$ in which $\tilde{g}(P)$ is not in $\Sigma^{\triangleright}(P_i)$. Therefore, $\tilde{g}$ is not a safe deviation and $DA^{\triangleright}$ is $\triangleright$-stable-transparent. □

It turns out that Theorem 2.1 can be used to derive an additional transparency property of $DA^{\triangleright}$. To do so, consider the following famous result on $\triangleright$-stable matchings.

Theorem (Lone Wolf Theorem (McVitie and Wilson, 1970))**.** For any type profile $P \in \mathcal{P}$ and any pair of matchings $\mu, \nu \in \Sigma^{\triangleright}(P)$, we have $\mu_X = \nu_X$ and $\mu_I = \nu_I$.

In words, for any given type profile, the set of assigned objects and agents is the same across all $\triangleright$-stable matchings. Hence given any $\triangleright$-stable mechanism g, any other $\triangleright$-stable mechanism $\tilde{g}$ is identity-equivalent and cardinality-equivalent to g and together with Theorem 2.1 the following result is immediate.

Corollary 2.1. *$DA^{\triangleright}$ is identity-transparent.*

The next result provides information on sufficient conditions for deviations to be safe under $DA^{\triangleright}$. More concretely, Theorem 2.1 and the result presented below together say that a deviation from $DA^{\triangleright}$ is safe if and only if the deviation is $\triangleright$-stable.

Proposition 2.1. *If $\tilde{g}$ is $\triangleright$-stable, then $\tilde{g}$ is a safe deviation with respect to $DA^{\triangleright}$.*

Proof. Consider an arbitrary $\triangleright$-stable deviation $\tilde{g}$ from announcement $DA^{\triangleright}$. 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 type profile $P'_{-i} \in \mathcal{P}_{-i}$ where for each $j \neq i$, j's top choice on P'_j is $\tilde{g}_j(P)$. Now note that since $\tilde{g}(P) \in \Sigma^{\rhd}(P)$, this implies that $\tilde{g}(P) \in \Sigma^{\rhd}(P_i, P'_{-i})$. Moreover, for each agent $j \neq i$, $\tilde{g}_j(P)$ is either the unique acceptable object on P'_j or the outside option and hence we obtain that $\Sigma^{\rhd}(P_i, P'_{-i})$ is a singleton. Thus, one reaches $\tilde{g}(P) = DA^{\rhd}(P_i, P'_{-i})$ and therefore P'_{-i} provides an innocent explanation for $o_i(\tilde{g}(P))$.

Note that agent i and type profile P were chosen arbitrarily and thus 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 $\rhd$-stable deviations was also arbitrary, the proof is complete. □

A direct consequence of Proposition 2.1 and Theorem 2.1 is that full-transparency of $DA^{\rhd}$ is achieved if and only if there exists a unique stable matching in $\Sigma^{\rhd}(P)$ for each $P \in \mathcal{P}$. However, as shown in the following, this can only be guaranteed if $DA^{\rhd}$ reduces to a *serial dictatorship* from Satterthwaite and Sonnenschein (1981) and Svensson (1994) defined as follows. A mechanism $g \in \mathcal{G}$ 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. Clearly, in case of $DA^{\rhd}$, this means that $\rhd_x = \rhd_y$ must hold for all $x, y \in X$.

Proposition 2.2. *$DA^{\rhd}$ is fully-transparent if and only if it is equivalent to a serial dictatorship.*

Proof. See Appendix 2.A. □

2.4 Efficient Mechanisms

In this section, I study the transparency features of efficient mechanisms. I start with two characterizations: First, I characterize the set of strategy-proof and efficient mechanisms that are fully transparent and in a next step, I identify the set of group strategy-proof and efficient mechanisms which are efficient-transparent. I then turn to a popular special case of group strategy-proof and efficient mechanisms, the *Top Trading Cycles (TTC) Mechanism* from Shapley and Scarf (1974). The version of *TTC* under study is known from Abdulkadiroğlu and Sönmez (2003) and operates on a fixed priority structure $\rhd \in \bar{Pr}$, such as is common in school choice. I characterize the domain of priority structures for which $TTC^{\rhd}$ is identity-transparent and cardinality-transparent.

2.4.1 Strategy-Proof and Efficient Mechanisms

I begin with an example that illustrates that announcing a strategy-proof and efficient mechanism may give the authority an opportunity to safely deviate to a non-efficient mechanism. As a motivation, consider an application such as school choice or college admission, where authorities tend to be interested in efficiency, but also to satisfy some distributional constraints that are in conflict with efficiency (e.g., equal distribution of different genders, meeting some regional quota, or other socioeconomic considerations).[9] Thus, a reasonable scenario is that an authority first announces an efficient and strategy-proof mechanism to encourage participation and honest revelation, but then deviates to a non-efficient mechanism that redistributes agents according to a distributional constraint it wants to satisfy. For illustrative purposes, I keep the size of the example small, but note that one can easily extend the example to a market of larger size.

Example 2.1. Let $I = \{i, j\}$ and $X = \{x, y\}$ and suppose that the authority announces g such that

- if i reports y as her top choice, then she is assigned to y and j gets her favorite object among the remaining ones.
- if j reports x as her top choice, then she is assigned to x and i gets her favorite object among the remaining ones.
- if i and j report two different top choices, assign both agents to their top choices.

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

It is easily checked that g is efficient and strategy-proof. Next, consider a selection of agents' types presented in the table below:

P_i	P'_i	P_j	P'_j
x	y	y	x
y	x	x	y
$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$

Assume that once g is announced, the authority uses mechanism $\tilde{g} \in \mathcal{G}$, where

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

Note that $\tilde{g}$ is a deviation from g and that $\tilde{g}$ is not efficient, since $g(P) \neq \tilde{g}(P)$ and, given type profile P, agents prefer to swap their assigned objects under $\tilde{g}(P)$. As motivated in the preface of the example, redistributing agent i and j might be in line with some distributional goal of the authority that depends on agents' characteristics.

To show that $\tilde{g}$ is a safe deviation, I set out to find innocent explanations for observations $o_i(\tilde{g}(P))$ and $o_j(\tilde{g}(P))$. First, type P'_j provides an innocent explanation for $o_i(\tilde{g}(P)) = o_i(g(P_i, P'_j))$. Symmetrically, P'_i leads to an innocent explanation for $o_j(\tilde{g}(P)) = o_j(g(P'_i, P_j))$. Finally, under any other type profile the matchings under g and $\tilde{g}$ coincide. Thus, $\tilde{g}$ is safe and we conclude that g is not efficient-transparent.

Note that the announced mechanism g essentially distributes guarantees to obtain objects y and x among i and j, respectively. These guarantees can be exploited by the authority once both agents rank all objects as acceptable and their top choices differ from their own guarantees. Specifically, in this case, agents can be assigned to the objects guaranteed to them, since this always provides an innocent explanation for the remaining agent. $\square$

As will become clear soon, the construction of the safe deviation in the example has some generality and will be helpful to prove the main results of this section.

Next, I introduce *sequential dictatorship mechanisms*—a class of Pareto efficient and group strategy-proof mechanisms known from Pápai (2001), Ehlers and Klaus (2003) and Pápai (2000) which play a key role in 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

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

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 2.6. A mechanism $g \in \mathcal{G}$ 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 type 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 2.2. *Let announcement $g \in \mathcal{SP}$ be efficient. Then, g is fully-transparent if and only if it is a sequential dictatorship.*

Proof. See Appendix 2.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. Note that 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. Now, if the authority deviates from some type profile, then following the correct ordering of dictators, there must be a first agent who infers the stage she must have been the dictator and notices that 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 and 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 also central to the next result, I briefly explain the basic line of reasoning here. Concretely, consider again how I constructed the safe deviation in Example 2.1. 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* which can be implemented via the *TC Algorithm*. Under each step of the *TC* 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.[10] Now, whenever a *TC* mechanism is not equivalent to a sequential dictatorship, then at some step of the *TC* algorithm, two different agents are pointed by the objects in the form of guarantees similar to those distributed to agents in Example 2.1. Once reaching this step, the authority can exploit agents' guarantees if the two agents prefer each others' guaranteed objects most and their own guarantees are their second preferred choices. 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 2.B for the formal statement of the characterization by Pycia and Ünver (2017) along with a description of the *TC* algorithm.

Next, note that the deviation described in the last paragraph is not efficient and that similar to Example 2.1, a possible way to motivate such a non-efficient deviation might be the tension between efficiency and some additional distributional constraint the authority wants to satisfy. Hence, the following result is immediate from applying the outlined arguments also used to prove Theorem 2.2.

Theorem 2.3. *Let announcement $g \in \mathcal{SP}$ be efficient and group strategy-proof. Then, the following three statements are equivalent: Announcement g is*

1. *fully-transparent.*
2. *efficient-transparent.*
3. *a sequential dictatorship.*

[10]The idea of the *TC* algorithm builds on Gale's *Top Trading Cycles (TTC) Algorithm* (Shapley and Scarf, 1974). However, pointing rules are more complex under *TC* compared to *TTC*. In fact, the mechanisms induced by the *TTC* algorithm are special cases of *TC* mechanisms and will be studied in more detail in the next subsection.

Proof. See Appendix 2.C. □

2.4.2 Top Trading Cycles

I proceed with *TTC mechanisms* as have been proposed for the school choice problem by Abdulkadiroğlu and Sönmez (2003). For the rest of this section, let $|I| \geq 3$ and $|X| \geq 3$. Denote the *TTC* mechanism operating on $\triangleright$ with $TTC^{\triangleright}$, where the *TTC* algorithm inducing $TTC^{\triangleright}$ is described in Appendix 2.D.

As a starting point for the discussion, I first apply the results of the previous subsection to $TTC^{\triangleright}$. Specifically, since $TTC^{\triangleright}$ is efficient and group strategy-proof (Pápai, 2000) one can apply Theorem 2.3. More concretely, note that sequential dictatorships are a natural generalization of serial dictatorships and as can be easily seen, $TTC^{\triangleright}$ is only equivalent to a sequential dictatorship, if it reduces to a serial dictatorship. Thus, one must have $\triangleright_x = \triangleright_y$ for all $x, y \in X$. Summarizing these observations thus leads to the following corollary.

Corollary 2.2. *The following three statements are equivalent:*

1. $TTC^{\triangleright}$ *is a serial dictatorship.*
2. $TTC^{\triangleright}$ *is fully-transparent.*
3. $TTC^{\triangleright}$ *is efficient-transparent.*

Next, I provide an example which shows that there exist priorities $\triangleright$ under which $TTC^{\triangleright}$ is not cardinality-transparent. To justify why an authority could be interested in a deviation as I construct in the example, consider a school choice setting where students expressed their desire for an efficient mechanism. It is well known from Balinski and Sönmez (1999) that efficiency is not generally compatible with stability and while stable mechanisms have proven successful in matching markets for decades, their unstable counterparts have often been replaced (Roth, 2002). Thus, a school choice authority may have conflicting goals if, on the one hand, it views stability as a desirable long-term goal, but on the other hand, it wants to give students the impression that their desire for efficiency is being taken into account. However, as shown in the example, the authority can accommodate this tension by first responding to participants' desire to announce an efficient mechanism, but then occasionally deviating to a stable matching when it is safe to do so. In fact, as the example shows, the stable matching deviated to assigns a different number of agents compared to what would follow from $TTC^{\rhd}$.

Example 2.2. There are four agents $I = \{i, j, k, l\}$ and four objects $X = \{w, x, y, z\}$. Relevant types for agents and priorities $\rhd$ are given in the following table, whereas the priorities for object z are specified arbitrarily:

P_i	P'_i	P_j	P_k	P'_k	P_l
w	w	x	x	y	w
y	$\emptyset$	z	y	$\emptyset$	x
$\emptyset$	y	$\emptyset$	$\emptyset$	x	y
x	x	y	z	z	z
z	z	w	w	w	$\emptyset$

$\rhd_w$	$\rhd_x$	$\rhd_y$
l	i	k
i	j	i
j	k	j
k	l	l

Suppose the authority announces $TTC^{\rhd}$. Under type profile P, the final matching produced by $TTC^{\rhd}$ is

$$TTC^{\rhd}(P) = \{(i, y), (j, z), (k, x), (l, w)\}.$$

Note that $TTC^{\rhd}(P)$ is not $\rhd$-stable since j blocks the matching at object x. Assume that the authority deviates to a stable matching under type profile P as motivated in the preface of this example. Specifically, consider deviation $\tilde{g}$ for which

$$\tilde{g}(P) = \{(i, \emptyset), (j, x), (k, y), (l, w)\}$$

and

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

Note that the deviation $\tilde{g}$ is not cardinality-equivalent since $\tilde{g}_I(P) < g_I(P)$. In fact, given type profile P, the authority refuses to execute a trading cycle between agent i and agent k. Instead, the authority enforces i to be matched with the outside option.

I now argue that deviation $\tilde{g}$ is safe, starting with innocent explanations for each agent's observation under type profile P:

- For agent i, type profile (P_i, P_j, P'_k, P_l) provides an innocent explanation for $o_i(\tilde{g}(P))$.
- For agent j, type profile (P'_i, P_j, P_k, P_l) provides an innocent explanation for $o_j(\tilde{g}(P))$.
- For agent k, type profile (P'_i, P_j, P_k, P_l) provides an innocent explanation for $o_k(\tilde{g}(P))$.
- For agent l, type profile (P'_i, P_j, P_k, P_l) provides an innocent explanation for $o_l(\tilde{g}(P))$.

For agent j, k and l the innocent explanation is that i finds only w acceptable. Specifically, once i is assigned to the outside option in the second step of the TTC algorithm (l left with w in the first step), agent j is pointed by x. Moreover, agent i can explain her observation with k having top choice y. Finally, note that except for type profile P, all observations produced under the deviation are the same as those under the announcement. Hence $\tilde{g}$ is safe and therefore $TTC^{\rhd}$ is not cardinality-transparent.

□

It turns out that a similar safe deviation as constructed in the example is always possible if the priority structure contains a cycle of the following type:

Definition 2.7 (Ergin (2002))**.** An *Ergin-cycle* consists of three agents $i, j, k \in I$ and two objects $x, y \in X$ such that $i \rhd_x j \rhd_x k \rhd_y i$. A priority structure $\rhd$ is *Ergin-acyclic* if there exists no Ergin-cycle in $\rhd$.

The result by Ergin (2002) presented next relates the acyclicity condition to $\rhd$-stability of $TTC^{\rhd}$.

Theorem (Ergin (2002)). $TTC^{\triangleright}$ is $\triangleright$-stable if and only if $\triangleright$ is Ergin-acyclic.

We are ready for the final result of this paper.

Proposition 2.3. *The following three statements are equivalent:*

1. $TTC^{\triangleright}$ *is* $\triangleright$*-stable.*

2. $TTC^{\triangleright}$ *is identity-transparent.*

3. $TTC^{\triangleright}$ *is cardinality-transparent.*

Proof. The sufficiency parts of the statements follow from applying Corollary 2.1 to $TTC^{\triangleright}$. The proofs for the necessity parts of the statements can be found in Appendix 2.E. □

2.5 Conclusion

Using newly developed transparency criteria as commitment devices, I studied to which extent a central authority can commit not to deviate from an announced matching mechanism. I established that *DA* provides full commitment to stability, whereas among efficient and group strategy-proof mechanisms only sequential dictatorships guarantee full commitment and commitment to efficiency. I also showed that in case of *DA* the authority can commit not to change identities (and the number) of matched agents, while *TTC* provides the same guarantee if and only if *TTC* is stable. However, for both candidates full-transparency is satisfied if and only if they reduce to a serial dictatorship.

As an open question for future research it would be interesting to see if the characterization of efficient-transparent mechanisms for group strategy-proof and efficient mechanisms can be extended to the more general class of all strategy-proof and efficient mechanisms.

2.A Proof of Proposition 2.2

($\Leftarrow$) If $DA^{\rhd}$ is a serial dictatorship, then $\rhd_x = \rhd_y$, for all $x, y \in X$. Given any $P \in \mathcal{P}$, following the ordering of $\rhd_x$ 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.[11] Hence for each $P \in \mathcal{P}$ it is clear that $\Sigma^{\rhd}(P)$ is a singleton. Therefore, Theorem 2.1 implies that there exists no safe deviation from $DA^{\rhd}$ and thus $DA^{\rhd}$ is fully-transparent.

($\Rightarrow$) Suppose that $DA^{\rhd}$ is not a serial dictatorship. By definition, there exist objects $x, y \in X$ such that $\rhd_x \neq \rhd_y$. This implies that there exist two agents $i, j \in I$ such that $i \rhd_y j$ and $j \rhd_x i$. Denote $I' = I \setminus \{i, j\}$ and let type 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. Moreover, consider the following relevant types for agents i and j.

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

- $x \, P_i \, y$ and for all $x' \in X \cup \{\emptyset\} \setminus \{x, y\}$: $y \, P_i \, x'$ and
- $y \, P'_i \, x$ and for all $x' \in X \cup \{\emptyset\} \setminus \{x, y\}$: $x \, P'_i \, x'$.

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

- $y \, P_j \, x$ and for all $x' \in X \cup \{\emptyset\} \setminus \{x, y\}$: $x \, P_j \, x'$ and
- $x \, P'_j \, y$ and for all $x' \in X \cup \{\emptyset\} \setminus \{x, y\}$: $y \, P'_j \, x'$.

Next, I construct a safe deviation $\tilde{g} \in \mathcal{G}$ from $DA^{\rhd}$. 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 $k \in I', \tilde{g}_k(P) = \emptyset$. Moreover, be $\tilde{g}$ such that

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

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

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

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

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

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 $DA^{\triangleright}$ is not fully-transparent.

2.B Trading Cycles and Characterizations of Group Strategy-Proofness

In this section, I introduce *Trading Cycles (TC) Mechanisms* (Pycia and Ünver, 2017) together with the main characterization of group strategy-proof and Pareto efficient mechanisms.[12] Moreover, I provide an additional characterization of group strategy-proof mechanisms by Pápai (2000) that will be useful for the proofs of Theorem 2.2 and Theorem 2.3.

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

[12] 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). The characterization of group strategy-proof and Pareto efficient mechanisms presented at the end of this section extends to the setting with outside options according to (Pycia and Ünver, 2017, Supplement, p.6).

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 2.8. 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 2.3. The interested reader is kindly referred to an excellent discussion and interpretation of these conditions in Pycia and Ünver (2017).

Definition 2.9. 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.

Moreover, for any two submatchings $\sigma, \sigma' \in \hat{\mathcal{M}}$ such that $\sigma \subset \sigma'$ with an agent $i \in \hat{I}_{\sigma'}$ who owns an object $x \in \hat{X}_{\sigma'}$ at σ it holds that

$(C4)$ agent i owns x at σ'.

$(C5)$ if i' brokers object x' at σ and $i' \in \hat{I}_{\sigma'}, x' \in \hat{X}_{\sigma'}$, then either i' brokers x' at σ' or i owns x' at σ'.

$(C6)$ If agent $i' \in \hat{I}_{\sigma'}$ controls $x' \in \hat{X}_{\sigma'}$ at σ, then i' owns x at $\sigma \cup \{(i, x')\}$.

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.

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 $c_{\sigma^{t-1}}(x)$. Let the outside option point to each agent in $i \in \hat{I}_{\sigma^{t-1}}$. 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. If there is a broker in $\hat{I}_{\sigma^{t-1}}$, then she is pointing to her most preferred object among all objects that are owned. For the remaining agents, let each agent $i \in \hat{I}_{\sigma^{t-1}}$ point to her top choice x among objects $\hat{X}_{\sigma^{t-1}} \cup \{\emptyset\}$.

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 $l = 1, \ldots, n$ 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.

A cycle exists in each step until each agent is matched and the number of steps is finite. Moreover, no pair of cycles intersects and there is at least one pair matched at each step.

The proof of Theorem 2.3 presented in Appendix 2.C builds on the following result.

Theorem (Pycia and Ünver (2017)). A mechanism $g \in \mathcal{G}$ 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) which requires that there is no agent who can change other agents' assignments by misreporting her type, without changing her own assignment. Formally, a mechanism $g \in \mathcal{G}$ 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 2.2 (Pápai (2000)). *A mechanism is group strategy-proof if and only if it is strategy-proof and non-bossy.*

2.C Proofs of Theorem 2.2 and Theorem 2.3

This section contains all results needed to obtain Theorem 2.2 and Theorem 2.3. Specifically, Lemma 2.3 presented first, implies the sufficiency parts of the statements. Necessity for Theorem 2.2 follows from applying Lemma 2.4 and Lemma 2.5, whereas the necessity parts of Theorem 2.3 only require Lemma 2.4.

Lemma 2.3. *If announcement $g \in \mathcal{SP}$ is a sequential dictatorship, then g is fully-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 type 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 2.6 (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 2.6 implies that $\pi_P^{-1}(i) = \pi_{(P_i,\tilde{P}_{-i})}^{-1}(i)$, 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 efficient-transparent. □

I now turn to two additional lemmas for the necessity parts of Theorem 2.2 and Theorem 2.3. 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 2.4. *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. By Pycia and Ünver (2017) there exists a TC mechanism $TC^{(c,b)}$ with control right structure $(c, b) \in \mathcal{C}$, which is equivalent to the announced mechanism g (See Appendix 2.B). Thus, it is sufficient to show that $TC^{(c,b)}$ is not efficient-transparent, i.e. there exists a non-efficient safe deviation from $TC^{(c,b)}$.

As has been shown by Pycia and Ünver (2017) (Theorem 6), given (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. Moreover, if at control right structure (c, b) there is a single owner at each submatching then, as can be easily shown, $TC^{(c,b)}$ is equivalent to a sequential dictatorship as defined by Definition 2.6. Conversely, since $TC^{(c,b)}$ is not equivalent to a sequential dictatorship, there must exist a submatching $\sigma \in \hat{\mathcal{M}}$ which has at least two owners.

Next, because the set of submatchings $\hat{\mathcal{M}}$ is finite, we can pick a smallest submatching $\sigma^* \in \hat{\mathcal{M}}$ with at least two owners. Let $P_{\sigma^*_I}$ be such that for each $k \in \sigma^*_I$, $\sigma^*(k)$ is k's top choice under P_k. Note 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 $\sigma^{|\sigma^*_I|} = \sigma^*$ in Step $|\sigma^*_I| + 1$. Moreover, note that any agent $k \in \sigma^*_I$ is matched to an object she owns at the step she is assigned or she is assigned to the outside option.

Now consider Step $|\sigma^*_I| + 1$ and let $i, j \in \hat{I}_{\sigma^*}$ denote two arbitrary owners at σ^*. Assume that i owns object $x \in \hat{X}_{\sigma^*}$ and j owns object $y \in \hat{X}_{\sigma^*}$.

The following types of agent i and j will be central. Let $P_i, P'_i \in \mathcal{P}_i$ be described by

- $y P_i x$ and for all $x' \in X \cup \{\emptyset\} \setminus \{x, y\}$: $x \, P_i \, x'$ and
- $x P'_i y$ and for all $x' \in X \cup \{\emptyset\} \setminus \{x, y\}$: $y \, P'_i \, x'$.

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

- $x \, P_j \, y$ and for all $x' \in X \cup \{\emptyset\} \setminus \{x, y\}$: $y \, P_j \, x'$ and
- $y \, P'_j \, x$ and for all $x' \in X \cup \{\emptyset\} \setminus \{x, y\}$: $x \, P'_j \, x'$.

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 $\sigma^{|\sigma_I^*|} = \sigma^*$ in Step $|\sigma_I^*| + 1$. 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 type profile $(P_{\sigma_I^*}, P_i, P_j, P_K)$ at Step $|\sigma_I^*| + 1$, 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 $|\sigma_I^*| + 1$, 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, for all $\tilde{P} \in \mathcal{P} \setminus (P_{\sigma_I^*}, P_i, P_j, P_K)$, since $\tilde{g}(\tilde{P}) = TC^{(c,b)}(\tilde{P})$, innocent explanations for the corresponding 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

$$o_i(TC^{(c,b)}(P_{\sigma_I^*}, P_i, P_j', P_K)) = o_i(\tilde{g}(P_{\sigma_I^*}, P_i, P_j, P_K)), \tag{2.1}$$
$$o_j(TC^{(c,b)}(P_{\sigma_I^*}, P_i', P_j, P_K)) = o_j(\tilde{g}(P_{\sigma_I^*}, P_i, P_j, P_K)). \tag{2.2}$$

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 $|\sigma_I^*| + 1$. Now consider Step $|\sigma_I^*| + 1$ 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 $\sigma^* = \sigma^{|\sigma_I^*|}$ and it is her favorite choice among the remaining objects according to P_i.[13] Symmetrically, at Step $|\sigma_I^*| + 1$ 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 $\sigma^* = \sigma^{|\sigma_I^*|}$ 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 2.2), it must be true that, for all $k \in K$, we have that

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

and as such conditions (1) and (2) are satisfied and each agent has an innocent explanation for any of her observations under $\tilde{g}$. We conclude that $\tilde{g}$ is safe and that $TC^{(c,b)}$ is not efficient-transparent. Since g is equivalent to $TC^{(c,b)}$, the same conclusion holds for announcement g. This completes the proof. □

To complete the proof, I next establish that each efficient mechanism allows safe deviations if it is

[13] 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.

strategy-proof but not group strategy-proof.

Lemma 2.5. *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 2.2. 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 we have $g(P_i, P_{-i}) \neq g(P'_i, P_{-i})$ 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 it is true that $g(P^*_i, P_{-i}) \neq g(P'_i, P_{-i})$ 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 $\tilde{g}$, where we have $\tilde{g}(P^*_i, P_{-i}) = g(P_i, P_{-i})$ and for all $\tilde{P} \in \mathcal{P} \backslash \{(P^*_i, P_{-i})\}$, $\tilde{g}(\tilde{P}) = g(\tilde{P})$. Since all observations under g and $\tilde{g}$ coincide except under type 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 type 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 type 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, under type 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. $\square$

2.D Algorithms

2.D.1 The (Agent-proposing) Deferred Acceptance Algorithm

Given any $\rhd \in \bar{Pr}$, the outcome of $DA^{\rhd}$ given any input profile $P \in \mathcal{P}$ is calculated as follows:

Step 1 Each agent $i \in I$ proposes to her most preferred object in $X \cup \{\emptyset\}$. Each object $x \in X$ considers all the proposals and tentatively accepts the candidate who applies to x and has the highest ranking on $\rhd_x$. The remaining proposals are rejected. Moreover, all agents that propose to the outside option $\emptyset$ are accepted.

Step k, $k \geq 2$ Each agent who was rejected at step $k-1$ applies to her most preferred object not yet applied to. Each object $x \in X$ considers all the new applicants including the agent tentatively assigned to it at step $k-1$. Among all applicants the object x accepts the highest ranked applicant according to $\rhd_x$. The remaining proposals are rejected. Moreover, all agents that propose to the outside option $\emptyset$ are accepted.

The algorithm terminates with the tentative assignments of the first step in which no agent is rejected.

2.D.2 The Top Trading Cycles Algorithm

Given any profile $P \in \mathcal{P}$, the outcome of $TTC^{\rhd}$ is calculated via the following algorithm:

Step 1 Each agent $i \in I$ points to her most preferred object $x \in X \cup \{\emptyset\}$ according to P_i, and each object $x \in X$ points to the agent who has the highest priority according to $\rhd_x$. There exists at least one cycle. Each agent in a cycle is assigned to the object she is pointing to and each such pair is removed. Each agent that points to $\emptyset$ is assigned to $\emptyset$ and is removed. If there is an agent left unmatched, move to Step 2; otherwise, the algorithm terminates.

Step $k \geq 2$ Each agent that has not been removed in a previous step points to her most preferred object according to P_i among those still unallocated. Each object $x \in X$ points to the agent who has the highest priority according to $\rhd_x$ among all agents that unmatched. There exists at least one cycle. Each agent in a cycle is assigned to the object she is pointing to and each such agent is removed with her object. Each agent that points to $\emptyset$ is assigned to $\emptyset$ and is removed. If any agents are left unmatched, move to the next step; otherwise, the algorithm terminates.

The algorithm terminates in a finite number of steps, whereas the outcome of the algorithm is the union of all assigned pairs at each step.

2.E Proof of Proposition 2.3

Proof. If $TTC^{\rhd}$ is stable, then we know that $TTC^{\rhd}$ and $DA^{\rhd}$ are equivalent. Thus, Corollary 2.1 implies that $TTC^{\rhd}$ is cardinality-transparent and identity-transparent.

To prove necessity of both parts of the statement, assume $\rhd$ such that $TTC^{\rhd}$ is not $\rhd$-stable. Thus, using the characterization by Ergin (2002), we know that $\rhd \in \bar{Pr}$ is not Ergin-acyclic. I aim to construct a safe deviation that is not cardinality-equivalent (i.e. also not identity-equivalent). First, since $\rhd$ is not Ergin-acyclic, there exist objects $x, y \in X$ and agents $i, j, k \in I$ such that $i \rhd_x j \rhd_x k \rhd_y i$.

Note that we either have $(C1')$ $k \rhd_y j$ or $(C2')$ $j \rhd_y k$. Thus, in both cases there exist three agents $\{l, m, n\} \in I$ and two objects $\{a, b\} \in X$ such that $l \rhd_a m \rhd_a n$ and $n \rhd_b l, m$:

- If condition $(C1')$ is satisfied, then for agents label $i = l, j = m$ and $k = n$ and for objects label $a = x, b = y$.
- If condition $(C2')$ is satisfied, then for agents label $j = l$, $k = m$ and $i = n$ and for objects label $a = y, b = x$.

In the following, suppose that either $(C1')$ or $(C2')$ are satisfied. Moreover, recall that $|X| \geq 3$, and hence there exists a third object $c \in X \backslash \{a, b\}$. I first describe a set of relevant types for agents $\{l, m, n\}$. For agent l, let $P_l, P'_l \in \mathcal{P}_l$ satisfy:

- $bP_l\emptyset$ and for all $x' \in X \setminus \{b\}$: $\emptyset\, P_l\, x'$.
- for all $x' \in X$, we have $\emptyset\, P'_l\, x'$.

For agent m, let $P_m \in \mathcal{P}_m$ be such that $a\, P_m\, c$ and for all $x' \in X \cup \{\emptyset\} \setminus \{a, c\}$, $c\, P_m\, x'$.

For agent n the types are P_n, P'_n are

- $a\, P_n\, b$ and for all $x' \in X \cup \{\emptyset\} \setminus \{a, b\}$, let $b\, P_n\, x'$;
- for all $x' \in X \cup \{\emptyset\} \setminus \{b\}$, let $b\, P'_n\, x'$.

For the remaining agents, denote $I' \equiv I \setminus \{l, m, n\}$ and let type profile $P_{I'} \in \mathcal{P}_{I'}$ be such that for each $o \in I'$ and all $x' \in X$, we have $\emptyset\, P_o\, x'$.

I now construct a candidate deviation $\tilde{g}$ that is not cardinality-equivalent. In the following, one can use a motivation for deviation $\tilde{g}$ as was for instance described in the preface of Example 2.2.[14]

Specifically, let $\tilde{g}$ be defined as follows: for profile $P = (P_l, P_m, P_n, P_{I'})$ suppose that $\tilde{g}(P)$ yields $\tilde{g}_l(P) = \emptyset, \tilde{g}_m(P) = a, \tilde{g}_n(P) = b$, and for all $o \in I', \tilde{g}_o(P) = \emptyset$. Moreover, be $\tilde{g}$ such that

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

It is easily checked that the TTC algorithm yields $TTC^{\triangleright}_l(P) = b$, $TTC^{\triangleright}_m(P) = c$, $TTC^{\triangleright}_n(P) = a$ and for all $o \in I'$, $TTC^{\triangleright}_o(P) = \emptyset$. In fact, note that whereas $TTC^{\triangleright}(P)$ is not $\triangleright$-stable since agent m blocks the matching at object a, $\tilde{g}(P)$ does indeed satisfy $\triangleright$-stability. Moreover, note that $\tilde{g}$ is not

[14] That is, one can imagine an authority conflicted between satisfying participants' preferences for an efficient mechanism through announcing a TTC mechanism on the one hand and supporting the long-term stability of the allocation process by inducing stable matchings on the other hand.

cardinality-equivalent (and not identity-equivalent) to $TTC^{\triangleright}$ since

$$|\tilde{g}_I(P)| < |TTC^{\triangleright}_I(P)|.$$

It remains to be shown that each observation possibly made under the deviation $\tilde{g}$ has an innocent explanation for the observing agent. Except for type 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.

To complete the proof, we need for each $i' \in I$ an innocent explanation for her observation $o_{i'}(\tilde{g}(P))$. Specifically, for agent l consider the types $(P_m, P'_n, P_{I'})$, for agent m consider $(P_l, P'_n, P_{I'})$ and for agent n consider $(P'_l, P_m, P_{I'})$. Last, for agents in I', consider for example the types (P_l, P_m, P'_n). It is easily checked that we obtain

$$TTC^{\triangleright}(P_l, P_m, P'_n, P_{I'}) = TTC^{\triangleright}(P'_l, P_m, P_n, P_{I'}) = \tilde{g}(P).$$

from which one can see that for each agent $i' \in I$, the observation $o_{i'}(\tilde{g}(P))$ has an innocent explanation.

Thus, $\tilde{g}$ is a safe deviation which is not cardinality-equivalent (and thus not identity-equivalent). Consequently, $TTC^{\triangleright}$ is not cardinality-transparent and not identity-transparent. This completes the proof. □

3

Regret-Free Truth-Telling in School Choice with Consent*

The *Efficiency Adjusted Deferred Acceptance Matching Rule (EDA)* is a promising candidate mechanism for public school assignment. A potential drawback of EDA is that it could encourage students to game the system since it is not strategy-proof. However, to successfully strategize, students typically need information that is unlikely to be available to them in practice. We model school choice under

*This chapter is based on Chen and Möller (2021). We thank especially our advisor, Alexander Westkamp. We are grateful to Christoph Schottmüller, Marcelo Ariel Fernandez, Kevin Breuer and Marius Gramb for helpful comments. All errors remain our own.

incomplete information and show that EDA is regret-free truth-telling, which is a weaker incentive property than strategy-proofness and was introduced by Fernandez (2020). We also show that there is no efficient matching rule that Pareto dominates a stable matching rule and is regret-free truth-telling.

3.1 Introduction

Efficiency and fairness are incompatible in the school choice problem.[1] The *Efficiency Adjusted Deferred Acceptance Rule (EDA)* (Kesten, 2010) elegantly circumvents this incompatibility by allowing students to give their consent to relax the fairness constraint. However, no compromise solution, including EDA, is strategy-proof (Abdulkadiroğlu et al., 2009).[2,3] We study whether EDA satisfies an incentive criterion by Fernandez (2020) which is weaker than strategy-proofness and is based on participants' wish to avoid regret.

We employ the many-to-one school choice model with consent (Kesten, 2010) under incomplete information. Students can reconsider their admission chances for alternative reports, through an observational structure that is based on the cutoff terminology. We express schools' priorities in the form of scores and for each school, the cutoff is the lowest score among all students that have been admitted to that school. Once the final matching has been determined, each student observes which student is assigned to which school and each school's cutoff. Based on her observation, a student can then draw inferences about *plausible scenarios*—pairs of underlying scores of schools and reports of other students that are consistent with the observation. We motivate our model through features common in the context of public school assignment. In practice, matching rules often use scores based on prox-

[1] A student has justified envy at a matching, if there exists a lower prioritized student assigned to a school and the corresponding school is preferred to her assignment (Abdulkadiroğlu and Sönmez, 2003). A matching is *fair* if no justified envy exists and a matching rule is fair if it only produces matchings which are fair. The trade-off between efficiency and fairness follows from Balinski and Sönmez (1999).

[2] Strategy-proofness requires that it is a weakly dominant strategy for students to report their true preferences.

[3] For related results, see also Erdil and Ergin (2008) and Alva and Manjunath (2019).

imity, walk-zone areas, sibling-status and other socioeconomic variables. The composition of scores is usually public information, whereas accurate information on other students' scores and reported preferences will generally be covered by privacy protection. Moreover, students typically receive feedback on the market outcome and cutoffs.

In this model, we adopt the incentive notion by Fernandez (2020). Specifically, a student *regrets* a report through an alternative report, once she finds her submitted report to be dominated by the alternative in any plausible scenario. A rule is *regret-free truth-telling* if no student would regret reporting her preferences truthfully.

The main finding of this chapter is that EDA is *regret-free truth-telling* (Theorem 3.1). Moreover, we show that under EDA, truth-telling is the *unique* option which never leads to regret (Proposition 3.2). Concretely, we show that for any misreport, there exists an observation such that the student regrets the misreport through her true preferences. Our last result concerns matching rules which Pareto dominate a stable matching rule.[4] A stable dominating rule always implements a matching that weakly Pareto dominates a stable matching (Alva and Manjunath, 2019). It is well known that all stable dominating rules, except the well known *Deferred Acceptance Matching Rule (DA)* (Gale and Shapley, 1962), are not strategy-proof (Abdulkadiroğlu et al., 2009).[5] We show that among the *efficient stable dominating rules* no matching rule is regret-free truth-telling (Theorem 3.2). Note that the original formulation of EDA considered in this chapter is not Pareto efficient since EDA respects improvements on efficiency only with students' consents for being exposed to justified envy.

All our results extend to the case where the students only observe their own assignment and the cutoffs. By showing that truth-telling is the unique regret-free strategy, we provide an appropriate statement for the intuition that truth-telling may be a focal strategy under EDA. Thus, our work

[4] A matching rule is *stable* if it produces outcomes which are *fair*, *individually rational* and *non-wasteful*. A matching is non-wasteful if there is no object that is unassigned although there is an agent that prefers it over her assignment. A matching is individually rational if no agent prefers her outside option over her final assignment.

[5] See also Erdil and Ergin (2008), Kesten (2010) and Alva and Manjunath (2019).

contributes to the strand of literature that outlines the many desirable features of EDA for practical implementation.

Related literature

To our knowledge, Fernandez (2020) is the first to introduce regret-based incentives in the matching literature.[6] In marriage markets, Fernandez (2020) shows that truth-telling is the unique regret-free strategy under DA for both men and women and that DA is the unique regret-free truth-telling rule among so-called quantile stable rules.[7] Fernandez (2020) sheds light on college admissions problems. He shows that the student-proposing variant of DA is regret-free truth-telling. However, under the college-proposing variant of DA, being truthful does not need to be free of regret for colleges. The key differences of our work to that of Fernandez (2020) is that only the student market side is strategic. Moreover, whereas in Fernandez (2020) participants only observe the realized matching, students in our model additionally observe cutoffs.

This chapter mainly contributes to the literature that deepens the understanding of EDA's incentive properties. Our results complement those of Troyan and Morrill (2020), who show that for cognitively limited participants beneficial misreporting under EDA is not *obvious* in the following sense: a profitable misreport is an *obvious manipulation* if the best-case outcome of the misreport is better than the best-case outcome of telling the truth or, if the worst-case outcome of the misreport is better than the worst-case outcome of telling the truth. The main difference between our work and that of Troyan and Morrill (2020) concerns the source of uncertainty that students face. A profitable

[6]Regret-based incentives have a long tradition in economic theory. For instance, in auction theory, regret-based incentives of bidders in first-price auctions have been studied by Filiz-Ozbay and Ozbay (2007) and Engelbrecht-Wiggans (1989). For a more detailed discussion we refer to Fernandez (2020). See Gilovich and Medvec (1995) and Zeelenberg and Pieters (2007) for psychological treatments of regret.

[7]Given any $q \in (0, 1]$, the q-quantile stable rule selects the $[qk]$ best stable school for each student given any report, where k is the number of stable matchings under this report. For more information on quantile stable mechanisms, we refer to Teo and Sethuraman (1998), Klaus and Klijn (2006), or Chen et al. (2015).

misreport is obvious if it is easy to recognize for students whose knowledge on the matching rule is imperfect, given that these students have full access to the scores of other students. That is, non-obvious manipulability is mainly driven by participants' limited understanding of the matching rule. By contrast, students in our model know how the matching rule works and our results are driven by students' incomplete access to the scores of other students. Notably, the positive result of Troyan and Morrill (2020) covers both EDA and stable dominating rules, where we reach a negative result for efficient stable dominating rules.

Previous results on EDA's incentive properties are inspired by the theoretical benchmark for low information environments from Roth and Rothblum (1999) and Ehlers (2008). Kesten (2010) studies Bayesian incentives of EDA in a setting where it is common knowledge that students' preferences over schools are ordered into shared quality classes and students' beliefs on how other students order schools within each quality class are symmetrically distributed. Kesten (2010) shows that if other students submit their true preferences, then truth-telling stochastically dominates any other strategy. The key difference to our model is that we do not specify any prior probability distribution regarding the beliefs or distribution on other participants' preferences and thus do not impose any symmetry assumptions or correlation of preferences over schools. Thus, in contrast to the approach of Kesten (2010) our information environment follows the 'Wilson doctrine' (Wilson, 1987).

The literature that is concerned with other theoretical properties of EDA is rapidly growing. Tang and Yu (2014), Ehlers and Morrill (2020), Bando (2014) and Dur et al. (2019) recently developed tractable alternatives to Kesten's initial formulation of EDA. Ehlers and Morrill (2020) generalize EDA to a school choice model where school priorities take the form of more flexible choice functions and Kwon and Shorrer (2020) propose a version of EDA for organ exchange.

Our work also relates to the line of literature that uses the cutoff terminology in school choice models. Most prominent in this regard is Azevedo and Leshno (2016) who characterize stable matchings in terms of cutoffs in a continuum school choice model. They show that cutoffs take the form of marketclearing prices that equalize supply and demand and can be used to perform comparative statics with respect to schools' incentives to invest in quality. When used to characterize stable matchings, cutoffs usually take the form of a guarantee for participants to be admitted at schools. In our framework, final assignments may not correspond to stable matchings. Therefore, the cutoffs do not necessarily provide a student with information about whether she will be admitted at a desired school. Moreover, in our model the cutoffs are incorporated into students' strategic reasoning.

The rest of this chapter is organized as follows. We introduce the basic model and EDA in Section 3.2. We model the informational environment and adopt regret-free truth-telling in Section 3.3. In Section 3.4, we present our main results. Our analysis regarding efficient stable dominating rules is provided in Section 3.5. Finally, Section 3.6 gives a short conclusion. The Appendix contains most of our proofs.

3.2 Model

There is a finite set of students I and a finite set of schools S. Each school $s \in S$ has a fixed capacity q_s and we collect the capacities in $q = (q_s)_{s \in S}$. We add a common outside option $s_\emptyset$ for students which has infinite capacity.

Each school $s \in S$ has a set of scores $g^s = \{g_i^s\}_{i \in I}$, where $g_i^s \in (0, 1)$ is i's score at s. We assume that $g_i^s \neq g_j^s$ for any $i, j \in I$ and any $s \in S$, and we say that for each pair of students $i, j \in I$, i has higher priority at s than j if and only if $g_i^s > g_j^s$. That is, for each school s, the school's scores induce a strict priority ranking over I.[8] For each $i \in I$, let $g_i = \{g_i^s\}_{s \in S}$ be the set of scores assigned to student i. Let a score structure $g = (g_i)_{i \in I}$ be a collection of scores for each student and let $g_{-i} = (g_j)_{j \in I \setminus \{i\}}$ be a collection of scores for students in $I \setminus \{i\}$. Moreover, set $\mathcal{G}_I$ as the domain of all possible score

[8] The incomplete information framework we introduce in Section 3.3 allows students to draw inferences about their admission chances. Our formulation of scores will then ensure that a student typically cannot infer her exact rank on a school's priority list just on the basis of her own score.

structures and $\mathcal{G}_{-i}$ as the domain of all score structures for students other than i.

For each student $i \in I$, let $\succ_i$ be a strict preference relation over $S \cup \{s_\emptyset\}$. The corresponding weak preference relation of $\succ_i$ is denoted by $\succeq_i$.[9] Let $\mathcal{P}$ denote the set of all possible strict preference relations over $S \cup \{s_\emptyset\}$. For any $\succ_i \in \mathcal{P}$, a school s is acceptable to i if $s \succ_i s_\emptyset$ and unacceptable if it is not acceptable. A preference profile $\succ = (\succ_i)_{i \in I}$ is a realization of $\mathcal{P}$ for each $i \in I$ and $\succ_{-i} = (\succ_j)_{j \in I \setminus \{i\}}$ is a preference profile for students in $I \setminus \{i\}$. We define $\mathcal{P}_I$ as the domain of all preference profiles and $\mathcal{P}_{-i}$ as the domain of all preference profiles for students in $I \setminus \{i\}$.

A *matching* $\mu : I \to S \cup \{s_\emptyset\}$ is a function such that for each $s \in S$, $|\mu^{-1}(s)| \leq q_s$. Given any μ, we set $\mu_i = \mu(i)$ as the assignment of i and $\mu_s = \mu^{-1}(s)$ as the set of students assigned to s. Denote the set of all possible matchings by $\mathcal{M}$.

In the following, fix any $\succ \in \mathcal{P}_I$. We say a matching μ *weakly Pareto dominates* another matching μ' if for all $i \in I$, $\mu_i \succeq_i \mu'_i$. A matching μ *Pareto dominates* μ' if μ weakly Pareto dominates μ' and for some $j \in I$, $\mu_j \succ_j \mu'_j$. A matching μ is *Pareto efficient* if there does not exist another matching μ' which Pareto dominates μ.

We now introduce two fairness notions, where we start with the well-known notion by Abdulkadiroğlu and Sönmez (2003). Given a matching μ, student i has *justified envy* towards student j at school μ_j under μ if $\mu_j \succ_i \mu_i$ and $g_i^{\mu_j} > g_j^{\mu_j}$. A matching μ is *fair* if no student has justified envy at μ. A matching μ is *individually rational* if for each student the assigned school is acceptable to her. A matching μ is *non-wasteful* if there does not exist a student i and a school s, such that $s \succ_i \mu_i$ and $|\mu_s| < q_s$. A matching μ is *stable* if it is fair, individually rational and non-wasteful.

We also consider a weaker fairness notion that was introduced by Kesten (2010). The notion takes students' willingness to consent for being exposed to justified envy into account. For each student i, the consent is parameterized by a binary variable $c_i \in \{0, 1\}$ where $c_i = 1$ means that i consents to any envy that is justified and otherwise to none. We say a matching μ *violates the priority* of student

[9] That is, for all $s, s' \in S$, $s \succeq_i s'$ if either $s \succ_i s'$ or $s = s'$.

i given c_i if $c_i = 0$ and if there exists another student $j \in I$ such that i has justified envy towards j at μ. Let $c = (c_i)_{i \in I}$ be a consent profile and let $\mathcal{C}_I$ be the domain of all consent profiles. Denote a consent profile of students other than i by $c_{-i} = (c_j)_{j \in I \setminus \{i\}}$ and the respective domain by $\mathcal{C}_{-i}$. Given a matching μ, a profile of preferences $\succ$ and a consent profile c, we say that a matching is *fair with consent* if there exists no student whose priority is violated at μ.

We call a collection $(I, S, q, g, \succ, c)$ a *school choice problem with consent* (or simply a *problem*). Throughout the main body of the chapter, we fix a problem $(I, S, q, g, \succ, c)$. A *report* of student i is pair $(\succ'_i, c'_i) \in \mathcal{P} \times \{0, 1\}$, and a report profile is described by $(\succ', c') \in \mathcal{P}_I \times \mathcal{C}_I$. Analogously, let $(\succ'_{-i}, c'_{-i}) \in \mathcal{P}_{-i} \times \mathcal{C}_{-i}$ be a report profile of students except i.

A *matching rule* $f : \mathcal{G}_I \times \mathcal{P}_I \times \mathcal{C}_I \to \mathcal{M}$ maps any triple of a score structure, preference profile and consent profile into a matching. Given a report profile $(\succ, c)$ and a score structure g, let the outcome of f be $f(g, \succ, c)$ and for each $i \in I$ let $f_i(g, \succ, c)$ denote student i's respective assignment. If the matching rule does not take consent decisions into consideration, we write $f(g, \succ)$ instead of $f(g, \succ, c)$. A matching rule f is *Pareto efficient* if each outcome of the matching rule is Pareto efficient. Similarly, a matching rule is *stable* if it produces a stable matching for any problem.

We proceed with the description of two incentive notions for students. A matching rule f is *consent-invariant* if $f_i(g, \succ, (c_i, c_{-i})) = f_i(g, \succ, (c'_i, c_{-i}))$ for all i and all c_i, c'_i. That is, each student's assignment is independent of her *own* consent decision. Note that the matching rules studied in this chapter are all consent-invariant. A matching rule f is *strategy-proof* if $f_i(g, (\succ_i, \succ_{-i}), c) \succeq_i f_i(g, (\tilde{\succ}_i, \succ_{-i}), c)$ for all i and all $\tilde{\succ}_i \in \mathcal{P}$. That means, for each student, reporting her true preferences is weakly better than reporting untruthfully regardless of other students' reports.

3.2.1 EDA

In this subsection, we present Kesten's *Efficiency Adjusted Deferred Acceptance Rule (EDA)* along with our first result. We use the *Top-Priority (TP) algorithm* (Dur et al., 2019) to calculate the outcomes of

EDA and start with some basic terminologies needed for its introduction. For the rest of this section, fix any $(\succ, c)$. For any matching $\mu \in \mathcal{M}$, any student i and any school s, we say that *i demands s at μ* if $s \succ_i \mu_i$. Moreover, we say that student i is *eligible* for s at μ if i demands s at μ and there exists no j who also demands s with $c_j = 0$ and $g_i^s < g_j^s$. In other words, the set of students eligible for s are those students who, once assigned to s, would not violate the priority of any other student at matching μ. Note that there could be more than one student who is eligible for a school and if two students i, i' are both eligible for s, then $g_i^s > g_{i'}^s$ implies $c_i = 1$.

Given a matching $\mu \in \mathcal{M}$, consider the directed graph $G(\mu) = (I, E(\mu))$, where $E(\mu) \subseteq I \times I$ is the set of (directed) edges such that $ij \in E(\mu)$ if and only if i is eligible for μ_j. A set of edges $\{i_1 i_2, i_2 i_3, ..., i_n i_{n+1}\}$ in $G(\mu)$ is a path if $i_1, i_2, ..., i_{n+1}$ are distinct and it is a cycle if $i_1, i_2, ..., i_n$ are distinct while $i_1 = i_{n+1}$.

A school s has *no demand* at μ if no student demands s at μ. A school s is *underdemanded* at μ if either it has no demand at μ or, there is no path in $G(\mu)$ that ends with some $i \in \mu_s$ which contains students who are part of a cycle in $G(\mu)$. We say that a student is *permanently matched* at μ if she is assigned to an underdemanded school at μ. Furthermore, a student is *temporarily matched* if she is not permanently matched.

Given $\mu \in \mathcal{M}$, we call $G^*(\mu) = (I, E^*(\mu))$ the *Top-priority graph* of μ and its set of edges $E^*(\mu)$ is defined as follows: we have $ij \in E^*(\mu)$ if and only if among the students who are temporarily matched at μ and are eligible for μ_j, student i has the highest score for μ_j. That is, for each $i \in I$, $E^*(\mu) \subseteq E(\mu)$ contains at most one edge pointing to i. Solving cycle $\gamma = \{i_1 i_2, i_2 i_3, ... i_n i_1\}$ in $G^*(\mu)$ is defined by the operation $\circ$ and yields matching $\nu = \gamma \circ \mu$, such that $\nu_i = \mu_j$ for each $ij \in \gamma$, and $\nu_{i'} = \mu_{i'}$ for each $i' \notin \{i_1, i_2, ..., i_n\}$.

The TP algorithm iteratively solves cycles based on the top-priority graphs, where one starts with the graph of the *Student Optimal Stable Matching (SOSM)*. The SOSM Pareto dominates all other stable matchings and can be calculated via the popular *Student-Proposing Deferred Acceptance Algo-*

rithm (DA) (Gale and Shapley, 1962) which is presented in Appendix 3.A. The TP algorithm works as follows:

Step 0**:** Calculate the SOSM and denote the matching by μ^0.

Step t, $t \geq 1$**:** Given matching μ^{t-1}:

t.1 If there is no cycle in $G^*(\mu^{t-1})$, then stop and let the final outcome be μ^{t-1}.

t.2 Otherwise, select one of the cycles in $G^*(\mu^{t-1})$, say γ^t, and let $\mu^t = \gamma^t \circ \mu^{t-1}$. Move to step $t+1$.

As has been shown in Lemma 6 of Dur et al. (2019), any cycle selection of the algorithm leads to the outcome of EDA and thus the TP algorithm induces EDA.

We now move to our discussion on EDA's incentive properties which is known to be consent-invariant but not strategy-proof (Kesten, 2010). Our first result, Proposition 3.1, states that a certain class of deviations of a student does not affect her own assignment. For any preference relation $\succ_i \in \mathcal{P}$ and school $s \in S$, let the weak lower contour set of $\succ_i$ with respect to s be $L_s^{\succ_i} = \{s' \in S \mid s \succeq_i s'\}$.

Proposition 3.1. *If $EDA(g, \succ, c) = \mu$ and $\tilde{\succ}_i \in \mathcal{P}$ is such that for all $s, s' \in L_{\mu_i}^{\succ_i}$, $s \succ_i s'$ only if $s \tilde{\succ}_i s'$, then $EDA_i(g, (\tilde{\succ}_i, \succ_{-i}), c) = \mu_i$.*

Proof. See Appendix 3.B. □

In words, Proposition 3.1 shows that if a student's deviation from her baseline report keeps the same order of the schools in the lower contour set with respect to the baseline assignment, then it yields the same outcome for the deviating student. Note that the set of deviations we consider in Proposition 3.1 is a subset of the monotonic transformations at the student's baseline assignment. Formally, $\succ_i'$ is a *monotonic transformation* of $\succ_i$ at $s \in S \cup \{s_\emptyset\}$ if $s' \succ_i' s$ implies that $s' \succ_i s$. Our main result presented in Theorem 3.1 can be used to illustrate that Proposition 3.1 does not hold for all monotonic transformations at μ_i.

3.3 Regret in school choice

In this section, we introduce the informational environment and regret-based incentives. We first describe the students' information and impose an observational structure. Assume that before submitting the report, each student i knows (I, S, q, g_i) and the matching rule f. After assignments have been determined by f, each student observes the matching and the cutoff at each school, i.e., the lowest score among all applicants matched to the school. More formally, given a report profile $(\hat{\succ}, \hat{c})$, student i observes $\mu = f(g, \hat{\succ}, \hat{c})$ and for each school $s \in S \cup \{s_\emptyset\}$, she observes $\pi_s(\mu, g) = \min_{j \in \mu_s} g_j^s$ when $|\mu_s| = q_s$ and $\pi_s(\mu, g) = 0$ otherwise. Let $\pi(\mu, g) = \{\pi_s(\mu, g)\}_{s \in S \cup \{s_\emptyset\}}$ and let an *observation* of student i be captured by $(\mu, \pi(\mu, g))$.

Next, define any triple $(\succ'_{-i}, c'_{-i}, g'_{-i}) \in \mathcal{P}_{-i} \times \mathcal{C}_{-i} \times \mathcal{G}_{-i}$ as a *scenario* for student i. If i submits $(\hat{\succ}_i, \hat{c}_i)$ and observes $(\mu, \pi(\mu, g))$, then scenario $(\succ'_{-i}, c'_{-i}, g'_{-i})$ is *plausible* if $\pi(\mu, g) = \pi(\mu, (g_i, g'_{-i}))$ and $f((g_i, g'_{-i}), (\hat{\succ}_i, \succ'_{-i}), (\hat{c}_i, c'_{-i})) = \mu$. The set of all plausible scenarios for student i is her *inference set* $\mathcal{I}(\mu, \hat{\succ}_i, \hat{c}_i)$. Moreover, for student $i \in I$ who reports $(\hat{\succ}_i, \hat{c}_i)$ to f, let

$$\mathcal{M}|_{(\hat{\succ}_i, \hat{c}_i)} = \{\mu \in \mathcal{M} \mid \exists (\succ'_{-i}, c'_{-i}) \in \mathcal{P}_{-i} \times \mathcal{C}_{-i} : f(g, (\hat{\succ}_i, \succ'_{-i}), (\hat{c}_i, c'_{-i})) = \mu\}$$

be the set of matchings that could be *observed* by student i. Note that g is fixed in $\mathcal{M}|_{(\hat{\succ}_i, \hat{c}_i)}$, since it is a primitive of the market and independent of the report profile.

Having defined our observational structure, we are ready to introduce the notions of regret and regret-free truth-telling adopted from Fernandez (2020). Recall that all matching rules we study are consent-invariant. To simplify our notation, we define regret with a fixed consent decision for the student under consideration.

Definition 3.1. Fix consent decision $\hat{c}_i$. Student i *regrets* submitting $\hat{\succ}_i$ at $\mu \in \mathcal{M}|_{(\hat{\succ}_i, \hat{c}_i)}$ through $\hat{\succ}'_i$ under f if

1. $\forall (\succ'_{-i}, c'_{-i}, g'_{-i}) \in \mathcal{I}(\mu, \hat{\succ}_i, \hat{c}_i)$: $f_i((g_i, g'_{-i}), (\hat{\succ}'_i, \succ'_{-i}), (\hat{c}_i, c'_{-i})) \succeq_i \mu_i$

2. $\exists(\tilde{\succ}_{-i}, \tilde{c}_{-i}, \tilde{g}_{-i}) \in \mathcal{I}(\mu, \hat{\succ}_i, \hat{c}_i)$: $f_i((g_i, \tilde{g}_{-i}), (\hat{\succ}'_i, \tilde{\succ}_{-i}), (\hat{c}_i, \tilde{c}_{-i})) \succ_i \mu_i$.

In words, a student regrets her report at an observation if there is an alternative report which guarantees her a weakly better assignment in all plausible scenarios and realizes a strict improvement in at least one plausible scenario.

Definition 3.2. Fix consent decision $\hat{c}_i$. A report $\hat{\succ}_i$ is *regret-free* under f if there does not exist a pair $(\mu, \hat{\succ}'_i) \in \mathcal{M}|_{(\hat{\succ}_i, \hat{c}_i)} \times \mathcal{P}$ such that i regrets $\hat{\succ}_i$ at μ through $\hat{\succ}'_i$.

That is, a regret-free report ensures that regardless of the realized observation, the student does not regret her report.

In this chapter, we only consider matching rules that are invariant in the unacceptable set and define reports as truth-telling if the report differs from a student's true preferences only in the order within the unacceptable set. Formally, let $A_i(\succ_i) = \{s \in S | s \succ_i s_\emptyset\}$ collect all acceptable schools and let $U_i(\succ_i) = S \setminus A_i(\succ_i)$ collect all unacceptable schools. Furthermore, let

$$T_i(\succ_i) = \{\succ'_i \in \mathcal{P} \mid A_i(\succ'_i) = A_i(\succ_i) \text{ and } s \succ'_i s' \Leftrightarrow s \succ_i s',\ \forall s, s' \in A_i(\succ_i) \cup \{s_\emptyset\}\}$$

be the set of preferences which differ from $\succ_i$ by only allowing for permutations in $U_i(\succ_i)$. We say that for any i and her true preferences $\succ_i$, a report $\succ'_i \in \mathcal{P}$ is *truth-telling* if $\succ'_i \in T_i(\succ_i)$.

Definition 3.3. A matching rule f is *regret-free truth-telling* if for each problem and for each student, truth-telling is regret-free under f.

Strategy-proofness is stronger than regret-free truth-telling. That is, once truth-telling is weakly dominant under a matching rule, it must also be regret-free. However, the converse is not true. Specifically, strategy-proofness means that truth-telling is the weakly best option under *any* scenario, whereas regret-freeness only requires that, given a students' observation, no alternative report weakly dominates the truth under all *plausible scenarios*.

3.4 Main results

In this section, we present our main result. We show that a student can avoid regret under EDA if she submits her true preferences (Theorem 3.1) and that there is no other reporting behavior that provides the same guarantee (Proposition 3.2). As will be apparent from the corresponding proofs, all our results hold under the assumption that each student can only observe her own assignment and the cutoffs.

Theorem 3.1. *EDA is regret-free truth-telling.*

Proof. See Appendix 3.C. □

The following exposition provides an overview of the main arguments used in the formal proof. Fix any student $i \in I$, suppose that she reports her true preferences $\succ_i$ and she observes $(\mu, \pi(\mu, g))$. Then, any misreport $\tilde{\succ}_i$ can be interpreted as a combination of the following types of permutations, where relative to $\succ_i$:

(A1) for all $s, s' \in S$, $s \succ_i s'$ and $s' \tilde{\succ}_i s$ only if $s \in S \setminus L_{\mu_i}^{\succ_i}$;

(A2) there exists $s' \in S$ such that $\mu_i \succ_i s'$ and $s' \tilde{\succ}_i \mu_i$, or;

(A3) there exists $s, s' \in L_{\mu_i}^{\succ_i}$ such that $s, s' \in L_{\mu_i}^{\tilde{\succ}_i}$, $s \succ_i s'$ and $s' \tilde{\succ}_i s$.

Type (A1) involves all permutations relative to $\succ_i$ which keep the same ranking of all schools that are truly less preferred to μ_i. Type (A2) considers the misreports which rank some schools that are truly less preferred to μ_i as more preferred and type (A3) considers the misreports which alter the rankings among the schools that are truly less preferred to μ_i.

First note that any permutation $\tilde{\succ}_i$ of type (A1) relates to Proposition 3.1. If $(\tilde{\succ}_{-i}, \tilde{c}_{-i}, \tilde{g}_{-i})$ is plausible, then we have $EDA((g_i, \tilde{g}_{-i}), (\succ_i, \tilde{\succ}_{-i}), (c_i, \tilde{c}_{-i})) = \mu$ and we can apply Proposition 3.1 to obtain $EDA_i((g_i, \tilde{g}_{-i}), (\tilde{\succ}_i, \tilde{\succ}_{-i}), (c_i, \tilde{c}_{-i})) = \mu_i$.

Next, let student i choose a misreport $\tilde{\succ}_i$ that contains permutations of type (A2) and we write $\tilde{S} = \{s' \in S \mid \mu_i \succ_i s' \text{ and } s' \tilde{\succ}_i \mu_i\}$. The key arguments in the proof can roughly be divided into two categories: The submission of $\tilde{\succ}_i$ either would not have effectively influenced the assignment process at all, meaning i's assignment remains μ_i; or there is at least one plausible scenario in which the student is finally assigned to some $s^* \in \tilde{S}$. Here, we discuss the latter and more interesting case. The starting point of our argument is to construct a plausible scenario $(\tilde{\succ}_{-i}, \tilde{c}_{-i}, \tilde{g}_{-i})$ where i is assigned to s^* under $DA((g_i, \tilde{g}_{-i}), (\tilde{\succ}_i, \tilde{\succ}_{-i}))$. Then, we show that either the potential improvements that involve i cannot be realized because the consent of a student is missing; or s^* has no demand under $DA((g_i, \tilde{g}_{-i}), (\tilde{\succ}_i, \tilde{\succ}_{-i}))$. If each student could fully observe the consent decisions of other students, EDA is no longer regret-free truth-telling. Conversely, the uncertainty regarding other students' consent decisions is necessary for our result to hold.[10]

Finally, suppose that the misreport $\tilde{\succ}_i$ contains permutations of type (A3). The key argument for such a misreport is similar to that for type (A2): By submitting $\tilde{\succ}_i$, student i faces the possibility to be assigned to a less preferred school s^* whose order is permuted in $\tilde{\succ}_i$ and which is underdemanded under $DA((g_i, \tilde{g}_{-i}), (\tilde{\succ}_i, \tilde{\succ}_{-i}))$ for a plausible scenario $(\tilde{\succ}_{-i}, \tilde{c}_{-i}, \tilde{g}_{-i})$. However, different from type (A2), here the target school s^* still ranks below μ_i on $\tilde{\succ}_i$. This difference brings an additional challenge to the proof. While for (A2) it is enough to consider a plausible scenario where under truth-telling, i was already assigned to μ_i under DA, for (A3) we need to construct a scenario where under truth-telling, i is involved in at least one solved cycle to improve her from some school $\hat{s} \in L_{\mu_i}^{\succ_i}$ to μ_i. Then, when i submits $\tilde{\succ}_i$, she is assigned to the underdemanded s^* under DA and thus loses the opportunity to be involved in any cycle.

Our final result in this section shows that truth-telling is the unique regret-free choice under EDA.

Proposition 3.2. *For any non-truthful report, there exists an observation at which the student regrets it through truth-telling.*

[10] See Case 3.2.1 in Lemma 3.3 in Appendix 3.C for details.

Proof. See Appendix 3.D. □

At first glance, it might appear that Proposition 3.1 and Proposition 3.2 are in conflict with each other. However, Proposition 3.1 only implies that a certain class of misreports does not change the student's assignment when we fixed an observation that follows from her true preferences. In Proposition 3.2, however, the observation is not fixed. Instead, we show that given any non-truthful report, we can find a corresponding observation, such that truth-telling guarantees weakly better assignments in all plausible scenarios.

As an intuition for Proposition 3.2 note that for every misreport there must exist a pair, say school s and $\tilde{s}$, which compared to the truth, reverse their rankings. Let student i prefer s to $\tilde{s}$ under truth. Now suppose that upon submission of the misreport, she is assigned to $\tilde{s}$ while a seat at s is vacant. Note that the vacant seat at s allows i to infer that the truth would have guaranteed her at worst s. As a result, she will regret not having been truthful. The key step in the proof is to construct an observation of the type just described for any misreport.

3.5 Efficient stable dominating rules

In this section, we extend our analysis to *efficient stable dominating rules*, which are Pareto efficient and only produce outcomes which weakly Pareto dominate a stable matching. In contrast to EDA, consent decisions do not play a role under efficient stable dominating rules and from now on we omit the corresponding notation.

Definition 3.4. A matching rule f is *efficient stable dominating* if for any problem $(I, S, q, g, \succ)$ the matching $f(g, \succ)$ is Pareto efficient and weakly Pareto dominates a stable matching.

Efficient stable dominating rules are a natural refinement of *stable dominating rules*, introduced by Alva and Manjunath (2019). It is well known that efficient rules which Pareto dominate a stable matching rule are not strategy-proof (Abdulkadiroğlu et al., 2009; Kesten, 2010). As we will show next, among efficient stable dominating rules also the weaker property of regret-free truth-telling cannot be fulfilled.

Theorem 3.2. *No efficient stable dominating rule is regret-free truth-telling.*

The proof below is constructive. We provide a problem with $|S| = 2$ and $|I| = 3$, and show that a student regrets submitting her true preferences under any efficient stable dominating rule. We only need small adaptions in the construction to apply the basic argument to any market with $|S| \geq 2$ and $|I| \geq 3$.

Proof. Consider a problem $(I, S, q, g, \succ)$ with two schools $S = \{s_1, s_2\}$ with capacities $q_{s_1} = q_{s_2} = 1$ and three students $I = \{i_1, i_2, i_3\}$. Suppose that i_1's true preferences $\succ_{i_1}$ are

$$s_2 \succ_{i_1} s_\emptyset \succ_{i_1} s_1.$$

Let $\succ_{-i} \in \mathcal{P}_{-i}$ be such that

$$s_1 \succ_{i_2} s_2 \succ_{i_2} s_\emptyset,$$

$$s_2 \succ_{i_3} s_1 \succ_{i_3} s_\emptyset.$$

and consider the following score structure g with

$$g_{i_1}^{s_1} > g_{i_3}^{s_1} > g_{i_2}^{s_1},$$

$$g_{i_2}^{s_2} > g_{i_1}^{s_2} > g_{i_3}^{s_2}.$$

The unique stable matching with respect to $\succ$ is

$$\nu = \{(\mathbf{i}_1, \mathbf{s}_\emptyset), (i_2, s_2), (i_3, s_1)\}$$

and that matching

$$\mu = \{(\mathbf{i}_1, \mathbf{s}_\emptyset), (i_2, s_1), (i_3, s_2)\},$$

is the unique Pareto efficient matching that Pareto dominates ν. Thus, for an arbitrary efficient stable dominating rule, denoted by f^{ESD}, we must have $f^{ESD}(\succ) = \mu$.

In the following, we construct a misreport $\tilde{\succ}_{i_1}$ through which i_1 regrets $\succ_{i_1}$ at observation $(\mu, \pi(\mu, g))$. Before we can make this misreport explicit, we need to describe i_1's inference set $\mathcal{I}(\mu, \succ_{i_1})$. To start, note that

$$g_{i_1}^{s_1} > \pi_{s_1}(\mu, g)\ ,\ g_{i_1}^{s_2} > \pi_{s_2}(\mu, g).$$

We now show that any $\tilde{g}^{s_2}$ must share its ordinal ranking with g^{s_2} for any plausible score structure $\tilde{g}_{-i}$. First, from the observation $(\mu, \pi(\mu, g))$ student i_1 observes that her top choice s_2 is assigned to a lower priority student i_3, i.e. $\tilde{g}_{i_1}^{s_2} > \tilde{g}_{i_3}^{s_2}$. Second, if i_1 would have top priority at s_2 this would imply that i_1 is assigned to s_2 under any stable matching ν' whenever s_2 is submitted as her top choice. Thus, this must also hold true for any Pareto Efficient matching μ' that improves on ν' and hence i_1 can infer that student i_2 must have top priority at s_2. In conclusion, for any plausible $(\tilde{\succ}_{-i_1}, \tilde{g}_{-i_1})$, the corresponding $\tilde{g}^{s_2}$ shares the same ordinal ranking with g^{s_2}.

Next, given $\tilde{g}^{s_2}$, it must hold $\tilde{\succ}_{i_2} = \succ_{i_2}$. First, i_2 must submit s_2 as acceptable since otherwise any stable matching would assign s_2 to i_1. Therefore, i_1 knows $s_2 \mathrel{\tilde{\succ}_{i_2}} s_\emptyset$. Second, note that since i_2 has top priority at s_2, f^{ESD} would have assigned s_2 to i_2 if i_2 would have submitted s_2 as her top choice. Thus, i_1 knows $s_1 \mathrel{\tilde{\succ}_{i_2}} s_2$. Combining the two relations i_1 can infer that $\tilde{\succ}_{i_2} = \succ_{i_2}$ is the unique candidate contained in any plausible $(\tilde{\succ}_{-i_1}, \tilde{g}_{-i_1})$.

Now, we describe the candidates for $\tilde{g}^{s_1}$. First, by observing $(\mu, \pi(\mu, g))$, student i_1 knows that s_1 is assigned to the lower priority student i_2, i.e., $\tilde{g}^{s_1}_{i_1} > \tilde{g}^{s_1}_{i_2}$. Second, we establish that given the information regarding $\tilde{g}^{s_2}$ and $\tilde{\succ}_{i_2}$, we must have $\tilde{g}^{s_1}_{i_3} > \tilde{g}^{s_1}_{i_2}$. Suppose by contradiction that $\tilde{g}^{s_1}_{i_2} > \tilde{g}^{s_1}_{i_3}$. In this case, in f^{ESD}, i_1 and i_2 must be assigned to their top choices s_2 and s_1, respectively. However, this is incompatible with μ. Thus, there are two remaining ordinal rankings

$$\text{either } \tilde{g}^{s_1}_{i_1} > \tilde{g}^{s_1}_{i_3} > \tilde{g}^{s_1}_{i_2} \text{ or } \tilde{g}^{s_1}_{i_3} > \tilde{g}^{s_1}_{i_1} > \tilde{g}^{s_1}_{i_2}$$

that are compatible with any plausible scenario $(\tilde{\succ}_{-i_1}, \tilde{g}_{-i_1})$.

At last, we show that only $\tilde{\succ}_{i_3} = \succ_{i_3}$ is compatible with i_1's observation. First, since i_3 is assigned to s_2 in μ, student i_1 can conclude that $s_2 \tilde{\succ}_{i_3} s_\emptyset$. If i_3 would have submitted $s_\emptyset \tilde{\succ}_{i_3} s_1$, then any stable matching would have assigned both i_1 and i_2 to their top choices, which is incompatible with the observation. Thus, it must be true that $s_1 \tilde{\succ}_{i_3} s_\emptyset$. Furthermore, suppose by contradiction that $s_1 \tilde{\succ}_{i_3} s_2$. Given that $s_\emptyset \succ_{i_1} s_1$ and $\tilde{g}^{s_1}_{i_3} > \tilde{g}^{s_1}_{i_2}$, student i_3 is assigned to s_1 under f^{ESD}, which is again incompatible with observing μ. Hence, student i_3 can only have submitted $\tilde{\succ}_{i_3} = \succ_{i_3}$.

As a result, we can classify i_1's inference set $\mathcal{I}(\mu, \succ_{i_1})$ into two cases that are distinguished by the remaining candidates of ordinal rankings for scores at s_1.

We now show that i_1 regrets reporting the truth $\succ_{i_1}$ at $(\mu, \pi(\mu, g))$ through

$$\tilde{\succ}_{i_1} : s_2 \tilde{\succ}_{i_1} \mathbf{s}_1 \tilde{\succ}_{i_1} \mathbf{s}_\emptyset.$$

We do so by establishing that among the two possible classes from the inference set, in one class i_1 is strictly better off through the misreport and she is not worse off in the remaining class.

Case 1 Suppose that $(\tilde{\succ}_{-i_1}, \tilde{g}_{-i_1}) \in \mathcal{I}(\mu, \succ_{i_1})$ satisfies $\tilde{g}^{s_1}_{i_1} > \tilde{g}^{s_1}_{i_3} > \tilde{g}^{s_1}_{i_2}$. In this case, we argue that f^{ESD} must assign i_1 to s_2 when i_1 submits $\tilde{\succ}_i$. Hence, student i_1 would strictly improve her assignment from $s_\emptyset$ under truth-telling to her top choice s_2. We first establish that there is a unique stable matching

$$\tilde{\nu} = \{(\mathbf{i}_1, \mathbf{s}_1), (i_2, s_2), (i_3, s_\emptyset)\}.$$

Note that in any stable matching i_1 cannot be assigned to $s_\emptyset$, since i_1 would have justified envy at s_1. This implies that whenever i_1 is not assigned to s_2, she must be assigned to s_1. Furthermore, if i_1 is matched with s_2, then i_2 must be assigned to s_1, which would mean that i_3 has justified envy at s_1. Thus, the unique stable matching corresponds to $\tilde{\nu}$. Hence, any efficient stable dominating rule must select

$$\tilde{\mu} = \{(\mathbf{i}_1, \mathbf{s}_2), (i_2, s_1), (i_3, s_\emptyset)\}$$

since it is the only Pareto efficient matching that dominates $\tilde{\nu}$. Thus, we conclude that conditional on her observation $(\mu, \pi(\mu, g))$, in this scenario, i_1 would have been better off if she had reported $\tilde{\succ}_{i_1}$ to f^{ESD}.

Case 2 It remains to show that given $(\tilde{\succ}_{-i_1}, \tilde{g}_{-i_1}) \in \mathcal{I}(\mu, \succ_{i_1})$ with $\tilde{g}^{s_1}_{i_3} > \tilde{g}^{s_1}_{i_1} > \tilde{g}^{s_1}_{i_2}$, student i_1 is not assigned to a worse option than under truth-telling (namely s_1). Clearly, in this case the unique stable matching is ν, while the unique matching that Pareto dominates ν is μ. Therefore, i_1 will be assigned to $s_\emptyset$ under f^{ESD}, which is the same assignment as under true preferences.

Since the choice of f^{ESD} was arbitrary, we have shown that for any efficient stable dominating rule, student i_1 regrets reporting the truth $\succ_{i_1}$ through misreport $\tilde{\succ}_{i_1}$ at $(\mu, \pi(\mu, g))$. This completes the proof. □

As mentioned before, this example allows us to illustrate one important feature of Theorem 3.1. Concretely, the observation $(\mu, \pi(\mu, g))$ at the beginning of the example is reached through EDA if one leaves reported preferences unchanged and additionally requires that $c_{i_1} = 1$. Notice that in Case 1 the improvement of i_1's assignment from $s_\emptyset$ to s_2 relies on the consent of student i_3. However, based on i_1's observation, the consent decision of i_3 cannot be inferred by i_1. Specifically, if $c_{i_3} = 0$, then i_1 would be assigned to s_1 in Case 1 which implies that she would not regret that she had told the truth. Thus, EDA being regret-free truth-telling relies partially on the uncertainty regarding other students' consent decisions.

All our results extend to the more restrictive case where instead of observing the full matching μ, each student i observes only her own assignment μ_i and the cutoffs. For Theorem 3.2 this can be explained as follows. In the problem constructed above, there is only one additional consistent matching if i_1 observes only μ_i. For this matching, which switches the assignments for student i_2 and i_3 compared to μ, a symmetric argument leads to the same conclusion as for μ.

3.6 Conclusion

Telling the truth is a safe choice under EDA if students wish to avoid regret their submitted reports. Strengthening this first result, we have also shown that truth-telling is the unique regret-free option under EDA. Moreover, among the class of efficient stable dominating rules—a class that covers natural alternatives for EDA in practice—no candidate is regret-free truth-telling. Our results open up several avenues for future research. For instance, it would be interesting to study whether EDA is the unique candidate among all non-strategy-proof and constrained Pareto-Efficient rules which is regret-free truth-telling. It is also an open question whether EDA is still regret-free if schools' priorities take the form of more flexible choice functions.[11]

[11] Ehlers and Morrill (2020) introduce a generalized version of EDA that might serve as a starting point for an investigation.

3.A Deferred acceptance rule

In this section, we first introduce the Student Proposing Deferred Acceptance Algorithm which induces the *Student-Proposing Deferred Acceptance Rule (DA)* due to Gale and Shapley (1962). Thereafter, we present a lemma on *DA* that is necessary to prove Proposition 3.1 and Theorem 3.1. In the following, fix a problem $(I, S, q, g \succ, c)$. The DA algorithm works as follows:

Step 1 Each student $i \in I$ proposes to her most preferred school in $S \cup \{s_\emptyset\}$. Each school $s \in S$ considers all the proposals and tentatively accepts the candidates who apply to s and are among the q_s-highest ranked applicants at that school. The remaining proposals are rejected. If there are fewer than q_s proposals, s accepts all of them. Moreover, all students that propose to the outside option $s_\emptyset$ are accepted.

Step k, $k \geq 2$ Each student who was rejected at step $k-1$ applies to her most preferred school not yet applied to. Each school $s \in S$ considers all the new applicants together with those who are tentatively assigned to it at step $k-1$. Each school s now tentatively accepts the q_s-highest ranked applicants and rejects all others. If there are fewer than q_s proposals, s accepts all of them. Moreover, all students that propose to the outside option $s_\emptyset$ are accepted.

The algorithm terminates with the tentative assignments of the first step in which no student is rejected. For our lemma presented below we define Weak Maskin Monotonicity as in Kojima and Manea (2010). We call $\succ'$ a monotonic transformation of $\succ$ at matching μ, if for each $i' \in I$, $\succ'_{i'}$ is a monotonic transformation of $\succ_{i'}$ at $\mu_{i'}$.

Definition 3.5. A matching rule f is *weakly Maskin monotonic* if, given any $\succ$ and for any $\succ'$ that is a monotonic transformation of $\succ$ at $f(g, \succ, c)$, $f(g, \succ', c)$ weakly Pareto dominates $f(g, \succ, c)$

Kojima and Manea (2010) show that DA is weakly Maskin monotonic. Furthermore, DA is strategy-proof (Dubins and Freedman, 1981; Roth, 1982b) and produces the SOSM for a given score structure and preference profile.

Lemma 3.1. *Let $\succ_i' \in \mathcal{P}$ be a monotonic transformation of $\succ_i$ at $DA_i(g, \succ)$, then $DA(g, (\succ_i', \succ_{-i}))$ weakly Pareto dominates $DA(g, \succ)$ and i's outcomes are identical, i.e., $DA_i(g, \succ) = DA_i(g, (\succ_i', \succ_{-i}))$.*

Proof. The first part follows from weak Maskin monotonicity of DA. The second part is proved by means of contradiction. Suppose that $DA_i(g, \succ) \neq DA_i(g, (\succ_i', \succ_{-i}))$, then by weak Maskin monotonicity of DA, $DA_i(g, (\succ_i', \succ_{-i})) \succ_i DA_i(g, \succ)$, which contradicts strategy-proofness of DA. □

3.B Proof of Proposition 3.1

For ease of presentation, we use $EDA(\succ)$ to refer to $EDA(g, (\succ_i, \succ_{-i}), c)$ and $EDA(\tilde{\succ})$ to refer to $EDA(g, (\tilde{\succ}_i, \succ_{-i}), c)$. In a similar way, we use $DA(\succ)$ to refer to $DA(g, (\succ_i, \succ_{-i}))$ and $DA(\tilde{\succ})$ to refer to $DA(g, (\tilde{\succ}_i, \succ_{-i}))$.

We first show that the outcomes of EDA are identical under both profiles given that i consents, i.e., we prove that $EDA(\succ) = EDA(\tilde{\succ})$ when $c_i = 1$. At the end of the proof we extend our arguments to cover the case where $c_i = 0$.

Let $pTP^{\succ} = \{\gamma^t\}_{t=1}^{T}$ be an arbitrary realized process of the TP algorithm with input $(\succ, c, g)$ that are captured by the series of solved top priority cycles $\{\gamma^t\}_{t=1}^{T}$. Specifically, for each $t \leq T$, γ^t is solved at step t of $pTP^{\succ}$ and we set $EDA^t(\succ) = \gamma^t \circ EDA^{t-1}(\succ)$ with $EDA^0(\succ) = DA(\succ)$.

Since the outcome of the TP algorithm is invariant in the choice of the cycle solved in each round, it suffices to construct one TP process with input $((\tilde{\succ}_i, \succ_{-i}), c, g)$, denoted by $pTP^{\tilde{\succ}}$, that leads to the same outcome as $pTP^{\succ}$. As a part of our construction, we make use of the algorithm presented next.

Initialize: Let $t = 1$. Also, let $\nu^0(\tilde{\succ}) = DA(\tilde{\succ})$ and $EDA^0(\succ) = DA(\succ)$.

Round $t \leq T$: Let $L^t = \{l \in I \mid \nu_l^{t-1}(\tilde{\succ}) \neq EDA_l^{t-1}(\succ)\}$.

- If each $jk \in \gamma^t$ satisfies that $j, k \in L^t$, let $\nu^t(\tilde{\succ}) = \nu^{t-1}(\tilde{\succ})$. Then, move to Round $t+1$ or terminate the algorithm if $t = T$.
- If there exists $jk \in \gamma^t$ such that $j \notin L^t$ or $k \notin L^t$, let $\nu^t(\tilde{\succ}) = \gamma^t \circ \nu^{t-1}(\tilde{\succ})$. Then, move to Round $t+1$ or terminate the algorithm if $t = T$.

Collect in $\{\tilde{\gamma}^t\}_{t=1}^{\tilde{T}}$ the series of cycles solved in the course of running the algorithm and note that, by construction, we have $\{\tilde{\gamma}^t\}_{t=1}^{\tilde{T}} \subseteq \{\gamma^t\}_{t=1}^{T}$. We now show that the generated cycle selection $\{\tilde{\gamma}^t\}_{t=1}^{\tilde{T}}$ allows to fully describe the desired $pTP^{\tilde{\succ}}$ which terminates at matching $EDA(\succ)$. Our strategy will be as follows. At the first step, we establish that the algorithm is well defined. At the second step, we will argue that $\nu^T(\tilde{\succ}) = EDA^T(\succ)$ and that $G^*(\nu^T(\tilde{\succ}))$ contains no cycles.

Step 1 We can generate the desired sequence of cycles $\{\tilde{\gamma}^t\}_{t=1}^{\tilde{T}}$ if for each round $t \leq T$, the following three statements are satisfied:

(B1) Either all agents involved in γ^t belong to L^t, or none of them does.

(B2) $\gamma^t \in G^*(\nu^{t-1}(\tilde{\succ}))$ when γ^t contains no agent from L^t.

(B3) $\nu^t(\tilde{\succ})$ weakly Pareto dominates $EDA^t(\succ)$, and $L^{t+1} \subseteq L^t$.

For each t, statement (B1) and statement (B2) ensure that we can find and solve the cycle as described in the algorithm in round t. Then, given that (B1) and (B2) are true, statement (B3), is needed to ensure that (B1) and (B2) will also be true for the next round $t+1$. To prove these three statements, we now argue via induction over t.

For the initial case we build on the following observations. First, it is immediate from Lemma 3.1, that $DA(\tilde{\succ})$ weakly Pareto dominates $DA(\succ)$ and $DA_i(\succ) = DA_i(\tilde{\succ})$. Thus, we can infer that $L^1 = \{l \in I \mid DA_l(\tilde{\succ}) \succ_l DA_l(\succ)\}$ and $i \notin L^1$. Furthermore, by the definition of ν it is true that $DA_l(\tilde{\succ}) = \nu_l^0(\tilde{\succ})$ for any $l \in I$. Note that these conditions resemble those in condition (B3). Moreover, let $S' = \{s \in S \mid s \succ_i \mu_i \text{ and } \mu_i \tilde{\succ}_i s\}$.

We present our arguments in their general form since they are also applicable to the inductive step. That is, for the initial case we do not explicitly insert $t = 1$.

Initial case (Let $t = 1$): *Statement (B1):* Since γ^t is a cycle, it suffices to show that $jk \in \gamma^t$ and $k \in L^t$ imply $j \in L^t$.

Towards this goal, we first establish that for $jk \in \gamma^t$, if $k \in L^t$, then $j \in L^t \cup \{i\}$. More generally, we show that for any $jk \in G^*(EDA^{t-1}(\succ))$, if $k \in L^t$, then we have $j \in L^t \cup \{i\}$. This generality will turn out to be useful proving other statements later on. By contradiction, let $j \notin L^t, j \neq i$ and $k \in L^t$. We aim at a contradiction towards the stability of $DA(\tilde{\succ})$. First, if $k \in L^t$, then there exists $l \in L^t$ such that $\nu_l^{t-1}(\tilde{\succ}) = EDA_k^{t-1}(\succ)$. Since $l \in L^t$, it holds $DA_l(\tilde{\succ}) = \nu_l^{t-1}(\tilde{\succ}) \succ_l EDA_l^{t-1}(\succ)$. Remarkably, for the initial case this argument is immediate since $DA_l(\tilde{\succ}) = \nu_l^0(\tilde{\succ}) \succ_l DA_l(\succ)$. When $t > 1$, the validity of this argument depends on the results we will establish later in the inductive step. Next, the previous observations and $jk \in G^*(EDA^{t-1}(\succ))$ imply that $g_j^{DA_l(\tilde{\succ})} > g_l^{DA_l(\tilde{\succ})}$ and $EDA_k^{t-1}(\succ) \succ_j EDA_j^{t-1}(\succ)$. Furthermore, $j \notin L^t$ implies $EDA_j^{t-1}(\succ) = \nu_j^{t-1}(\tilde{\succ}) \succeq_j DA_j(\tilde{\succ})$ while $j \neq i$ implies $\succ_j = \tilde{\succ}_j$. Combining the relations derived so far, leads to

$$DA_l(\tilde{\succ}) = \nu_l^{t-1}(\tilde{\succ}) = EDA_k^{t-1}(\succ) \tilde{\succ}_j EDA_j^{t-1}(\succ) = \nu_j^{t-1}(\tilde{\succ}) \tilde{\succeq}_j DA_j(\tilde{\succ}).$$

However, this implies that j has justified envy towards l at $DA(\tilde{\succ})$. Hence we arrive at a contradiction to the stability of $DA(\tilde{\succ})$ with respect to $\tilde{\succ}$.

It remains to show that $jk \in \gamma^t$ and $k \in L^t$ imply $j \neq i$. When $EDA_k^{t-1}(\succ) \notin S'$, the arguments above ensure $ik \notin G^*(EDA^{t-1}(\succ))$, and therefore also $ik \notin \gamma^t$. Consider the remaining case where $EDA_k^{t-1}(\succ) \in S'$. Here, if $ik \in \gamma^t$, then it implies that $EDA_k^{t-1}(\succ) = EDA_i^t(\succ) \succ_i \mu_i$. However, this is a contradiction to μ being the final matching. Thus, we must have $j \neq i$.

Now, we can conclude that once there is an edge $jk \in \gamma^t$ with $k \in L^t$, then $j \in L^t$. Therefore, either all agents involved in γ^t belong to L^t, or no such agent does. Statement (B1) is satisfied at round t.

Statement (B2): Given that (B1) is true at round t, we proceed to prove (B2). Suppose that for each $jk \in \gamma^t, j, k \notin L^t$. Thus, we get $EDA_j^{t-1}(\succ) = \nu_j^{t-1}(\tilde{\succ})$ and $EDA_k^{t-1}(\succ) = \nu_k^{t-1}(\tilde{\succ})$. This implies that

$$\nu_k^{t-1}(\tilde{\succ}) \mathrel{\tilde{\succ}_j} \nu_j^{t-1}(\tilde{\succ}).$$

Note that this is also true if $j = i$, since in this case $\nu_k^{t-1}(\tilde{\succ}) \notin S'$. Hence, we obtain that student j must still desire $\nu_k^{t-1}(\tilde{\succ})$ at $\nu^{t-1}(\tilde{\succ})$. Note that the last argument is true for all j such that $jk \in \gamma^t$. Thus, we have that all students involved in γ^t are temporarily matched at $\nu^{t-1}(\tilde{\succ})$. Next, since $\nu^{t-1}(\tilde{\succ})$ weakly Pareto dominates $EDA^{t-1}(\succ)$, there are weakly fewer temporarily matched students who desire $\nu_k^{t-1}(\tilde{\succ})$ at $\nu^{t-1}(\tilde{\succ})$ compared to $EDA^{t-1}(\succ)$. As a result, j still has the highest score among all temporarily matched students pointing to k. Hence $jk \in G^*(\nu^{t-1}(\tilde{\succ}))$. Since this holds for all edges in γ^t, it follows that $\gamma^t \in G^*(\nu^{t-1}(\tilde{\succ}))$.

Statement (B3): We start with showing that the desired weak Pareto dominance relation holds at the end of round t. To begin with, note that $\nu^{t-1}(\tilde{\succ})$ weakly Pareto dominates $EDA^{t-1}(\succ)$ and that if any, only students in γ^t change their assignments in round t of our algorithm (and also in round t of $pTP^{\succ}$). Thus, to conclude that $\nu^t(\tilde{\succ})$ weakly Pareto dominates $EDA^t(\succ)$, it is sufficient to show that for each $jk \in \gamma^t$:

$$\nu_j^t(\tilde{\succ}) \succeq_j EDA_j^t(\succ).$$

Of the two cases we have to consider, we start with the simpler one, in which for any $jk \in \gamma^t$, we have

$j, k \notin L^t$. In this case, γ^t is solved in both $\nu^{t-1}(\tilde{\succ})$ and $EDA^{t-1}(\succ)$. Therefore, $\nu_j^t(\tilde{\succ}) = EDA_j^t(\succ)$ and we obtain the desired result.

In the remaining case, any $jk \in \gamma^t$ satisfies that $j, k \in L^t$. Clearly, we can solve a cycle of this form only if $L^t \neq \emptyset$. Moreover, note that $EDA^t(\succ) = \gamma^t \circ EDA^{t-1}(\succ)$ and $\nu^t(\tilde{\succ}) = \nu^{t-1}(\tilde{\succ})$. We proceed by contradiction and assume that $EDA_j^t(\succ) \succ_j \nu_j^t(\tilde{\succ})$. We derive a contradiction to the stability of $DA(\tilde{\succ})$ with respect to $\tilde{\succ}$. We make the following observations: First, since $k \in L^t$, there must exist $l \in L^t$ such that we have $\nu_l^{t-1}(\tilde{\succ}) = EDA_k^{t-1}(\succ)$. Second, $l \in L^t$ implies that $DA_l(\tilde{\succ}) = \nu_l^{t-1}(\tilde{\succ}) \succ_l EDA_l^{t-1}(\succ)$. Therefore, $jk \in \gamma^t$ also means that $g_j^{DA_l(\tilde{\succ})} > g_l^{DA_l(\tilde{\succ})}$ and $EDA_k^{t-1}(\succ) = EDA_j^t(\succ)$. Third, the algorithm guarantees that $\nu_j^t(\tilde{\succ}) \succeq_j DA_j(\tilde{\succ})$. If we combine all relations above with $\succ_j = \tilde{\succ}_j$, we obtain

$$DA_l(\tilde{\succ}) = \nu_l^{t-1}(\tilde{\succ}) = EDA_k^{t-1}(\succ) = EDA_j^t(\succ) \,\tilde{\succ}_j\, \nu_j^t(\tilde{\succ}) \,\tilde{\succeq}_j\, DA_j(\tilde{\succ})$$

and reach a contradiction, since j has justified envy towards l at $DA(\tilde{\succ})$. Thus, $\nu^t(\tilde{\succ})$ weakly Pareto dominates $EDA^t(\succ)$. Moreover, based on the Pareto dominance result, we can also write L^{t+1} as $L^{t+1} = \{l \in I \mid \nu_l^t(\tilde{\succ}) \succ_l EDA_l^t(\succ)\}$.

To finish the proof for statement (B3) we need to show that $L^{t+1} \subseteq L^t$ for which we again have two cases to consider. If any $jk \in \gamma^t$ satisfies $j, k \notin L^t$, then it is immediate that $L^{t+1} = L^t$. On the contrary, if any $jk \in \gamma^t$ satisfies $j, k \in L^t$, then we make the following two observations. First, for each such j, as $j \in L^t$, we have $\nu_j^{t-1}(\tilde{\succ}) \succ_j EDA_j^{t-1}(\succ)$ and $\nu_j^t(\tilde{\succ}) \succeq_j EDA_j^t(\succ)$. This implies that while j is contained in L^t, she might not be in L^{t+1}. Second, for each $j' \in I$ not involved in γ^t, we have $\nu_{j'}^t(\tilde{\succ}) = \nu_{j'}^{t-1}(\tilde{\succ})$ and $EDA_{j'}^t(\succ) = EDA_{j'}^{t-1}(\succ)$, which implies that $j' \in L^t$ if and only if $j' \in L^{t+1}$. In conclusion, we can infer that $L^{t+1} \subseteq L^t$. Hence statement (B3) is satisfied.

Inductive step: Let $t > 1$ and assume that for all $t' < t$, (B1) - (B3) are satisfied. By assumption of the inductive step and from (B3), we have that $L^{t'+1} \subseteq L^{t'}$ for any $t' < t$, which implies $L^t \subseteq L^{t'}$. Second, through the description of the algorithm, we know that given any $t' < t$, assignments at $\nu^{t'}(\tilde{\succ})$ and $\nu^{t'-1}(\tilde{\succ})$ are identical for each student in $L^{t'}$. Therefore, since $L^t \subseteq L^{t'}$, we can infer that for each $l \in L^t$, $DA_l(\tilde{\succ}) = \nu_l^{t-1}(\tilde{\succ})$. Together with the observations above, the arguments we already presented for (B1) in the initial case also apply to the inductive step. Furthermore, the same holds for (B2). Finally, given we established (B1) and (B2), also (B3) follows again from the same arguments as in the initial case. This completes the induction.

Step 2: We show that $EDA^T(\succ) = \nu^T(\tilde{\succ})$. Let $t_i \leq T$ be the first step in $pTP^{\succ}$ where i is permanently matched and consider round t_i of our algorithm.

If $EDA^{t_i-1}(\succ) = \nu^{t_i-1}(\tilde{\succ})$, we have that $L^t = \emptyset$ and that γ^t is solved in each round $t > t_i$ of the algorithm. Consequently, it is true that $EDA^T(\succ) = \nu^T(\tilde{\succ})$.

If $EDA^{t_i-1}(\succ) \neq \nu^{t_i-1}(\tilde{\succ})$, then L^{t_i} is non-empty. In this case, we show that there exists $\hat{t} > t_i$ such that $EDA^{\hat{t}}(\succ) = \nu^{\hat{t}}(\tilde{\succ})$. As shown above, this leads to $EDA^T(\succ) = \nu^T(\tilde{\succ})$.

We show that there must be a cycle in $G^*(EDA^{t_i-1}(\succ))$ that solely consists of elements in L^{t_i}. We begin with showing that for any $k \in L^{t_i}$, there exists an edge $jk \in G^*(EDA^{t_i-1}(\succ))$ for some $j \in I$. Since $k \in L^{t_i}$, there exists $l \in L^{t_i}$ such that $EDA_k^{t_i-1}(\succ) = \nu_l^{t_i-1}(\tilde{\succ}) \succ_l EDA_l^{t_i-1}(\succ)$. That is, at $EDA^{t_i-1}(\succ)$, for each student in L^{t_i}, her assignment is desired by at least one student in L^{t_i} whose assignment is further desired by some other student in L^{t_i}. Now, recall that we assume $c_1 = 1$. Since i is permanently matched at step t_i and i consents, then even if i prefers $EDA_k^{t_i-1}(\succ)$ to μ_i, she cannot prevent any agent from being eligible for $EDA_k^{t_i-1}(\succ)$. In other words, at least one edge that is pointing to k, namely lk, is contained in $G(EDA^{t_i-1}(\succ))$. Therefore, we can infer that k is temporarily matched in $EDA^{t_i-1}(\succ)$ and thus there must be $jk \in G^*(EDA^{t_i-1}(\succ))$ for some $j \in I$.

Next, for any such jk, our arguments from (B1) will be sufficient to conclude that $j \in L^{t_i}$. First, we have already shown $j \in L^{t_i} \cup \{i\}$. Second, we know that $j \neq i$, since i is permanently matched. Thus, we can infer that each student in L^{t_i} is pointed by another student in L^{t_i} in $G^*(EDA^{t_i-1}(\succ))$. Since L^{t_i} is finite, the existence of the desired cycle is guaranteed. Notably, according to (B3) and by iteratively applying the same argument, we can eventually reach a round $\hat{t} > t_i$ where $EDA^{\hat{t}}(\succ) = \nu^{\hat{t}}(\tilde{\succ})$.

We next claim that no cycles can be found in $G^*(\nu^T(\tilde{\succ}))$. Notably, if $G^*(\nu^T(\tilde{\succ}))$ has a cycle, by similar arguments in (B2), we can infer that $G^*(EDA^T(\succ))$ must also have a cycle. However, this contradicts the fact that exactly T cycles are solved in $pTP^{\succ}$.

Based on the statements provided so far, we can construct the desired $pTP^{\tilde{\succ}}$ as $pTP^{\tilde{\succ}} = \{\tilde{\gamma}^t\}_{t=1}^{\tilde{T}}$. This leads to

$$EDA(\succ) = EDA(\tilde{\succ})$$

which completes the proof for $c_i = 1$.

Finally, it remains to prove that our results extend to the case where $c_i = 0$. Note that since EDA is consent-invariant, the following two relations are true: $EDA_i(\succ) = EDA_i(g, (\succ_i, \succ_{-i}), (\tilde{c}_i, c_{-i}))$ and $EDA_i(\tilde{\succ}) = EDA_i(g, (\tilde{\succ}_i, \succ_{-i}), (\tilde{c}_i, c_{-i}))$ for $\tilde{c}_i = 1$. Since we just showed when i consents, submitting $\tilde{\succ}_i$ will not alter the outcome: $EDA(g, (\succ_i, \succ_{-i}), (\tilde{c}_i, c_{-i})) = EDA(g, (\tilde{\succ}_i, \succ_{-i}), (\tilde{c}_i, c_{-i}))$. This allows us to conclude $EDA_i(\succ) = EDA_i(\tilde{\succ})$, which completes the proof.

3.C Proof of Theorem 3.1

Fix an arbitrary problem $(I, S, q, g, \succ, c)$ and consider an arbitrary student $i \in I$. Since EDA only takes acceptable schools into account, for any tuple $(g, \succ_{-i}, c)$ and any $\succ'_i \in T_i$, we can claim that $EDA(g, (\succ'_i, \succ_{-i}), c) = EDA(g, (\succ_i, \succ_{-i}), c)$. Hence, if student i does not regret reporting her true preferences $\succ_i$, she does not regret to report any $\succ'_i \in T_i$. Thus, we show that i does not regret to report $\succ_i$.

Lemma 3.2, 3.3 and 3.7 each consider a distinct class of misreports of student i and jointly imply that i cannot regrets submitting her true preferences. In the following exposition, take an arbitrary observation $(\mu, \pi(\mu, g))$ where $\mu \in \mathcal{M}|_{(\succ_i, c_i)}$. We fix i's scores g_i and i's consent decision c_i throughout the proof. From now on, we use $\tilde{g}$ to refer to $(g_i, \tilde{g}_{-i})$ and $\tilde{c}$ to refer to $(c_i, \tilde{c}_{-i})$.

We first show that a misreport is not profitable for i, if it shares the same relative ranking of schools weakly below her own assignment under truth-telling.

Lemma 3.2. *Consider $\tilde{\succ}_i \in \mathcal{P}$ such that for all $s, s' \in L^{\succ_i}_{\mu_i}$, $s \tilde{\succ}_i s'$ if and only if $s \succ_i s'$. For any $(\tilde{\succ}_{-i}, \tilde{c}_{-i}, \tilde{g}_{-i}) \in \mathcal{I}(\mu, \succ_i, c_i)$, it is true $EDA_i(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i}), \tilde{c}) = \mu_i$.*

Proof. Select any $(\tilde{\succ}_{-i}, \tilde{c}_{-i}, \tilde{g}_{-i}) \in \mathcal{I}(\mu, \succ_i, c_i)$. By definition, $EDA(\tilde{g}, (\succ_i, \tilde{\succ}_{-i}), \tilde{c}) = \mu$ and using Proposition 3.1, we know $EDA_i(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i}), \tilde{c}) = \mu_i$. □

We proceed with misreports in which some schools ranked below μ_i under truth permute their order with μ_i. Our next Lemma shows that the student can either infer that she would have possibly been worse off, or that the misreport would not have affected her assignment in any plausible scenario.

Lemma 3.3. *Consider $\tilde{\succ}_i \in \mathcal{P}$ such that $\mu_i \succ_i s$ and $s \tilde{\succ}_i \mu_i$ for some $s \in S$. Then,*
(1) *either there exists $(\tilde{\succ}_{-i}, \tilde{c}_{-i}, \tilde{g}_{-i}) \in \mathcal{I}(\mu, \succ_i, c_i)$ such that $\mu_i \succ_i EDA_i(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i}), \tilde{c})$;*
(2) *or for any $(\tilde{\succ}_{-i}, \tilde{c}_{-i}, \tilde{g}_{-i}) \in \mathcal{I}(\mu, \succ_i, c_i)$: $EDA_i(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i}), \tilde{c}) = \mu_i$.*

Proof. Let $\tilde{S} = \{s' \in S \mid \mu_i \succ_i s' \text{ and } s' \tilde{\succ}_i \mu_i\}$. We start by considering the case where $\tilde{S} = \{s^*\}$ is a singleton. We will explain how to generalize the arguments to cases where $\tilde{S}$ contains more elements at the end of the proof. Given that $\tilde{S}$ is a singleton, we distinguish the following exhaustive cases based on i's observation $(\mu, \pi(\mu, g))$:

Case 1: $\pi_{s^*}(\mu, g) = 0$. If $\pi_{s^*}(\mu, g) = 0$, then s^* has not exhausted its capacity at the observed matching. We use the following argument repeatedly throughout the proof: Note that students assigned to a school that has not exhausted its capacity under the observed matching cannot be involved in a cycle in any corresponding TP process for any plausible scenario. This implies that at this school also under DA the same set of students must have been assigned there. Furthermore, since DA is non-wasteful, we can conclude that at any plausible scenario the school has no demand under the DA matching. Concretely, since for any $(\tilde{\succ}_{-i}, \tilde{c}_{-i}, \tilde{g}_{-i}) \in \mathcal{I}(\mu, \succ_i, c_i)$, s^* has vacant seat at $EDA(\tilde{g}, (\succ_i, \tilde{\succ}_{-i}), \tilde{c}) = \mu$, then s^* must also have vacant seat at $DA(\tilde{g}, (\succ_i, \tilde{\succ}_{-i}))$ and that for any $i' \in I$, i' weakly prefers $DA_{i'}(\tilde{g}, (\succ_i, \tilde{\succ}_{-i}))$ to s^* given preference profile $\succ$. That is, s^* has no demand.

Next, if i submits $\tilde{\succ}_i$ then we obtain $DA_i(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i})) = s^*$. Now notice that before being matched to the final assignment, the set of applications i sends to reach $DA_i(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i}))$ is a subset of those sent to reach $DA_i(\tilde{g}, (\succ_i, \tilde{\succ}_{-i}))$. Therefore, $DA_{i'}(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i})) \succeq_{i'} DA_{i'}(\tilde{g}, (\succ_i, \tilde{\succ}_{-i}))$ holds for all $i' \neq i$. Accordingly, each agent $i' \in I$ still weakly prefers $DA_{i'}(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i}))$ to s^* given preference profile $\tilde{\succ}$. Hence s^* has again no demand at $DA(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i}))$ and thus no agent is pointing to i in $G^*(DA(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i})))$. As a result, i cannot be involved in any solved cycle during the TP process and thus $EDA_i(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i}), \tilde{c}) = DA_i(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i})) = s^*$. Statement (1) holds.

Case 2: $\pi_{s^*}(\mu, g) \neq 0$, $\pi_{\mu_i}(\mu, g) = 0$ and $g_i^{s^*} < \pi_{s^*}(\mu, g)$. Under this condition, we show that statement (2) is satisfied. Take an arbitrary $(\tilde{\succ}_{-i}, \tilde{c}_{-i}, \tilde{g}_{-i}) \in \mathcal{I}(\mu, \succ_i, c_i)$. To start, note that whenever a student j improves her assignment from one school to another at one step of the TP algorithm, another student with lower priority is assigned to the school that j left at that step. Since $g_i^{s^*} < \pi_{s^*}(\mu, g)$, this implies that student i must have a lower score than any student assigned to s^* at $DA_i(\tilde{g}, (\succ_i, \tilde{\succ}_{-i}))$. Thus, compared to the DA procedure of i submitting $\succ_i$, i's additional application to s^* by submitting $\tilde{\succ}_i$ has no influence on the outcome and we reach $DA(\tilde{g}, (\succ_i, \tilde{\succ}_{-i})) = DA(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i}))$. Moreover, since $\pi_{\mu_i}(\mu, g) = 0$ and as argued in Case 1, μ_i must have vacant seat at $DA(\tilde{g}, (\succ_i, \tilde{\succ}_{-i}))$, thus also at $DA(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i}))$. As a result, μ_i has no demand at $DA(\tilde{g}, (\tilde{\succ}_i.\tilde{\succ}_{-i}))$ and this implies that $EDA_i(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i}), \tilde{c}) = DA_i(\tilde{g}, (\tilde{\succ}_i.\tilde{\succ}_{-i})) = \mu_i$. Hence, statement (2) holds.

Case 3: $\pi_{s^*}(\mu, g) \neq 0$ and either (C1) $g_i^{s^*} > \pi_{s^*}(\mu, g)$; or (C2) $\pi_{\mu_i}(\mu, g) \neq 0$ and $g_i^{s^*} < \pi_{s^*}(\mu, g)$. Throughout the discussion, we will make it explicit whenever (C1) and (C2) are in need to be distinguished.[12] Furthermore, except for the last subcase (Case 3.2.2.2), statement (1) will apply in Case 3 and our approach for each subcase except this last subcase will be standardized going through the following steps:

Step 1: We construct a candidate scenario $(\tilde{\succ}_{-i}, \tilde{c}_{-i}, \tilde{g}_{-i})$.

Step 2: We show that $(\tilde{\succ}_{-i}, \tilde{c}_{-i}, \tilde{g}_{-i}) \in \mathcal{I}(\mu, \succ_i, c_i)$.

Step 3: We argue that $EDA_i(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i}), \tilde{c}) = s^*$.

Let $j \in I$ be such that $\mu_j = s^*$ and $g_j^{s^*} = \pi_{s^*}(\mu, g)$. Let $\hat{S} = \{s_1, \ldots, s_T\}$ be the set of schools for which i has justified envy at μ and assume without loss of generality $s_1 \succ_i s_2 \succ_i \ldots \succ_i s_T$. Note that our constructions of the candidate scenarios $(\tilde{\succ}_{-i}, \tilde{c}_{-i}, \tilde{g}_{-i})$ below varies for different cardinalities of $\hat{S}$.

In the following, for any $\succ_i' \in \mathcal{P}$ and any $s \in S$, we denote the strict lower contour set of $\succ_i'$ at s by $SL_s^{\succ_i'} = \{s' \in S \mid s \succ_i' s'\}$ and the strict upper contour set of $\succ_i'$ at s by $SU_s^{\succ_i'} = \{s' \in S \mid s' \succ_i' s\}$. Notably, the following observations on $\hat{S}$ will be helpful:

- $\hat{S} = \emptyset$ whenever $c_i = 0$, since EDA does not allow for any priority violations for i.
- If $\hat{S} \neq \emptyset$, non-wastefulness of EDA implies that for each $s' \in \hat{S}$, $\pi_{s'}(\mu, g) \neq 0$.
- Since $\hat{S} \subseteq SU_{\mu_i}^{\succ_i}$ and $s^* \in SL_{\mu_i}^{\succ_i}$, $s^* \notin \hat{S}$.

[12] Note that since we assume that $g_i^s \neq g_j^s$ for any $i, j \in I$ and any $s \in S$, it cannot be true that $\pi_{s^*}(\mu, g) = g_i^{s^*}$, when $i \notin \mu_{s^*}$.

Next, for each $t \in \{1, \ldots, T\}$, let $i_t \in \mu_{s_t}$ be such that $g^{s_t}_{i_t} = \pi_{s_t}(\mu, g)$. Collect all such students in $\hat{I} = \{i_1, \ldots, i_T\}$. For each $i_t \in \hat{I}$, in any TP process corresponding to a plausible scenario, there must exist a solved cycle γ such that $i_t k \in \gamma$ for some $k \in I$ and i_t is assigned to s_t when γ is solved. Moreover, solving γ must be the last step in that TP process in which i_t is improved.

Case 3.1: $|\hat{S}| \neq 1$. For now, assume that (C2) is satisfied. At the end of this subcase, we present a slight modification needed in the construction for (C1).

Step 1: We start with the candidate score structure $\tilde{g}_{-i}$:

- let $\tilde{g}^{\mu_i}_i \geq \pi_{\mu_i}(\mu, g) > \tilde{g}^{\mu_i}_j$ and;
- for any $s' \in S \setminus \{\hat{S} \cup \mu_i\}$ let $\tilde{g}^{s'} = g^{s'}$.

Let $i_0 = i_T$ and $s_{T+1} = s_1$. In case that $\hat{S} \neq \emptyset$:

- for each $s_t \in \hat{S}$, let $\tilde{g}^{s_t}$ be such that $\tilde{g}^{s_t}_{i_{t-1}} > \tilde{g}^{s_t}_i > \tilde{g}^{s_t}_{i_t}$ and for all $l \in \mu_{s_t}$ with $l \neq i_t$, let $\tilde{g}^{s_t}_l > \tilde{g}^{s_t}_{i_{t-1}}$.

Next, select an arbitrary $\tilde{c}_{-i}$ and consider the following preferences $\tilde{\succ}_{-i}$:

$$\mu_i \,\tilde{\succ}_j\, s^* \,\tilde{\succ}_j\, s_\emptyset \,\tilde{\succ}_j\, \ldots,$$

$$s_t \,\tilde{\succ}_{i_t}\, s_{t+1} \,\tilde{\succ}_{i_t}\, s_\emptyset \,\tilde{\succ}_{i_t}\, \ldots \quad \forall t \in \{1, \ldots, T\},$$

$$\mu_k \,\tilde{\succ}_k\, s_\emptyset \,\tilde{\succ}_k\, \ldots \quad \forall k \in I \backslash (\hat{I} \cup \{i, j\}),$$

Step 2: As one can easily see, we have $\pi(\mu, (g_i, \tilde{g}_{-i})) = \pi(\mu, g)$. We now show that the constructed scenario $(\tilde{\succ}_{-i}, \tilde{c}_{-i}, \tilde{g}_{-i})$ yields μ under the TP algorithm. We have two cases to consider: First, if $\hat{S} = \emptyset$, we get $DA(\tilde{g}, (\succ_i, \tilde{\succ}_{-i})) = \mu$ and the TP process terminates with μ since there are no cycles $G^*(\mu)$. Second, suppose that $\hat{S} \neq \emptyset$. We describe how we arrive at the corresponding DA outcome: $DA_k(\tilde{g}, (\succ_i, \tilde{\succ}_{-i})) = \mu_k$ for all $k \in I \setminus \hat{I}$ and $DA_{i_t}(\tilde{g}, (\succ_i, \tilde{\succ}_{-i})) = s_{t+1}$ for all $i_t \in \hat{I}$. Note that every student $k \in I \setminus \{i, j\}$ considers her assigned school (μ_k) as the top choice in $\tilde{\succ}_{-i}$ and each such student k gets accepted by her top choice at the first step of the corresponding DA process. Moreover, at some step, student i applies to s_1 and gets tentatively accepted. This triggers a series of rejections. Specifically, for each $t \in \{1, \ldots, T\}$, i_t gets rejected by s_t and applies to s_{t+1} in the next step, causing i_{t+1} being rejected by s_{t+1} and so forth. This rejection chain ends with i_T applying to s_1 which leads i to be rejected by s_1. Thereafter, i applies to all schools in $SU_{\mu_i}^{\succ_i} \setminus SU_{s_1}^{\succ_i}$ and is rejected until finally being accepted by μ_i. At last, j is rejected by μ_i and applies to s^* to which she is finally assigned in DA.

There is a unique cycle $\gamma = \{i_T i_{T-1}, \ldots, i_2 i_1, i_1 i_T\}$ in $G^*(DA(\tilde{g}, (\succ_i, \tilde{\succ}_{-i})))$ which, once solved, produces μ. According to $(\succ_i, \tilde{\succ}_{-i})$, i and j are the only students who do not receive their top choice in μ and therefore the TP algorithm terminates with μ.

Step 3: First, be aware that the outcome $DA(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i}))$ may vary in the position of s^* on $\tilde{\succ}_i$:

- If $s^* \tilde{\succ}_i s_1$, then $DA_i(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i})) = s^*$, $DA_j(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i})) = \mu_i$, $DA_k(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i})) = \mu_k$ for any $k \in I \setminus \{i, j\}$.
- If $s_1 \tilde{\succ}_i s^*$, then $DA_i(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i})) = s^*$, $DA_j(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i})) = \mu_i$, $DA_{i_t}(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i})) = s_{t+1}$ for any $i_t \in \hat{I}$ and $DA_k(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i})) = \mu_k$ for any $k \in I \setminus (\{i, j\} \cup \hat{S})$.

In both instances above s^* has no demand at $DA(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i}))$. As a result, we have that $EDA_i(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i}), \tilde{c}) = s^*$ and thus the argument for (C2) is complete.

Now suppose that (C1) holds. The construction above does not work here generally, since when $\pi_{\mu_i}(\mu, g) = 0$, both i and j get finally assigned to μ_i in $EDA(\tilde{g}, (\succ_i, \tilde{\succ}_{-i}), \tilde{c})$. We make the following adjustments in the construction:

Step 1: Modify the preferences of j to be

$$s^* \tilde{\succ}_j s_\emptyset \tilde{\succ}_j \ldots,$$

and keep all other details of our construction the same as in instance (C2) above.

Step 2 and Step 3: The arguments resemble those in instance (C2) above.

Case 3.2: $|\hat{S}| = 1$. The construction in Case 3.1 does not work here. Specifically, we cannot construct a cycle that consists of students in $\hat{I}$ when $|\hat{I}| = |\hat{S}| = 1$. Cycles will therefore contain students not in $\hat{I}$ and moreover, from (C1) to (C2), we need to alter the identity of students involved in the cycle:

Case 3.2.1: $g_i^{s^*} > \pi_{s^*}(\mu, g)$. That is, (C1) holds and we have $g_i^{s^*} > g_j^{s^*}$.

Step 1: Let $\tilde{g}_{-i}$ be such that

- $\tilde{g}_j^{s_1} > \tilde{g}_i^{s_1} > \tilde{g}_{i_1}^{s_1}$;
- $\tilde{g}_i^{s^*} > \tilde{g}_{i_1}^{s^*} > \tilde{g}_j^{s^*}$;
- $\tilde{g}^{s'} = g^{s'}$ for any $s' \in S \setminus \{s^*, s_1\}$.

Now, let $\tilde{c}_{-i}$ be such that $\tilde{c}_{i_1} = 0$[13] and consider the following profile $\tilde{\succ}_{-i}$:

$$s^* \mathrel{\tilde{\succ}_j} s_1 \mathrel{\tilde{\succ}_j} s_\emptyset \ldots,$$

$$s_1 \mathrel{\tilde{\succ}_{i_1}} s^* \mathrel{\tilde{\succ}_{i_1}} s_\emptyset \ \ldots,$$

$$\mu_k \mathrel{\tilde{\succ}_k} s_\emptyset \mathrel{\tilde{\succ}_k} \ \ldots \ \ \forall k \in I \backslash \{i, j, i_1\}.$$

Step 2: First, it is easily checked $\pi(\mu, (g_i, \tilde{g}_{-i})) = \pi(\mu, g)$. Following a similar application procedure as in Case 3.1, the DA algorithm leads to $DA_j(\tilde{g}, (\succ_i, \tilde{\succ}_{-i})) = s_1$, $DA_{i_1}(\tilde{g}, (\succ_i, \tilde{\succ}_{-i})) = s^*$ and $DA_k(\tilde{g}, (\succ_i, \tilde{\succ}_{-i})) = \mu_k$ for all $k \in I \setminus \{j, i_1\}$. There is a unique cycle $\gamma = \{i_1 j, j i_1\}$ in

[13] It is worth mentioning that this is the only place in the proof of Theorem 3.1, where we need a scenario where a student does not consent.

$G^*(DA(\tilde{g}, (\succ_i, \tilde{\succ}_{-i})))$ and once this cycle is solved we obtain μ. In this instance, all students except i receive their top choice in μ. The TP algorithm thus terminates and $EDA(\tilde{g}, (\succ_i, \tilde{\succ}_{-i}), \tilde{c}) = \mu$.

Step 3: Note that the DA algorithm arrives at $DA_i(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i})) = s^*$, $DA_j(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i})) = s_1$, $DA_{i_1}(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i})) = s_\emptyset$ and $DA_k(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i})) = \mu_k$ for all $k \in I \backslash \{i, j, i_1\}$. Also, j is not eligible for s^* since $\tilde{c}_{i_1} = 0$. Therefore, we cannot add ji to the graph and thus there is no cycle in $G^*(DA(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i})))$. In conclusion, $EDA_i(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i}), \tilde{c}) = s^*$.

Case 3.2.2: $\pi_{\mu_i}(\mu, g) \neq 0$ and $g_i^{s^*} < \pi_{s^*}(\mu, g)$ That is, (C2) holds and we thus have $g_i^{s^*} < g_j^{s^*}$.

Case 3.2.2.1: There exists $s' \in S \setminus \{s_1, \mu_i, s^*\}$ such that $\pi_{s'}(\mu, g) \neq 0$. Pick an arbitrary such s' and denote with j' the student who has the lowest score among all students being assigned to s' under μ.

Step 1: Let $\tilde{g}_{-i}$ be such that

- $\tilde{g}_{j'}^{s_1} > \tilde{g}_i^{s_1} > \tilde{g}_{i_1}^{s_1}$;
- $\tilde{g}_{i_1}^{s'} > \tilde{g}_{j'}^{s'}$;
- $\tilde{g}_i^{\mu_i} > g_j^{\mu_i}$;
- $\tilde{g}^{s''} = g^{s''}$ for any $s'' \in S \setminus \{s_1, \mu_i, s'\}$.

Next, fix an arbitrary $\tilde{c}_{-i}$ and consider the following profile $\tilde{\succ}_{-i}$:

$$\mu_i \,\tilde{\succ}_j\, s^* \,\tilde{\succ}_j\, s_\emptyset \,\tilde{\succ}_j\, \ldots,$$

$$s_1 \,\tilde{\succ}_{i_1}\, s' \,\tilde{\succ}_{i_1}\, s_\emptyset \,\tilde{\succ}_{i_1}\, \ldots,$$

$$s' \,\tilde{\succ}_{j'}\, s_1 \,\tilde{\succ}_{j'}\, s_\emptyset \,\tilde{\succ}_{j'}\, \ldots,$$

$$\mu_k \,\tilde{\succ}_k\, s_\emptyset \,\tilde{\succ}_k\, \ldots \quad \forall k \in I \backslash \{i, i_1, j, j'\}.$$

Step 2 and Step 3: We omit the arguments for Step 2 and Step 3, since they are similar to those in Case 3.1.

Case 3.2.2.2: There does not exist $s' \in S \setminus \{s_1, \mu_i, s^*\}$ such that $\pi_{s'}(\mu, g) \neq 0$. Since $\pi_{s^*}(\mu, g) \neq 0$ and $\pi_{\mu_i}(\mu, g) \neq 0$, there are only three schools, namely s_1, μ_i, s^*, which exhaust their capacity under μ. In this last subcase, we show that statement (2) is satisfied.

We first argue that in any plausible scenario, there is only one top priority cycle and it consists of i_1 and one student assigned to s^*. To start, since i has justified envy for s_1 at μ, there exists a cycle containing i_1 that is solved in the EDA process. Second, by non-wastefulness of EDA, we know that if a school is contained in one solved cycle, it exhausts its capacity under the final matching. Recall that only s_1, μ_i, s^* exhaust their capacity at μ. Thus, the candidate student for forming a cycle can only be assigned to s^*. Therefore, we can construct exactly one cycle with i_1 and some $l \in \mu_{s^*}$.

Now select any $(\tilde{\succ}_{-i}, \tilde{c}_{-i}, \tilde{g}_{-i}) \in \mathcal{I}(\mu, \succ_i, c_i)$. Since $g_i^{s^*} < \pi_{s^*}(\mu, g)$ and by our arguments made above, it must be true $\tilde{g}_{i_1}^{s^*} > \tilde{g}_l^{s^*} > g_i^{s^*}$ and $DA_{i_1}(\tilde{g}, (\succ_i, \tilde{\succ}_{-i})) = s^*$. However, this implies that i will be rejected by s^* under DA when she reports $\tilde{\succ}_i$. As a result, we can claim that $DA_i(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i})) = EDA_i(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i}), \tilde{c}) = \mu_i$ and statement (2) thus holds.

This completes the proof for the case in which $\tilde{S}$ is a singleton. To finish the proof, suppose now that $\tilde{S}$ contains multiple elements.

We denote the top ranked school on $\tilde{\succ}_i$ among all schools in $\tilde{S}$ by s_1. Specifically, let $\succ_i^1$ be such that $s_1 \succ_i^1 \mu_i$ and $s \succ_i^1 s'$ if $s \succ_i s'$ for all $S \setminus \{s_1\}$. Since s_1 is the only permuted school on $\succ_i^1$ compared to $\succ_i$, we can apply the arguments above (for singleton $\tilde{S}$) to $\succ_i^1$. Here, we distinguish two cases. In the first case, suppose that the observation $(\mu, \pi(\mu, g))$ is such that statement (1) holds for $\succ_i^1$. That is, we find $(\succ_{-i}^1, c_{-i}^1, g_{-i}^1) \in \mathcal{I}(\mu, \succ_i^1, c_i)$ such that $EDA_i(g^1, (\succ_i^1, \succ_{-i}^1), c^1) = s_1$. Note that all our constructions above satisfy that $DA_i(g^1, (\succ_i^1, \succ_{-i}^1)) = EDA_i(g^1, (\succ_i^1, \succ_{-i}^1), c^1) = s_1$.

Since $SU_{s_1}^{\tilde{\succ}_i} = SU_{s_1}^{\succ_i^1}$, we obtain $DA_i(g^1, (\tilde{\succ}_i, \succ_{-i}^1)) = EDA_i(g^1, (\tilde{\succ}_i, \succ_{-i}^1), c^1) = s_1$. Thus, we can conclude that statement (1) also holds for misreport $\tilde{\succ}_i$ for the first case. In the second case, suppose that the observation $(\mu, \pi(\mu, g))$ falls into the case where statement (2) holds for $\succ_i^1$. Then, we need further consider the second ranked school among $\tilde{S}$ on $\tilde{\succ}$, denoted by s_2.

Specifically, we construct $\succ_i^2$ such that $s_1 \succ_i^2 s_2 \succ_i^2 \mu_i$ and $s \succ_i^2 s'$ if $s \succ_i s'$ for all $s, s' \in S \setminus \{s_1, s_2\}$. Since we assume that $\succ_i^1$ has no influence on the result at all, we can again apply the arguments for the singleton case to $\succ_i^2$. That is, we consider whether statement (1) or statement (2) applies to $\succ_i^2$. If statement (1) holds for $\succ_i^2$, then as explained above we can conclude that statement (1) holds for $\tilde{\succ}_i$. Otherwise, we further consider the third ranked school among $\tilde{S}$ on $\tilde{\succ}$. In the following, we iteratively apply the above arguments by adding a new school from $\tilde{S}$ through each iteration. Once we arrive at a step where statement (1) holds, we stop and conclude that statement (1) holds for $\tilde{\succ}_i$. On the contrary, if for all schools in $\tilde{S}$ the observation (2) holds, then we conclude that statement (2) holds for the misreport $\tilde{\succ}_i$. □

We move to the final class of misreports in which all schools that are truly less preferred to μ_i still rank lower than μ_i. That is, in the rest of the proof, we consider $\tilde{\succ}_i \in \mathcal{P}$ such that $SU_{\mu_i}^{\tilde{\succ}_i} \subseteq SU_{\mu_i}^{\succ_i}$ and for which there exists $s, s' \in SL_{\mu_i}^{\succ_i}$ such that $s \succ_i s'$ and $s' \tilde{\succ}_i s$. Our strategy is to show that if a student could have been improved upon truth through such a misreport $\tilde{\succ}_i$ in a plausible scenario, then the misreport could also have made the misreporting student worse off in another plausible scenario.

Before we formally show the above argument, we provide three auxiliary results. The first result states that a student can improve upon μ_i via reporting $\tilde{\succ}_i$ only if μ_i is not her SOSM assignment under true preferences. Throughout the remaining discussion, we fix some $(\succ'_{-i}, c'_{-i}, g'_{-i}) \in \mathcal{I}(\mu, \succ_i, c_i)$. Also, for any $\succ'_i \in \mathcal{P}$ and any $s \in S$, we denote the weak upper contour set of $\succ'_i$ at s by $U_s^{\succ'_i} = \{s' \in S \,|\, s' \succeq'_i s\}$.

Lemma 3.4. *If $EDA_i(g', (\tilde{\succ}_i, \succ'_{-i}), c') \succ_i \mu_i$, then $\mu_i \succ_i DA_i(g', (\succ_i, \succ'_{-i}))$.*

Proof. EDA guarantees that $\mu_i \succeq_i DA_i(g', (\succ_i, \succ'_{-i}))$. We now prove the contrapositive statement: If $DA_i(g', (\succ_i, \succ'_{-i})) = \mu_i$, then $EDA_i(g', (\tilde{\succ}_i, \succ'_{-i}), c') = \mu_i$. Towards this goal, construct $\hat{\succ}_i \in \mathcal{P}$ such that

(D1) for each $s_1, s_2 \in L^{\succ_i}_{\mu_i}$, $s_1 \hat{\succ}_i s_2$ if and only if $s_1 \succ_i s_2$;

(D2) for each $s_3, s_4 \in U^{\tilde{\succ}_i}_{\mu_i}$, $s_3 \hat{\succ}_i s_4$ if and only if $s_3 \tilde{\succ}_i s_4$ and;

(D3) for all $s \in SU^{\succ_i}_{\mu_i} \setminus SU^{\tilde{\succ}_i}_{\mu_i}$, $s \in SL^{\hat{\succ}_i}_{\mu_i}$.

Since $SU^{\tilde{\succ}_i}_{\mu_i} \subseteq SU^{\succ_i}_{\mu_i}$ and $SU^{\succ_i}_{\mu_i} \cap SL^{\succ_i}_{\mu_i} = \emptyset$, one obtains $L^{\succ_i}_{\mu_i} \cap U^{\tilde{\succ}_i}_{\mu_i} = \{\mu_i\}$. Therefore, (D1) and (D2) consider distinct sets of schools, and more concretely, (D1) - (D3) defines the full order of $\hat{\succ}_i$. With (D1) we can immediately apply Proposition 3.1 such that we reach $EDA_i(g', (\hat{\succ}_i, \succ'_{-i}), c') = EDA_i(g', (\succ_i, \succ'_{-i}), c') = \mu_i$. Clearly, (D1) means that $\hat{\succ}_i$ is a monotonic transformation of $\succ_i$ at μ_i. Thus, according to Lemma 3.1 we have $DA_i(g', (\hat{\succ}_i, \succ'_{-i})) = DA_i(g', (\succ_i, \succ'_{-i}))$. Thus, we obtain $EDA_i(g', (\hat{\succ}_i, \succ'_{-i}), c') = DA_i(g', (\hat{\succ}_i, \succ'_{-i})) = \mu_i$. Moreover, (D2) and (D3) jointly ensure that $DA(g'_{-i}, (\tilde{\succ}_i, \succ'_{-i})) = DA(g'_{-i}, (\hat{\succ}_i, \succ'_{-i}))$.

Now, note that $DA_i(g', (\hat{\succ}_i, \succ'_{-i})) = EDA_i(g', (\hat{\succ}_i, \succ'_{-i}), c')$, which implies that i cannot be improved to any school more preferred than μ_i on $\hat{\succ}_i$ by EDA. Since by (D2) and (D3) we know that $\tilde{\succ}_i$ and $\hat{\succ}_i$ share the same ranking for schools more preferred than μ_i. Thus, it follows that $DA_i(g', (\tilde{\succ}_i, \succ'_{-i})) = EDA_i(g', (\tilde{\succ}_i, \succ'_{-i}), c')$. Thus, we obtain $EDA_i(g', (\tilde{\succ}_i, \succ'_{-i}), c') = \mu_i$. This completes the proof. □

Next, we show that if a student could improve upon μ_i via a misreport $\tilde{\succ}_i$, then at least one school satisfies the following three conditions: First, the student prefers her assignment to this school. Second, the relative ranking of this school is lowered under the misreport compared to truth-telling. Third, the student's score at this school is higher than this school's cutoff.

Let $S' = \{s \in SL_{\mu_i}^{\succ_i} \mid \exists \tilde{s} \in SL_{\mu_i}^{\succ_i} : s \succ_i \tilde{s} \text{ and } \tilde{s} \tilde{\succ}_i s\}$. Recall that we now consider misreport $\tilde{\succ}_i$ of the last class where $SU_{\mu_i}^{\tilde{\succ}_i} \subseteq SU_{\mu_i}^{\succ_i}$. According to Proposition 3.1, we know that S' must be non-empty since $EDA_i(g', (\tilde{\succ}_i, \succ'_{-i}), c') \neq EDA_i(g', (\succ_i, \succ'_{-i}), c')$.

Lemma 3.5. *If* $EDA_i(g', (\tilde{\succ}_i, \succ'_{-i}), c') \succ_i \mu_i$*, then there exists* $s' \in S'$ *such that* $g_i^{s'} > \pi_{s'}(\mu, g) > 0$.

Proof. We prove by means of contradiction. That is, given $EDA_i(g', (\tilde{\succ}_i, \succ'_{-i}), c') \succ_i \mu_i$, we suppose that for each $\hat{s} \in S'$ either $\pi_{\hat{s}}(\mu, g) > g_i^{\hat{s}}$ or $\pi_{\hat{s}}(\mu, g) = 0$. We aim at a contradiction by showing that we arrive at $EDA_i(g', (\tilde{\succ}_i, \succ'_{-i}), c') = \mu_i$.

Select any TP process with input $(g', (\succ_i, \succ'_{-i}), c')$ and denote it by $pTP^{\succ}$. Let $EDA^t(\succ)$ be the outcome of the t_{th} step in $pTP^{\succ}$. Since $EDA_i(g', (\tilde{\succ}_i, \succ'_{-i}), c') \succ_i \mu_i$, by Lemma 3.4 we have that $\mu_i \succ_i DA_i(g', (\succ_i, \succ'_{-i}))$. Then, we collect the set of schools to which i is (temporarily) assigned during $pTP^{\succ}$ in $S_i = \{\hat{s} \in S \mid \exists t \in \mathbb{N} : EDA_i^t(\succ) = \hat{s}\}$. As mentioned before, during the process of the TP algorithm, scores of assigned students are weakly decreasing at each school from step to step. Thus, for any $s'' \in S_i$ we have $g_i^{s''} \geq \pi_{s''}(\mu, g)$. Also, schools in S_i must have positive cutoffs. Therefore, by assumption of S', we have $S' \cap S_i = \emptyset$. Hence, we can use the following two features:

1. for any $s' \in S_i$, $SU_{s'}^{\tilde{\succ}_i} \subseteq SU_{s'}^{\succ_i}$; and

2. for any $s', s'' \in S_i$, $s' \tilde{\succ} s''$ if and only if $s' \succ_i s''$.

In the following, we first assume that $c_i = 1$. Under this assumption, we claim that with the above two features of $\succ_i$ and $\tilde{\succ}_i$, we can implement the algorithm in proof of Proposition 3.1 with profiles $(g', (\tilde{\succ}_i, \succ'_{-i}), c')$ to construct a process $pTP^{\tilde{\succ}}$ that yields the same outcome as $pTP^{\succ}$ does. Concretely, compared to the misreports studied in Proposition 3.1, the misreport $\tilde{\succ}_i$ considered here allows for additional permutations which move some $s \in L_{\mu_i}^{\succ_i}$ from above some $s'' \in S_i$ to below. Note that the first feature above ensures that all cycles solved in $pTP^{\succ}$ which do not involve i are no different from those already covered by the algorithm in Proposition 3.1. Concretely, although agent i demands some additional schools in S', since i consents and i has a lower score than the cutoff at each of these schools, the additional demand of student i does not change the formation of cycles at each step of the algorithm. Next, note that for cycles solved in $pTP^{\succ}$ which contain i, the second feature above guarantees that such a cycle can still be solved at the corresponding step. Therefore, we arrive at $EDA_i(g', (\tilde{\succ}_i, \succ'_{-i}), c') = EDA_i(g', (\succ_i, \succ'_{-i}), c') = \mu_i$, which contradicts our initial assumption.

Next, assume $c_i = 0$. Recall that in Proposition 3.1, we extend the conclusions to the case where $c_i = 0$. Here, we can also conclude $EDA_i(g', (\tilde{\succ}_i, \succ'_{-i}), c') = EDA_i(g', (\succ_i, \succ'_{-i}), c')$ with the same line of reasoning. Again, we reach the desired contradiction. □

From now on, assume that $\mu_i \succ_i DA_i(g'_{-i}, (\succ_i, \succ'_{-i}))$. The reason for this assumption is that, as shown in Lemma 3.5, misreporting $\tilde{\succ}_i$ could potentially be profitable only if this assumption is satisfied. If reporting $\tilde{\succ}_i$ is not profitable at all, then the agent will never regret telling the truth through such a misreport. Notably, this assumption also implies that we have $\pi_{\mu_i}(\mu, g) \neq 0$ in the rest of the proofs. Moreover, Lemma 3.5 shows that there exists a maximal and non-empty set $S_1 \subseteq S'$ such that $s' \in S_1$ if and only if $g_i^{s'} > \pi_{s'}(\mu, g) > 0$. For the rest of the proof, let $s^* \in S_1$ be such that $s^* \succeq_i s'$ for any $s' \in S_1$. Furthermore, we collect in $S_2 = \{r' \in L_{\mu_i}^{\succ_i} \mid s^* \succ_i r', r' \tilde{\succ}_i s^*\}$ and denote with $r^* \in S_2$ the school such that $r^* \tilde{\succeq}_i r'$ for any $r' \in S_2$. For our construction for the last class of misreports, we rely on the following property of r^*.

Lemma 3.6. *If $EDA_i(g', (\tilde{\succ}_i, \succ'_{-i}), c') \succ_i \mu_i$, then $\pi_{r^*}(\mu, g) \neq 0$.*

Proof. We aim to show the contrapositive statement. That is, given $\pi_{r^*}(\mu, g) = 0$, we prove that $\mu_i \succeq_i EDA_i(g', (\tilde{\succ}_i, , \succ'_{-i}), c')$. Let $DA(g', (\succ_i, \succ'_{-i})) = \nu_i$. Since we assume $\pi_{\mu_i}(\mu, g) \neq 0$, it follows immediately $\pi_{\nu_i}(\mu, g) \neq 0$. That is, $\nu_i \neq r^*$. In the following, we consider two cases that are distinguished by the relative ranking of r^* and ν_i on $\tilde{\succ}_i$.

In the first case, suppose $\nu_i \tilde{\succ}_i r^*$. We claim $EDA_i(g', (\tilde{\succ}_i, \succ'_{-i}), c') = \mu_i$ here and will use similar arguments as in the proof of Lemma 3.5. Recall that we have $SU_{\mu_i}^{\tilde{\succ}_i} \subseteq SU_{\mu_i}^{\succ_i}$. Together with the assumption $\nu_i \tilde{\succ}_i r^*$ and the fact $\mu_i \succeq_i \nu_i$, we can infer $\mu_i \tilde{\succ}_i r^*$. Now, select an arbitrary TP process with input $(g', (\succ_i, \succ'_{-i}), c')$, denoted by $pTP^{\succ'}$ and let $EDA^t(\succ')$ be the outcome of the t_{th} step in $pTP^{\succ'}$. Also, let $S'_i = \{s' \in S \mid \exists t \in \mathbb{N} : EDA^t_i(\succ') = s'\}$ be the set of schools to which i is (temporarily) assigned during $pTP^{\succ'}$. As argued before, for each $s' \in S'_i$, it is true $g_i^{s'} > \pi_{s'}(\mu, g) > 0$. Note that $\nu_i \in S'_i$ and by assumption $\nu_i \tilde{\succ}_i r^*$, the selection of r^* ensures that:

1. for any $s' \in S'_i, SU_{s'}^{\tilde{\succ}_i} \subseteq SU_{s'}^{\succ_i}$; and
2. for any $s', s'' \in S'_i, s' \tilde{\succ} s''$ if and only if $s' \succ_i s''$.

Notably, as argued in Lemma 3.5, this leads to $EDA_i(g', (\tilde{\succ}_i, \succ'_{-i}), c') = \mu_i$.

In the second case, suppose $r^* \tilde{\succ}_i \nu_i$. We show $\mu_i \succ_i EDA_i(g', (\tilde{\succ}_i, \succ'_{-i}), c')$ here. Towards this goal, we first argue $SU_{r^*}^{\tilde{\succ}_i} \subseteq SU_{\nu_i}^{\succ_i}$. By contradiction, suppose that there exists $s' \in S$ such that $r' \in SU_{r^*}^{\tilde{\succ}_i}$ and $r' \notin SU_{\nu_i}^{\succ_i}$. Then, we know that (1) $\nu_i \succ_i r'$, (2) $r' \tilde{\succ}_i \nu_i$ and (3) $r' \tilde{\succ}_i r^*$. Since $g_i^{\nu_i} > \pi_{\nu_i}(\mu, g) > 0$, by (1) and (2) we can infer $\nu_i \in S_1$. Thus, the selection of s^* ensures that $s^* \succeq_i \nu_i$, which combined with (1) shows $s^* \succ_i r'$. Moreover, from (3) and the construction of S_2 we have $r' \tilde{\succ}_i r^* \tilde{\succ}_i s^*$. Note that $s^* \succeq_i \nu_i$ and $r' \tilde{\succ}_i r^* \tilde{\succ}_i s^*$, we reach a contradiction to how r^* is selected. Thus, we have $SU_{r^*}^{\tilde{\succ}_i} \subseteq SU_{\nu_i}^{\succ_i}$. Next, since by assumption r^* has vacant seat at $EDA(g', (\succ_i, \succ'_{-i}), c')$, it also has vacant seat at $DA(g', (\succ_i, \succ'_{-i}))$. With the two findings above, we can implement the arguments in Case 2 of Lemma 3.3 and conclude that $DA_i(g', (\tilde{\succ}_i, \succ'_{-i})) = r^*$ is underdemanded. Thus, student i cannot improve her assignment above r^* and we reach $EDA_i(g', (\tilde{\succ}_i, \succ'_{-i}), c') = r^*$. Since $\mu_i \succ_i r^*$, this completes the proof. □

We now show the formal arguments for the last class of misreports. Concretely, we show that when i reports $\tilde{\succ}_i$, she could have been worse off by being assigned to r^*.

Lemma 3.7. *If $EDA_i(g', (\tilde{\succ}_i, \succ'_{-i}), c') \succ_i \mu_i$, there exists $(\tilde{\succ}_{-i}, \tilde{c}_{-i}, \tilde{g}_{-i}) \in \mathcal{I}(\mu, \succ_i, c_i)$ such that $\mu_i \succ_i EDA_i(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i}), \tilde{c}) = r^*$.*

Proof. Note that by Lemma 3.6, we only need to construct such a scenario for case $\pi_{r^*}(\mu, g) > 0$. Similar as in the proof of Lemma 3.3, we go through a series of steps to show the desired result:

Step 1: We construct a candidate scenario $(\tilde{\succ}_{-i}, \tilde{c}_{-i}, \tilde{g}_{-i})$.

Step 2: We show that $(\tilde{\succ}_{-i}, \tilde{c}_{-i}, \tilde{g}_{-i}) \in \mathcal{I}(\mu, \succ_i, c_i)$.

Step 3: We argue that $EDA_i(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i}), \tilde{c}) = r^*$.

Recall that $s^* \in S_1$ is the school that ranks highest on $\succ_i$ among all schools in S_1. Let $j \in I$ be such that $\mu_j = r^*$, and let $l \in I$ be such that $\mu_l = s^*$ and $g_l^{s^*} = \pi_{s^*}(\mu, g)$. Moreover, consider the set $\bar{S} = \{s' \in SU_{s^*}^{\succ_i} \mid g_i^{s'} > \pi_{s'}(\mu, g)\}$ and denote $\bar{S} = \{s_1, s_2, \ldots, s_T\}$. Without loss of generality, let $s_1 \succ_i s_2 \succ_i \ldots \succ_i s_T$. Since $r^* \in SL_{s^*}^{\succ_i}$, we know that $r^* \notin \bar{S}$. For each $t \in \{1, ..., T\}$, denote the student with the lowest score assigned to s_t in μ by i_t and collect all such students in $\bar{I} = \{i_1, \ldots, i_T\}$. Similar to Lemma 3.3, we make a case distinction based on different observations. However, since we already know that $\pi_{\mu_i}(\mu, g) \neq 0$ and $\pi_{r^*}(\mu, g) \neq 0$, it suffices to consider different cardinalities of $\bar{S}$.

Case 1: $|\bar{S}| \neq 1$. *Step 1:* We start with the candidate score structure. Let $\tilde{g}_{-i}$ be such that

- $\tilde{g}_l^{\mu_i} > \tilde{g}_j^{\mu_i} > \tilde{g}_i^{\mu_i}$; and for any $k \in \mu_{\mu_i} \setminus \{i\}, g_k^{\mu_i} > g_l^{\mu_i}$;
- $g_i^{s^*} > g_l^{s^*}$; and for any $k \in \mu_{s^*} \setminus \{l\}, \tilde{g}_k^{s^*} > \tilde{g}_i^{s^*}$;
- $\tilde{g}^{s'} = g^{s'}$ for any $s' \in S \setminus \{s_1, \ldots, s_T, \mu_i, s^*\}$.

Let $i_0 = i_T$ and $s_{T+1} = s_1$. In case that $\bar{S} \neq \emptyset$, for any $s_t \in \bar{S}$:

- $\tilde{g}_{i_{t-1}}^{s_t} > \tilde{g}_i^{s_t} > \tilde{g}_{i_t}^{s_t}$; and for any $k \in \mu_{s_t} \setminus \{i_t\}, \tilde{g}_k^{s_t} > \tilde{g}_{i_{t-1}}^{s_t}$.

Next, we specify $\tilde{c}_{-i}$ such that for all $i' \in I \setminus \{i\}$ it holds that $\tilde{c}_{i'} = 1$ and consider preference profile $\tilde{\succ}_{-i} \in \mathcal{P}_{-i}$:

$$s_t \mathrel{\tilde{\succ}_{i_t}} s_{t+1} \mathrel{\tilde{\succ}_{i_t}} s_\emptyset \mathrel{\tilde{\succ}_{i_t}} \ldots \quad \forall t \in \{1, \ldots, T\},$$

$$s^* \mathrel{\tilde{\succ}_l} \mu_i \mathrel{\tilde{\succ}_l} s_\emptyset \mathrel{\tilde{\succ}_l} \ldots,$$

$$\mu_i \mathrel{\tilde{\succ}_j} r^* \mathrel{\tilde{\succ}_j} s_\emptyset \mathrel{\tilde{\succ}_j} \ldots,$$

$$\mu_k \mathrel{\tilde{\succ}_k} s_\emptyset \mathrel{\tilde{\succ}_k} \ldots \quad \forall k \in I \backslash (\bar{I} \cup \{i, j, l\}).$$

Step 2: It is easily checked that $\pi(\mu, (g_i, \tilde{g}_{-i})) = \pi(\mu, g)$. Next, we show that DA leads to $DA_i(\tilde{g}, (\succ_i, \tilde{\succ}_{-i})) = s^*, DA_j(\tilde{g}, (\succ_i, \tilde{\succ}_{-i})) = r^*, DA_l(\tilde{g}, (\succ_i, \tilde{\succ}_{-i})) = \mu_i, DA_{i_t}(\tilde{g}, (\succ_i, \tilde{\succ}_{-i})) = s_{t+1}$ for each $t \in \{1, \ldots, T\}$ and $DA_k(\tilde{g}, (\succ_i, \tilde{\succ}_{-i})) = \mu_k$ for $k \in I \backslash (\bar{I} \cup \{i, j, l\})$.

Consider the corresponding application process of DA under the constructed scenario. For each student $k \in I \setminus (\bar{I} \cup \{i, j, l\})$, either μ_k is in $U_{s^*}^{\succ_i}$ and k is among the top q_{μ_k} scored students at μ_k; or μ_k is in $SL_{s^*}^{\succ_i}$ and at most q_{μ_k} students apply to μ_k according to $(\succ_i, \tilde{\succ}_{-i})$. Therefore, at the first step of the DA process, each such k applies to μ_k and is finally assigned to μ_k. Furthermore, the following students will be tentatively accepted at the first step:

- student j applies to μ_i,
- student l applies to s^*,
- for all $t \in \{1, \ldots, T\}$, student i_t applies to s_t.

At the first step of the application process, also i applies to her top choice. If i's top choice is not s_1, let $t_1 \in \mathbb{N}$ be the step in the application process, in which i applies to s_1. In all the previous steps $t < t_1$, student i is rejected at each school she proposes to. However, at step t_1 student i is tentatively accepted at s_1 and student i_1 is rejected. In fact, being initial for student i_1 being rejected at s_1, student i induces a sequence of rejections. This sequence ends in student i being rejected at s_1 and for all $t \in \{2, \ldots, T\}$ student i_t is rejected from school s_t in favor of student i_{t-1} at step $t_1 + t$. Finally at step $t_1 + T$, student i_T applies to s_1 such that student i gets rejected. In the following steps, only student i makes new applications until she gets accepted. Precisely, student i proposes to each remaining school in $SU_{s^*}^{\tilde{\succ}_i}$ that she has not yet proposed to and gets immediately rejected at each of these schools. Finally, student i applies to s^* and gets accepted in favor of student l. Student l being rejected at s^* applies now to μ_i such that student j gets rejected. Next, j applies to r^* and gets accepted. Notice that at this step no student is rejected, the application process ends and the algorithm terminates.

Starting with the final outcome $DA_i(\tilde{g}, (\succ_i, \tilde{\succ}_{-i}))$ of the just described process, we now show that the cycle selection under a TP process ends in the observed matching μ. Since j is permanently matched in $DA(\tilde{g}, (\succ_i, \tilde{\succ}_{-i}))$ and $\tilde{c}_j = 1$, we know that $G^*(DA(\tilde{g}, (\succ_i, \tilde{\succ}_{-i})))$ contains the cycle $\gamma^1 = \{il, li\}$ and solving it yields $EDA^1(\tilde{g}, (\succ_i, \tilde{\succ}_{-i}), \tilde{c}) = \gamma^1 \circ DA(\tilde{g}, (\succ_i, \tilde{\succ}_{-i}))$, where compared to in $DA(\tilde{g}, (\succ_i, \tilde{\succ}_{-i}))$, only i and l switch their assignments.

Next, since $c_i = 1$ and i is permanently matched to μ_i in $EDA^1(\tilde{g}, (\succ_i, \tilde{\succ}_{-i}), \tilde{c})$, whenever $\bar{S}$ is non-empty, $G^*(EDA^1(\tilde{g}, (\succ_i, \tilde{\succ}_{-i}), \tilde{c}))$ contains a unique cycle

$$\gamma^2 = \{i_T i_{T-1}, i_{T-1} i_{T-2}, \ldots, i_{t+1} i_t, \ldots i_2 i_1, i_1 i_T\}$$

which once solved yields matching μ. Since all students except i and j get their top-choice, and both i, j are permanently matched, there is no cycle in $G^*(\mu)$. Therefore, $EDA(\tilde{g}, (\succ_i, \tilde{\succ}_{-i}), \tilde{c}) = \mu$.

Step 3: Reviewing the application process above, we get $DA_i(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i})) = r^*$. Moreover, note that apart from the students who are matched with school r^* at $DA(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i}))$, student j is the only one who ranks r^* above $s_\emptyset$ in $\tilde{\succ}_{-i}$. However, notice that $DA_j(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i})) = \mu_i \tilde{\succ}_j r^*$ and thus school r^* is underdemanded in $DA(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i}))$. As a result, i is permanently matched with r^* at $DA(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i}))$, which implies $EDA_i(\tilde{g}, (\tilde{\succ}_i, \tilde{\succ}_{-i}), \tilde{c}) = r^*$. This completes the proof for Case 1.

Case 2: $|\bar{S}| = 1$. *Step 1:* Let $\tilde{g}_{-i}$ be such that

- $\tilde{g}^{s_1}_{l} > \tilde{g}^{s_1}_{i} > \tilde{g}^{s_1}_{i_1}$; and for all $k \in \mu_{s_1} \setminus \{i_1\}, \tilde{g}^{s_1}_{k} > \tilde{g}^{s_1}_{l}$;
- $\tilde{g}^{\mu_i}_{i_1} > \tilde{g}^{\mu_i}_{j} > \tilde{g}^{\mu_i}_{i}$; and for all $k \in \mu_{\mu_i} \setminus \{i\}, \tilde{g}^{\mu_i}_{k} > \tilde{g}^{\mu_i}_{i_1}$;
- $\tilde{g}^{s^*}_{i} > \tilde{g}^{s^*}_{i_1} > \tilde{g}^{s^*}_{l}$; and for all $k \in \mu_s \setminus \{l\}, \tilde{g}^{s^*}_{k} > \tilde{g}^{s^*}_{i}$;
- $\tilde{g}^{s'} = g^{s'}$ for any $s' \in S \setminus \{s_1, \mu_i, s^*\}$.

Also, let $\tilde{c}_{-i}$ be such that for all $i' \in I \setminus \{i\}$ it holds that $\tilde{c}_{i'} = 1$ and consider preference profile $\tilde{\succ}_{-i} \in \mathcal{P}_{-i}$:

$$s_1 \,\tilde{\succ}_{i_1}\, s \,\tilde{\succ}_{i_1}\, \mu_i \,\tilde{\succ}_{i_1}\, s_\emptyset \,\tilde{\succ}_{i_1} \,\ldots,$$

$$s^* \,\tilde{\succ}_{l}\, s_1 \,\tilde{\succ}_{l}\, s_\emptyset \,\tilde{\succ}_{l} \,\ldots,$$

$$\mu_i \,\tilde{\succ}_{j}\, r^* \,\tilde{\succ}_{j}\, s_\emptyset \,\tilde{\succ}_{j} \,\ldots,$$

$$\mu_k \,\tilde{\succ}_{k}\, s_\emptyset \,\tilde{\succ}_{k} \,\ldots \quad \forall k \in I \setminus \{i, j, l, i_1\}.$$

Step 2 and Step 3: We can resemble the arguments in Step 2 and Step 3 for Case 1 to conclude that i is worse off by being finally assigned to r^* in this constructed scenario. □

Since the conclusion holds for any observation, any student and any problem, we conclude that EDA is regret-free truth-telling.

3.D Proof of Proposition 3.2

With a similar technique as in the proof of Proposition 1 in Fernandez (2020), we now show that any non-truthful report is regretted through the truth at some observation. Throughout the discussion, fix an arbitrary problem $(I, S, q, g, \succ, c)$ and fix an arbitrary $i \in I$. We divide the set of possible misreports into three exhaustive cases. In each case, we consider an arbitrary misreport $\succ_i'$. We then construct an observation following $\succ_i'$ such that the truth $\succ_i$ would have granted i a weakly better assignment in any plausible scenario. Moreover, there exists at least one plausible scenario in which the improvement is strict.

Case 1 Suppose that for $\succ_i'$ there exists $s \in S$ such that $s_\emptyset \succ_i s$ and $s \succ_i' s_\emptyset$. Let i submit $\succ_i'$ and consider the pair $(\mu, \pi(\mu, g))$ such that $\mu_i = s$ and $g_i^{s'} < \pi_{s'}(\mu, g)$ for all $s' \in SU_s^{\succ_i'}$. At first, we show that $\mu \in \mathcal{M}|_{(\succ_i', c_i)}$ by constructing $(\tilde{\succ}_{-i}, \tilde{c}_{-i}, \tilde{g}_{-i})$ that leads to $(\mu, \pi(\mu, g))$: That is, we show that $(\mu, \pi(\mu, g))$ is an observation under EDA. Let $\tilde{g}_{-i}$ be such that, for each $s' \in SU_{\mu_i}^{\succ_i'}$, each student in $\mu_{s'}$ is among the top $q_{s'}$'s scored students at school s'. Let i rank highest on $\tilde{g}^s$ and suppose that the remaining scores are arbitrary. Let $\tilde{\succ}_{-i}$ be such that for each $j \in I \setminus \{i\}$, $\tilde{\succ}_j$ only ranks μ_j as acceptable and suppose that $\tilde{c} = c$. Apparently, we have $\pi(\mu, (g_i, \tilde{g}_{-i})) = \pi(\mu, g)$ and $EDA(\tilde{g}, (\succ_i', \tilde{\succ}_{-i}), \tilde{c}) = \mu$. Thus, $\mu \in \mathcal{M}|_{(\succ_i', c_i)}$.

It remains to be shown that student i regrets $\succ_i'$ through the truth $\succ_i$. Note that since EDA is individually rational, it holds that $EDA_i(\tilde{g}, (\succ_i, \tilde{\succ}_{-i}), \tilde{c}) \succeq_i s_\emptyset$ for any $(\tilde{\succ}_{-i}, \tilde{c}_{-i}, \tilde{g}_{-i}) \in \mathcal{I}(\mu, \succ_i, c_i)$. Since $s_\emptyset \succ_i s$, student i thus regrets $\succ_i'$ through the truth at $(\mu, \pi(\mu, g))$.

Case 2 Let for $\succ_i'$ exist $s \in S$ such that $s_\emptyset \succ_i' s$ and $s \succ_i s_\emptyset$. Suppose i submits $\succ_i'$ and consider $(\mu, \pi(\mu, g))$ such that $\mu_i = s_\emptyset$, $\pi_s(\mu, g) = 0$ and $g_i^{s'} < \pi_{s'}(\mu, g)$ for all $s' \in SU_{s_\emptyset}^{\succ_i'}$. Notably, by doing the same construction $(\tilde{\succ}_{-i}, \tilde{c}_{-i}, \tilde{g}_{-i})$ as in Case 1, we can infer $\mu \in \mathcal{M}|_{(\succ_i', c_i)}$.

It remains to be shown that student i regrets $\succ_i'$ through the truth $\succ_i$. To see this, note that since EDA is non-wasteful, it holds that $EDA_i(\tilde{g}, (\succ_i, \tilde{\succ}_{-i}), \tilde{c}) = s$ for any $(\tilde{\succ}_{-i}, \tilde{c}_{-i}, \tilde{g}_{-i}) \in \mathcal{I}(\mu, \succ_i, c_i)$. Since $s \succ_i s_\emptyset$, student i thus regrets $\succ_i'$ through the truth at $(\mu, \pi(\mu, g))$.

Case 3 In this last case consider $\succ_i'$ which only contains permutations in the acceptable and unacceptable set, i.e., $A_i(\succ_i') = A_i(\succ_i)$ and $U_i(\succ_i') = U_i(\succ_i)$.

The following labeling for any $\succ_i'' \in \mathcal{P}$ in the acceptable set $A_i(\succ_i'')$ ensures that a school's index corresponds to its position in $\succ_i''$. Precisely, we denote s_1'' as the $\succ_i''$-maximal element on $A_{i,1}(\succ_i'') = A_i(\succ_i'')$ and s_2'' as the $\succ_i''$-maximal element on $A_{i,2}(\succ_i'') = A_{i,1}(\succ_i'') \setminus \{s_1''\}$, and so forth.

Now suppose that $|A_i(\succ_i)| = N \in \mathbb{N}$ is the number of acceptable schools under true preferences of student i and consider a permutation $\succ_i'$ as described above. Since $\succ_i'$ is a permutation, there exists $n^* = \arg\min_n \{n \leq N \,|\, s_n' \neq s_n\}$.

Next, let student i observe $(\mu, \pi(\mu, g))$ such that $\mu_i = s_{n^*}'$, $\pi_{s_{n^*}}(\mu, g) = 0$ and $g_i^{s'} < \pi_{s'}(\mu, g)$ for all $s' \in UC_{s_{n^*}'}^{\succ_i'}$. Again, by doing the same construction $(\tilde{\succ}_{-i}, \tilde{c}_{-i}, \tilde{g}_{-i})$ as in Case 1, we can infer $\mu \in \mathcal{M}|_{(\succ_i', c_i)}$.

It remains to be shown that student i regrets $\succ_i'$ through the truth $\succ_i$. Since s_{n^*} has capacity left, this allows us to conclude that if i would have reported $\succ_i$ then, for any $(\tilde{\succ}_{-i}, \tilde{c}_{-i}, \tilde{g}_{-i}) \in \mathcal{I}(\mu, \succ_i, c_i)$, student i would have been matched to s_{n^*}. Since $s_{n^*} \succ_i s_{n^*}'$, we conclude that i regrets $\succ_i'$ through $\succ_i$ at $(\mu, \pi(\mu, g))$. This completes the proof. $\square$

4

Envy and Strategic Choice in Matching Markets*

4.1 Introduction

The reduction of envy among participants in matching markets usually addresses fairness or legal concerns. A goal that is primarily achieved by enforcing normative notions of envy minimization on the

*This chapter is based on Möller (2021b). I am grateful to my advisor Alexander Westkamp for his continuous advice and support. Thanks also to Yiqiu Chen and Marius Gramb for helpful comments. All remaining errors are of course my own.

market outcome.[1] However, as a psychological phenomenon, the occurrence and dislike of envy has a strong idiosyncratic component and also guides human decision making and incentives to envy-avoiding behavior. For instance, it has been shown in different economic settings that agents are willing to incur personal costs to avoid envy.[2] This paper examines the conditions under which matching mechanisms render strategic considerations by market participants aimed at avoiding envy obsolete.

I study a one-to-one priority-based matching model between agents and objects, where agents may experience envy towards each other. To account for the evidence that humans only envy those to which they have a sufficient degree of self-reference (Salovey and Rodin, 1984), the set of agents another agent can potentially envy is expressed as an (arbitrary) subset of other agents in the market, called her *base*. Given any two agents i and j and a matching, agent i has *envy* towards agent j if agent j is in her base and is assigned to an object agent i likes more than the object assigned to herself. In this framework, I develop the notion of *inevitable envy*. Concretely, given a mechanism, agent i's envy towards an agent j that is matched to object x is inevitable, if keeping the reports of agents other than i fixed, agent i has no report where she does not envy j at object x.

I introduce an incentive concept named *envy-proofness*, which accounts for agents' aversion to experiencing envy. Envy-proofness requires that for each agent, any instance of envy she experiences upon reporting her true preference ranking is inevitable. In other words, given the reports of other

[1] In fact, the focus lies on the envy that occurs subject to the violation of priority criteria set by central authorities or by law. The standard fairness-based envy-notion used in the matching context is *justified envy* (Abdulkadiroğlu and Sönmez, 2003). This notion takes an object-specific priority ranking over agents as given, and an agent has justified envy at a matching, only if there exists another agent assigned to an object preferred to her own and that agent is ranked lower at that object than herself. A matching is *fair* if there is no justified envy. For other envy-based fairness notions in matching see, for instance, Morrill (2015) Troyan et al. (2020), Alcalde and Romero-Medina (2015), Ehlers and Morrill (2020) or Nesterov (2017) and Kondratev and Nesterov (2022).

[2] For reference, see Zizzo and Oswald (2001) in the context of money burning, Bolton (1991) in bargaining, or Mui (1995) in retailing. More generally, the influence of envy on human behavior is well documented, both experimentally and empirically, across various domains in economics, sociology and psychology. For reference, see for example, Grund and Sliwka (2005) and Eisenkopf and Teyssier (2013) in tournaments, Kirchsteiger (1994) in ultimatum games or Wenninger et al. (2019) in social networking. For a comprehensive review on the psychological literature on envy, see Smith and Kim (2007) and Crusius et al. (2020) and the references therein.

agents, the instances of envy an agent experiences while being truthful is a weak subset of the instances of envy the agent experiences under any other report she can submit. I show that envy-proofness is stronger than strategy-proofness and thus ensures that agents' incentives with regard to receiving their most preferred objects and those to avoid envy are aligned.[3]

This paper focuses on two prominent strategy-proof matching mechanisms, the Top Trading Cycle (*TTC*) mechanism (Shapley and Scarf, 1974) and the (agent-proposing) Deferred Acceptance (*DA*) mechanism (Gale and Shapley, 1962). DA is the unique strategy-proof mechanism that satisfies the normative notion of justified envy-freeness known also as stability (Alcalde and Barberà, 1994), whereas *TTC* is strategy-proof and efficient (Roth, 1982b) but does not satisfy stability.[4] The main results of this paper are the following. I show that *TTC* is envy-proof, while *DA* is not. In addition, I offer a characterization of envy-proof mechanisms as strategy-proof mechanisms, where given any agent and given the reports of other agents, the agent can only affect the assignments of those agents that match objects that the agent herself can obtain with one of her reports.

In the second part of the paper, I develop an envy-based incentive criterion which is weaker than strategy-proofness. Specifically, a matching mechanism is *weakly envy-invariant* if in any of its Nash-equilibria, each agent's envy is inevitable. Thus, following the spirit of envy-proofness, weak envy-invariance implies that in equilibrium, an agent's instances of envy are a weak subset of instances she could be confronted with when changing her report. The results for this weaker criterion are as follows. I establish that the well-known and non strategy-proof Boston mechanism is weakly envy-invariant, whereas no stable mechanism satisfies this weaker criterion.

The results of this paper suggest a trade-off between achieving certain forms of normative envy-

[3] Strategy-proofness means that it is a weakly dominant strategy for each agent to report her true preferences over individual objects, given the agent cares only about her own assignment.

[4] A matching mechanism is *stable* if it produces matchings which are *fair*, *individually rational* and *non-wasteful*. A matching is non-wasteful if there is no object that is unassigned although there is an agent that prefers it over her assignment. A matching is individually rational if no agent prefers her outside option over her final assignment.

freeness, such as stability, and susceptibility to envy-driven strategic behavior. In fact, under stable mechanisms, avoiding envy and obtaining the best possible individual object may be conflicting strategic goals for participants. However, unstable mechanisms like *TTC* and the Boston mechanism turn out to be strategically robust to envy-avoiding behavior.

Related Literature

To my knowledge, this paper is the first in the matching context to incorporate envy into agents' incentives. Yet, envy has been extensively studied in matching frameworks through the lens of normative envy-freeness and envy-minimization. This includes some recent and related work that has explored in more detail the relationship between envy-minimization and the incentives that a mechanism provides to agents. Specifically, using various comparative methods, Abdulkadiroğlu et al. (2020), Kwon and Shorrer (2020), and Dogan and Ehlers (2021) show that *TTC* minimize justified envy among strategy-proof and Pareto efficient mechanisms.[5] Despite the different focus and the differences in the definition of envy, the present work can be seen as complementing these positive findings on *TTC* in the sense that it additionally ascribes strategic resilience to *TTC* with regard to envy-avoidance.[6]

In addition to the literature dealing with (justified) envy-freeness and envy-minimization in the spirit of Abdulkadiroğlu and Sönmez (2003), there is also a literature examining envy-freeness in random assignment problems following Hylland and Zeckhauser (1979), Zhou (1990), and Bogomolnaia and Moulin (2001). Recent work in the latter tradition by Nesterov (2017) shows that envy is usually a concomitant of achieving efficiency and good incentives. However, as for the case of deterministic matching, the literature that allows for random mechanisms does not consider the influence of envy avoidance on agents' behavior.

Next, since envy-proofness is stronger than strategy-proofness, this paper also contributes to re-

[5] Roth (1982a) has shown that efficiency and no justified envy are incompatible.

[6] Note that the positive findings about the inevitability of envy under *TTC* also apply to any justified envy experienced under *TTC*.

search that has developed incentive notions which are stronger than strategy-proofness. A promiment such property is *group strategy-proofness* which requires that no group of agents can jointly misrepresent their preferences, such that each agent in the group is weakly better off and at least one agent in the group is strictly better off. That is, different from envy-proofness, group strategy-proofness focuses on group deviations rather than unilateral deviations. Pápai (2000) shows that group strategy-proofness is characterized by strategy-proofness and *non-bossiness* defined by Satterthwaite and Sonnenschein (1981). Specifically, non-bossiness means that agents cannot change other agents' assignments by changing their own report, without thereby changing also their own assignment. Envy-proofness, on the other hand, constrains the influence of an agent to assignments of other agents that can be matched with objects that are obtainable by herself, and extends the constraint to situations in which the agent is allowed to change her own assignment. In this sense, envy-proofness and group strategy-proofness impose conceptually similar invariance properties on mechanisms.[7]

To the incentive notions that are stronger than strategy-proofness counts also Li (2017)'s work on *obvious-strategy-proofness (OSP)*. A mechanism is OSP if each agent has an obviously dominant strategy which requires that the worst possible outcome resulting from following the truth is better than the best outcome resulting from any possible untruthful report. Envy-proofness and OSP are not directly related. While the main objective of OSP is to account for cognitive limitations in agents' strategic reasoning, envy-proofness is about addressing agents' concerns about other agents' outcomes. The two approaches also differ in terms of results. Li (2017) finds that *TTC* is not OSP-implementable in some situations and following his work, Troyan (2019) shows that *TTC* is OSP-implementable only if objects' priorities over agents satisfy an acyclicity condition. Ashlagi and Gonczarowski (2018) shows that also DA is not always OSP-implementable, while Thomas (2020) characterizes the priorities under which this is the case.

[7]Note that also the results on envy-proofness and group strategy-proofness are similar: it is well-known that *TTC* is group strategy-proof, and *DA* is not. It is an interesting open question how closely group strategy-proofness and envy-proofness are related.

Finally, this work is also part of a growing literature that integrates nonstandard preferences that account for behavioral biases into the matching context. Fernandez (2020) introduces an incentive concept called regret-free truth-telling that is weaker than strategy-proofness which is based on participants' desire to avoid regret. Meisner and von Wangenheim (2021) and Dreyfuss et al. (2019) examine the behavior of loss averse parents in the context of school choice. Pan (2019) studies the role of agents' overconfidence in a school choice framework. Apart from the differences in the addressed behavioral biases, a main departure is that the agents in the present work are not only concerned about their own assignment, but also about the assignments of other agents.

The rest of this paper is organized as follows. Section 4.2 describes the basic model and Section 4.3 introduces the incentive criterion envy-proofness. Section 4.4 contains the main results. Section 4.5 considers weak envy-invariance and the corresponding results. Section 4.6 concludes.

4.2 Basic Model

4.2.1 Primitives

Let I be a set of agents and X be a set of objects. There is a common outside option $\emptyset$ with infinite capacity. Equip each object $x \in X$ with a strict priority ranking $\rhd_x$ over the set I. A priority structure $\rhd \equiv (\rhd_x)_{x \in X}$ is a profile of priority rankings and the domain of such structures is denoted with $\bar{Pr}$.

For each $i \in I$, let P_i be a strict preference relation over objects $X \cup \{\emptyset\}$. Let R_i be the corresponding weak preference relation of P_i.[8] Let $\mathcal{P}_i$ denote the set of all possible strict preference relations over $X \cup \{\emptyset\}$. For any $P_i \in \mathcal{P}_i$, an object $x \in X$ is *acceptable* to i if $xP_i\emptyset$ and *unacceptable* if it is not acceptable. Refer to the collection $P = (P_i)_{i \in I}$ as a preference profile of agents I and let $\mathcal{P}$ be the full domain of preference profiles. For any $J \subset I$, $P_J = (P_j)_{j \in J}$ is a preference profile for agents J, where $\mathcal{P}_J = \times_{j \in J} \mathcal{P}_j$ is the corresponding domain of such profiles. Denote with $-i$, the set of all agents

[8] That is, for all $x, x' \in X$, xR_ix', if either xP_ix' or $x = x'$.

except i.

A *matching* is a function $\mu : I \to X \cup \{\emptyset\}$ where each object $x \in X$ is assigned to at most one agent and each agent $i \in I$ who does not receive an object $x \in X$ is assigned to the outside option $\emptyset$. Let μ_i denote the object that is assigned to agent $i \in I$ under μ and collect in $\mathcal{M}$ the set of all possible matchings.

In the following, fix some matching $\mu \in \mathcal{M}$ and a preference profile $P \in \mathcal{P}$. Say that μ is *non-wasteful* if there exists no $i \in I$ and no object $x \in X$ such that $x \, P_i \, \mu_i$ and x is unassigned under μ. A matching μ is *individually rational* if, for each $i \in I$, $\mu_i \, R_i \, \emptyset$. Given any $i \in I$, agent i has *justified envy* towards an agent $j \in I$ at $\mu_j \in X$ under μ if $\mu_j \, P_i \, \mu_i$ and $i \rhd_{\mu_j} j$. A matching μ is *fair* if no agent has justified envy. A matching μ is *stable*, if it is individually rational, fair and non-wasteful.

Another matching $\nu \in \mathcal{M}$ *weakly Pareto dominates* matching μ if, for each $i \in I$, $\nu_i \, R_i \, \mu_i$. A matching ν *strictly Pareto dominates* μ, if ν weakly Pareto dominates μ and there exists an agent $j \in I$ with $\nu_j \, P_j \, \mu_j$. A matching μ is *Pareto efficient* if there exists no matching that strictly Pareto dominates it.

A *mechanism* $\psi : \mathcal{P} \to \mathcal{M}$ maps any reported preference profile (hereafter referred to as the report profile) into a matching. Let the result of ψ given a report profile P be denoted with $\psi(P)$, and for each $i \in I$, let $\psi_i(P)$ denote the respective assignment of i under $\psi(P)$. Consider the following desirable properties of mechanisms. First, a mechanism ψ is Pareto efficient if each outcome of the mechanism is Pareto efficient with respect to the report profile. Second, a mechanism is stable if it produces only matchings that are stable with respect to the report profile. Third, a mechanism ψ is non-wasteful (individually rational) if for each reported preference profile, the matching induced is non-wasteful (individually rational). In the remainder of this paper, I restrict attention to mechanisms that are both non-wasteful and individually rational. Denote with $\mathcal{G}$ the full domain of such mechanisms.

Consider the following incentive criterion known as *strategy-proofness*. It requires that an agent interested only in obtaining her best possible match according to her preference ranking should never lie about her preferences. Formally, a mechanism $\psi \in \mathcal{G}$ is strategy-proof if, given any preference profile $P \in \mathcal{P}$, there is no agent $i \in I$ and $\tilde{P}_i \in \mathcal{P}_i$ such that $\psi_i(\tilde{P}_i, P_{-i})\ P_i\ \psi_i(P)$.

4.2.2 Envy

As described in Parrott and Smith (1993) "envy occurs when a person lacks another's superior quality, achievement, or possession and either desires it or wishes that the other lacked it". Moreover, the occurrence and strength of a person's envy towards another typically relies on the degree of self-reference (e.g., a personal relationship, belonging to a reference group, competition) towards the envied person (Salovey and Rodin, 1984). Following these principles, I define the set of other agents an agent can envy in an idiosyncratic manner. Concretely, for any agent $i \in I$, let $\tau_i \subseteq -i$ describe a *base* of agent i and say that i *tracks* agent $j \in I$ if and only if $j \in \tau_i$. Let $\mathcal{T}_i$ be the domain of i's bases. Then, $(P_i, \tau_i) \in \mathcal{P}_i \times \mathcal{T}_i$ is a *type* of agent i. Let $\tau = (\tau_i)_{i \in I}$ be a *base profile* and let $\mathcal{T}$ be the corresponding domain. A *type profile* for agents is a pair $(P, \tau) \in \mathcal{P} \times \mathcal{T}$ that contains a type for each agent.

To define envy, fix a type profile $(P, \tau) \in \mathcal{P} \times \mathcal{T}$ and a matching $\mu \in \mathcal{M}$. An agent $i \in I$ *envies* another agent $j \in I$ at μ_j under μ if i tracks j and $\mu_j\ P_i\ \mu_i$. That is, an agent's envy is always evaluated with respect to the agent's true preferences and an agent envies only agents which are in her base. Refer to a quintuple $(I, X, \rhd, P, \tau)$ as an *(envy) problem*. Fix a triple $(I, X, \rhd)$ and a mechanism $\psi \in \mathcal{G}$ from now on.

4.3 Envy-Proofness

In this section, I introduce the incentive concept *envy-proofness* that promotes truthful disclosure of preferences when agents wish to avoid the experience of envy.

I start with some necessary and helpful terminology. Take any type profile (P, τ) and any agent $i \in I$. Given a report profile $\hat{P} \in \mathcal{P}$, let agent i envy some agent $j \in I$ at $\psi_j(\hat{P})$ under matching $\psi(\hat{P})$.

I say that agent i's envy towards j at $\psi(\hat{P})$ is *inevitable* if for each report $\tilde{P}_i \in \mathcal{P}_i$, i envies j at $\psi_j(\hat{P})$ under $\psi(\tilde{P}_i, \hat{P}_{-i})$. In words, given the reports of other agents, i envies j at the same object for any of her own submitted reports. A report $\hat{P}_i \in \mathcal{P}_i$ is *envy-neutral* for i, if for any $\tilde{P}_{-i} \in \mathcal{P}_{-i}$ and for each agent-object pair $(j, \psi_j(\hat{P}_i, \tilde{P}_{-i}))$ for which i envies j at $\psi_j(\hat{P}_i, \tilde{P}_{-i})$ under $\psi(\hat{P}_i, \tilde{P}_{-i})$, the envy is inevitable for i.

Definition 4.1. A mechanism $\psi \in \mathcal{G}$ is *envy-proof* if, given any type profile $(P, \tau) \in \mathcal{P} \times \mathcal{T}$ and for any agent $i \in I$, being truthful is envy-neutral.

In other words, envy-proofness ensures that agents who lie about their preferences experience every instance of envy that they would experience if they were truthful, and possibly additional instances. Envy-proofness thus implies that agents with the goal to avoid envy, cannot do better than reporting their true preferences. This holds irrespective of whether agents would be willing to sacrifice a better match to avoid envy or not. In fact, the only requirement on agents' attitude towards envy that is needed for the application of envy-proofness is that agents perceive fewer instances of envy as weakly more desirable than more instances of envy in terms of set-inclusion. That is, no additional assumptions are required about how an agent's dislike of envying another agent differs across objects, or how dislike of envy differs across tracked agents.[9]

As shown next, envy-proofness is a stronger incentive requirement than strategy-proofness.

Lemma 4.1. *If $\psi \in \mathcal{G}$ is envy-proof, then it is strategy-proof.*

Proof. Consider mechanism ψ that is not strategy-proof. Thus, there exists a preference profile P and an agent $i \in I$ with a report $\tilde{P}_i \in \mathcal{P}_i$ such that

$$\psi_i(\tilde{P}_i, P_{-i})\; P_i\; \psi_i(P).$$

[9]For instance, a weaker version of inevitability could require that an agent's envy towards another agent persists at *some* weakly less desirable object instead of persisting at the same object. In this case, all negative results in this paper would also extend to the corresponding weaker version of envy-proofness.

Since ψ is non-wasteful, there must exist some agent $j \in I$ with $\psi_j(P) = \psi_i(\tilde{P}_i, P_{-i})$ and $\psi_j(P) \neq \emptyset$. Take the type profile $(P, \tilde{\tau})$ with $j \in \tilde{\tau}_i$ and note that i being truthful implies that i envies j at $\psi_j(P)$ under $\psi(P)$. Now, if i reports $\tilde{P}_i$, then $\psi_i(\tilde{P}_i, P_{-i}) = \psi_j(P)$ and thus $\psi_j(\tilde{P}_i, P_{-i}) \neq \psi_j(P)$. Hence i's envy towards j at $\psi_j(P)$ is not inevitable. We conclude that ψ is not envy-proof. $\square$

The essence of mechanisms being envy-proof, thus, is that truthfulness is the weakly dominant action for agents who want their best possible match while avoiding envy. So there is no strategic trade-off for agents between envy-avoidance and getting the best possible match.

In the remainder of this section, I provide a straightforward characterization of envy-proof mechanisms. To do so, some additional terminology is needed. Take any agent $i \in I$ and a preference profile for other agents $P_{-i} \in \mathcal{P}_{-i}$. Let

$$\bar{X}_i(\psi, P_{-i}) \equiv \{x \in X \,|\, \nexists \tilde{P}_i \in \mathcal{P}_i \text{ with } \psi_i(\tilde{P}_i, P_{-i}) = x\}$$

be the set of i's *unattainable* objects. Moreover, let

$$\bar{I}_i(\psi, P_{-i}) \equiv \{j \in I \,|\, \exists \tilde{P}_i \in \mathcal{P}_i \text{ and } x \in \bar{X}_i(\psi, P_{-i}) \text{ with } \psi_j(\tilde{P}_i, P_{-i}) = x\}$$

be the set of agents (other than i) that can be matched to an object unattainable for agent i.

Definition 4.2. A mechanism $\psi \in \mathcal{G}$ is *fenced*, if for each $i \in I$, each preference profile $P \in \mathcal{P}$, $\tilde{P}_i \in \mathcal{P}_i$ and for each $j \in \bar{I}_i(\psi, P_{-i})$, we obtain $\psi_j(P) = \psi_j(\tilde{P}_i, P_{-i})$.

In other words, under a fenced mechanism, an agent cannot influence the assignments of agents that can be assigned to an object that, given the reports of others, is unattainable for herself. We reach the following characterization of envy-proof mechanisms.

Proposition 4.1. *A mechanism $\psi \in \mathcal{G}$ is envy-proof if and only if it is strategy-proof and fenced.*

Proof. See Appendix 4.B. □

4.4 Main Results

In this section, I present the main results of this paper. I show that the *Top Trading Cycles (TTC)* mechanism (Shapley and Scarf, 1974) is envy-proof and that the *Deferred Acceptance (DA) Mechanism* (Gale and Shapley, 1962) is not.

4.4.1 TTC

Denote the *TTC* mechanism that is induced with the *TTC* algorithm operating on $\rhd \in \bar{Pr}$ with $TTC^{\rhd}$. The description of the algorithm to induce $TTC^{\rhd}$ can be found in Appendix 2.D of Chapter 2.

Before I present the result, consider the following helpful properties of $TTC^{\rhd}$. First, it is well known that $TTC^{\rhd}$ is strategy-proof and Pareto efficient (Roth, 1982b). Furthermore, $TTC^{\rhd}$ satisfies *non-bossiness* (Satterthwaite and Sonnenschein, 1981) which requires that there is no agent who can change other agents' assignments by misreporting her preferences, without thereby also changing her own assignment (Pycia and Ünver, 2017). Formally, a mechanism $\psi \in \mathcal{G}$ is non-bossy, if for all $P \in \mathcal{P}$, there is no $i \in I$, and $\tilde{P}_i \in \mathcal{P}_i$, such that $\psi_i(P) = \psi_i(\tilde{P}_i, P_{-i})$, but $\psi(P) \neq \psi(\tilde{P}_i, P_{-i})$. A mechanism $\psi \in \mathcal{G}$ is *bossy* if it is not non-bossy. For each report profile $P \in \mathcal{P}$, let $TTCA^{\rhd}(P)$ denote the corresponding process under *TTC* algorithm. The *TTC* algorithm has the following helpful property.

Lemma 4.2 (Roth (1982b))**.** *Take an agent $i \in I$ and a preference profile $P \in \mathcal{P}$. If, given any $\tilde{P}_i \in \mathcal{P}_i$, agent i is unmatched at the beginning of step t of $TTCA^{\rhd}(P)$ and $TTCA^{\rhd}(\tilde{P}_i, P_{-i})$, then the same cycles are solved in each step $t' < t$ in $TTCA^{\rhd}(P)$ and $TTCA^{\rhd}(\tilde{P}_i, P_{-i})$.*

In words, under the TTC algorithm there is no agent who can affect cycles that have been solved before that agent is involved in a cycle herself.

We are ready for the main result of this paper.

Theorem 4.1. *$TTC^{\triangleright}$ is envy-proof.*

As an intuition for the proof, first note that non-bossiness of $TTC^{\triangleright}$ implies that an agent who would be envious under truth and wishes to avoid it by influencing others assignment, must choose a non-truthful report that grants her an object, that is worse than the one she receives under truth. Lemma 4.2 then implies that if under the non-truthful report the agent is involved in a cycle, given all the objects for which she had envy have already been assigned, then those objects must be matched in the same cycles as if the agent had been truthful. Now consider worse objects with which the agent can achieve to be in a cycle earlier than the step in which she is involved in a cycle under truth. In this case, strategy-proofness of *TTC* ensures that an agent can delay to be assigned such a worse object, at least until the step where all objects she could be envious at have already been assigned.[10] This works, for instance, with a report where relative to the true ranking the worse object in consideration exchanges positions with the object the agent is assigned to under truth. By Lemma 4.2, under the delayed scenario all envied agents are involved in the same cycles as under truth. Finally, using the non-bossiness of $TTC^{\triangleright}$ whether or not to delay the match to the worse object does not affect other agents' assignments. Thus, if the agent matches a worse object, then all instances of envy the agent experiences under truth are still present.

4.4.2 DA

Denote the *DA* mechanism that is induced with the *DA* algorithm operating on $\triangleright \in \bar{Pr}$ with $DA^{\triangleright}$. The algorithm can be found in Appendix 2.D of Chapter 2. Note that $DA^{\triangleright}$ is the unique stable and

[10] A feature of the TTC algorithm is that if the pointing of an object to an agent began, the pointing continues until the agent leaves with her match.

strategy-proof mechanism (Alcalde and Barberà, 1994) and that by stability, agents cannot experience justified envy under $DA^{\triangleright}$. However, envy as introduced in subsection 4.2.2 is also present under $DA^{\triangleright}$. As will be demonstrated below, this envy is not always inevitable if agents are truthful.

Theorem 4.2. *If $|X| \geq 3$ and $|I| \geq 3$, then there exists $\triangleright$ such that $DA^{\triangleright}$ is not envy-proof.*

Proof. Let $|X| \geq 3$ and $|I| \geq 3$. We first specify a type profile (P, τ) and describe the priorities $\triangleright$. To start, for agents $i, j, k \in I$ and objects $x, y, z \in X$ let the relevant preferences and priorities be depicted in the following tables:

P_i	$\tilde{P}_i$	P_j	P_k
y	z	y	x
z	y	x	y
x	x	z	z
$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
$\vdots$	$\vdots$	$\vdots$	$\vdots$

$\triangleright_x$	$\triangleright_y$	$\triangleright_z$
j	k	i
k	i	j
i	j	k
$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$

Let $\tau_i = \{k\}$ and for all other agents $i' \in I \setminus \{i\}$ let $\tau_{i'}$ be arbitrary. Let the priorities $\triangleright_{x'}$ for each remaining object $x' \in X \setminus \{x, y, z\}$ be arbitrary and assume that for each $l \in I \setminus \{i, j, k\}$ that P_l ranks the outside option $\emptyset$ first and the ranking below $\emptyset$ is arbitrary. Taking as an input the true preferences of each agent, one reaches $DA_i^{\triangleright}(P) = z$, $DA_j^{\triangleright}(P) = x$, $DA_k^{\triangleright}(P) = y$ and for all $l \in I \setminus \{i, j, k\}$, $DA_l^{\triangleright}(P) = \emptyset$. Since $\tau_i = \{k\}$, agent i envies agent k at y under $DA^{\triangleright}(P)$. To see that i's envy towards k at y is not inevitable, let agent i instead report $\tilde{P}_i$ as displayed in the table above. In this case, one reaches $DA_i^{\triangleright}(\tilde{P}_i, P_{-i}) = z$, $DA_j^{\triangleright}(\tilde{P}_i, P_{-i}) = y$, $DA_k^{\triangleright}(\tilde{P}_i, P_{-i}) = x$ and for all $l \in I \setminus \{i, j, k\}$, $DA_l^{\triangleright}(\tilde{P}_i, P_{-i}) = \emptyset$. Thus, i does not envy k since k is now assigned to x, whereas i's assignment remains unchanged. That is, being truthful is not envy-neutral for agent i and hence $DA^{\triangleright}$ is not envy-proof. This completes the proof.

$\square$

In the example above, note that through i's choice to avoid envy actually a Pareto improvement for the other agents is generated and that i's assignment is not affected by her choice to be untruthful. This is not generally true. In fact, one can construct similar examples where the welfare effect on other agents goes in the opposite direction, the effect on other agents is ambiguous or where the envying agent has to change her assignment to avoid envy. Such an example is presented in the next section, in which I introduce an envy-based incentive concept that is also applicable to non-strategy mechanisms.

4.5 Weak Envy-Invariance

In this section, I extend the analysis to non strategy-proof mechanisms. The key difference compared to the analysis of strategy-proof mechanisms is the following. Under non strategy-proof candidates there may be agents who can avoid instances of envy they experience while being truthful by improving their own match through lying about their preferences. A straightforward way to accommodate for these differences, is to focus on instances of envy that remain subject to an equilibrium condition.

Following these ideas, I now delineate an incentive criterion called *weak envy-invariance*, which will be weaker than strategy-proofness and takes care of instances of envy present in the Nash-equilibria of a mechanism. Take any type profile (P, τ). For an agent $i \in I$, the report $\hat{P}_i \in \mathcal{P}_i$ is a *best response* to other agents' reports $\hat{P}_{-i} \in \mathcal{P}_{-i}$ if $\psi_i(\hat{P})\ R_i\ \psi_i(\tilde{P}_i, \hat{P}_{-i})$ for all $\tilde{P}_i \in \mathcal{P}_i$. A report profile $\hat{P} \in \mathcal{P}$ constitutes a *Nash-equilibrium* under ψ, if for each $i \in I$, $\hat{P}_i$ is a best response to other agents' reports $\hat{P}_{-i}$. Now consider the following incentive criterion. A mechanism $\psi \in \mathcal{G}$ is weakly envy-invariant, if for each type profile (P, τ), for each $i \in I$, under any associated Nash-equilibrium $\hat{P} \in \mathcal{P}$ and for each agent-object pair $(j, \psi_j(\hat{P}))$ for which i envies j at $\psi_j(\hat{P})$ under $\psi(\hat{P})$, the envy is inevitable for agent i.

4.5.1 Stable Mechanisms

Since DA is not envy-proof (Theorem 4.2), one may ask whether there exist stable mechanisms that are weakly envy-invariant irrespective of what the underlying priorities are. As shown below, the answer is no.

Proposition 4.2. *If* $|X| \geq 3$ *and* $|I| \geq 4$*, then there exists* $\rhd$ *such that there is no stable and weakly envy-invariant mechanism.*

Proof. Let $|X| \geq 3$ and $|I| \geq 4$. We first specify a type profile (P, τ) and priorities $\rhd$. To start, for agents $i, j, k, l \in I$ and objects $x, y, z \in X$ let the relevant preferences and priorities be depicted in the following tables:

P_i	$\tilde{P}_i$	P_j	P_k	P_l
y	z	y	z	x
x	y	x	y	$\emptyset$
z	$\emptyset$	$\emptyset$	$\emptyset$	y
$\emptyset$	x	z	x	z
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

$\rhd_x$	$\rhd_y$	$\rhd_z$
i	k	i
l	j	k
j	i	j
k	l	l
$\vdots$	$\vdots$	$\vdots$

Suppose that i only tracks j, i.e. $\tau_i = \{j\}$ and for all other agents $i' \in I \setminus \{i\}$ let $\tau_{i'}$ be arbitrary. Assume that for each $m \in I \setminus \{i, j, k, l\}$ that P_m ranks the outside option $\emptyset$ first and the ranking below $\emptyset$ is arbitrary. Let the priorities $\rhd_{x'}$ for each remaining object $x' \in X \setminus \{x, y, z\}$ be arbitrary. I now claim that given input P any stable mechanism ψ must produce $\psi_i(P) = x$, $\psi_j(P) = y$, $\psi_k(P) = z$, $\psi_l(P) = \emptyset$ and for all $m \in I \setminus \{i, j, k, l\}$, $\psi_m(P) = \emptyset$.

To start, note that for all $m \in I \setminus \{i, j, k, l\}$, $\psi_m(P) = \emptyset$, by individual rationality of ψ. Next, observe that i cannot be assigned to her top choice y in any stable matching since j would have justified envy towards i at y. This implies that i must get her second choice x where she has highest priority. Once her top choice is gone with agent i, by individual rationality of stable matchings, l must receive the outside option which is her second choice. Moreover, once i is assigned to x, k is assured to get z, since no agent but i has higher priority than her at z and z is k's top choice. Hence j is left with y, which is also her top choice. Note that it is clear then, that the report profile P constitutes a Nash-equilibrium under ψ. In fact, one can see that agents who are not matched to their top choices and unilaterally deviate cannot be made better off, without at the same time inducing justified envy at the resulting candidate matchings. However, this would contradict stability of the mechanism.

As a next step, note that i envies j at y under $\psi(P)$ since $y \; P_i \; x$ and $j \in \tau_i$. We now establish that the envy towards j at y is not inevitable for i. Concretely, suppose that i reports preferences $\tilde{P}_i$ as displayed in the table above. Then, given report profile $(\tilde{P}_i, P_{-i})$ any stable mechanism ψ must produce $\psi_i(\tilde{P}_i, P_{-i}) = z$, $\psi_j(\tilde{P}_i, P_{-i}) = \emptyset$, $\psi_k(\tilde{P}_i, P_{-i}) = y$, $\psi_l(\tilde{P}_i, P_{-i}) = x$ and for all $m \in I \setminus \{i, j, k, l\}$, $\psi_m(\tilde{P}_i, P_{-i}) = \emptyset$. This can be seen using the following arguments. First, by individual rationality of ψ for all $m \in I \setminus \{i, j, k, l\}$, $\psi_m(\tilde{P}_i, P_{-i}) = \emptyset$. Next, under any stable matching with respect to $(\tilde{P}_i, P_{-i})$, i must be assigned to z since it is her top choice and she is the highest priority agent at z. However, then k must be assigned to y, for which she has highest priority priority and which is her best choice remaining once i left with z. In this case, any stable matching must assign l to x since j has lower priority at x than l. Hence j must remain with the outside option.

Finally, note that i has no envy under $\psi(\tilde{P}_i, P_{-i})$ and thus the envy towards j at y is not inevitable. Thus, ψ is not weakly envy-invariant. This completes the proof. □

4.5.2 Immediate Acceptance

Another interesting and well-known candidate mechanism in practice is the *Immediate Acceptance (IA) mechanism*, also known as the Boston mechanism. Denote the *IA* mechanism that operates on $\triangleright$ with $IA^{\triangleright}$. The algorithm to calculate the matchings under *IA* can be found in Appendix 4.A. The efficient and non-bossy *IA* mechanism is not strategy-proof and also not stable. In fact, the *IA* mechanism was criticized as being vulnerable to manipulation of preferences and was therefore replaced by the Boston School Choice Committee in 2005 with the strategy-proof and stable *DA* (Pathak and Sönmez, 2008). However, as will be shown below, unlike stable mechanisms, the *IA* mechanism satisfies the weak envy-invariance criterion.

Proposition 4.3. *$IA^{\triangleright}$ is weakly envy-invariant.*

Proof. Fix an agent $i \in I$ and a type profile (P, τ) with some agent $j \in \tau_i$. Let report profile $\hat{P} \in \mathcal{P}$ constitute a Nash-equilibrium under $IA^{\triangleright}$ and let i envy agent $j \in I$ at $IA_j^{\triangleright}(\hat{P})$ under $IA^{\triangleright}(\hat{P})$. We show that i's envy towards j at $IA_j^{\triangleright}(\hat{P})$ is inevitable for i.

To start, since $\hat{P}_i$ is a best response to $\hat{P}_{-i}$ and $IA^{\triangleright}$ is non-bossy[11], for any other best response $P_i^* \in \mathcal{P}_i$ of i to $\hat{P}_{-i}$, we reach $IA^{\triangleright}(P_i^*, \hat{P}_{-i}) = IA^{\triangleright}(\hat{P})$. Thus, for all other best responses, i still envies j at $IA_j^{\triangleright}(\hat{P})$.

Next, let $\tilde{P}_i \in \mathcal{P}$ be an arbitrary report of i which is not a best response to $\hat{P}_{-i}$. Thus, we have $IA_i^{\triangleright}(\tilde{P}_i, \hat{P}_{-i}) \neq IA_i^{\triangleright}(\hat{P})$. To show that j's match will not be affected, first note that j must rank $IA_j^{\triangleright}(\hat{P})$ as her top choice in $\hat{P}_j$. Otherwise, $\hat{P}_i$ cannot be a best response to $\hat{P}_{-i}$, since i could submit a report where $IA_j^{\triangleright}(\hat{P})$ is ranked first, which would imply that i must receive $IA_j^{\triangleright}(\hat{P})$. However, note that by reporting $\tilde{P}_i$, agent i can not lead a new agent $k \in I \setminus \{i, j\}$ to apply to $IA_j^{\triangleright}(\hat{P})$ at the first step of the *IA* algorithm with input $(\tilde{P}_i, \hat{P}_{-i})$. This implies that j must receive $IA_j^{\triangleright}(\hat{P})$ and as such i's envy towards j at $IA_j^{\triangleright}(\hat{P})$ is inevitable.

[11] The proof for non-bossiness of $IA^{\triangleright}$ can be found in Harless (2014) (Proposition 4).

Since all choices made were arbitrary, we conclude that any instance of envy under a best response in an associated Nash equilibrium is inevitable. Hence $IA^{\triangleright}$ is weakly envy-invariant. This completes the proof.

□

4.6 Conclusion

In this paper, I introduced a new incentive concept called envy-proofness that takes into account agents' aversion to envy and which is stronger than strategy-proofness. The candidates an agent possibly envies are specified idiosyncratically and are incorporated into the agent's type. Envy-proofness requires that agents cannot avoid instances of envy they are exposed to while being truthful about their preferences. I applied the concept to *TTC* and *DA* and established that the former is envy-proof, while the latter is not.

It is an open question how envy-proofness relates to group strategy-proofness. Related to this, future research could explore more general frameworks for finding envy-proof mechanisms, such as when group strategy-proofness is difficult to achieve or infeasible. For example, group strategy-proofness is incompatible with Pareto efficiency on the preference domain with indifferences (Ehlers, 2002) and it would be interesting to see whether the positive findings of this paper also apply to strategy-proof versions of *TTC* in these more general settings (Jaramillo and Manjunath, 2012). Another example is the class of *Component-Wise Individually Rational Mechanisms* proposed by Manjunath and Westkamp (2021) suitable in particular for the context of shift-exchange. These mechanisms are strategy-proof, individually rational and Pareto efficient, but it is unclear whether the stronger incentive requirement of envy-proofness can be satisfied for some members of this class.[12]

[12] In Manjunath and Westkamp (2021) agents preferences are *trichotomous*. Specifically, agents with *trichotomous* preferences divide the set of available objects into those they find desirable, those they are endowed with but are undesirable and those they find undesirable but they are not endowed with.

4.A The Immediate Acceptance Algorithm

The outcome of $IA^{\triangleright}$ given any preference profile $P \in \mathcal{P}$ is calculated as follows:

Step 1 Each agent $i \in I$ proposes to her most preferred object in $X \cup \{\emptyset\}$. Each object $x \in X$ considers all the proposals and accepts the candidate who applies to x and has the highest ranking on $\triangleright_x$. The remaining proposals are rejected. Moreover, all agents that propose to the outside option $\emptyset$ are accepted.

Step k, $k \geq 2$ Each agent who was rejected at step $k-1$ applies to her most preferred object not yet applied to. Each object $x \in X$ considers all the new applicants and accepts the highest ranked applicant according to $\triangleright_x$ in case no agent has applied to the object in some previous round. Otherwise, all proposals are rejected. All agents that propose to the outside option $\emptyset$ are accepted.

The algorithm terminates with the assignments of the first step in which no agent is rejected.

4.B Proofs

4.B.1 Proof of Proposition 4.1

($\Rightarrow$) By Lemma 4.1, strategy-proofness is necessary for envy-proofness. Hence it remains to show that each envy-proof mechanism is fenced. Take any $\psi \in \mathcal{G}$ that is strategy-proof and not fenced. Thus, there exists report profile $P \in \mathcal{P}$, $i \in I$ and $\tilde{P}_i \in \mathcal{P}_i$ such that there exists $j \in \bar{I}_i(\psi, P_{-i})$ with $\psi_j(P) \in \bar{X}_i(\psi, P_{-i})$ and $\psi_j(P) \neq \psi_j(\tilde{P}_i, P_{-i})$. I show that there exists a type profile, for which being truthful is not envy-neutral for agent i.

Consider type profiles (P, τ) and $((\tilde{P}_i, P_{-i}), \tau)$, where $\tau_i = I \backslash \{i\}$. We distinguish two basic cases.

Case 1: $\psi_j(P)\; P_i\; \psi_i(P)$ or $\psi_j(\tilde{P}_i, P_{-i})\; \tilde{P}_i\; \psi_i(\tilde{P}_i, P_{-i})$. If $\psi_j(P)\; P_i\; \psi_i(P)$, then i envies j at $\psi_j(P)$ under $\psi(P)$ for type profile (P, τ). Now, since $\psi_j(P) \neq \psi_j(\tilde{P}_i, P_{-i})$, i's envy towards j at $\psi_j(P)$ is not inevitable and thus ψ is not envy-proof. If $\psi_j(\tilde{P}_i, P_{-i})\; \tilde{P}_i\; \psi_i(\tilde{P}_i, P_{-i})$, then $\psi_j(\tilde{P}_i, P_{-i}) \neq \emptyset$ by non-wastefulness of ψ. Then, a symmetric argument works with i envying j at $\psi_j(\tilde{P}_i, P_{-i})$ under $\psi(\tilde{P}_i, P_{-i})$ for type profile $((\tilde{P}_i, P_{-i}), \tau)$.

Case 2: $\psi_i(P)\; P_i\; \psi_j(P)$ and $\psi_i(\tilde{P}_i, P_{-i})\; \tilde{P}_i\; \psi_j(\tilde{P}_i, P_{-i})$. Consider a type profile $((P^*_i, P_{-i}), \tau)$ with P^*_i such that $\psi_j(P)\; P^*_i\; \psi_i(P)$ and $\psi_i(P)\; R^*_i\; x'$ for all $x' \in X \cup \{\emptyset\} \setminus \{\psi_j(P)\}$.

By strategy-proofness of ψ and $\psi_j(P) \in \bar{X}_i(\psi, P_{-i})$, we have $\psi_i(P^*_i, P_{-i}) = \psi_i(P)$. Moreover, by non-wastefulness of ψ, there exists $k \in I$ such that $\psi_k(P^*_i, P_{-i}) = \psi_j(P)$. Hence by definition of $\bar{X}_i(\psi, P_{-i})$, $k \in \bar{I}_i(\psi, P_{-i})$ and since $k \in \tau_i$, i envies k at $\psi_k(P^*_i, P_{-i})$ under $\psi(P^*_i, P_{-i})$. If $k \neq j$, then i's envy towards k at $\psi_k(P^*_i, P_{-i})$ is not inevitable since $\psi_k(P^*_i, P_{-i}) \neq \psi_k(P)$. If $k = j$, then i's envy towards j is not inevitable, since $\psi_j(P^*_i, P_{-i}) \neq \psi_j(\tilde{P}_i, P_{-i})$. Thus, ψ is not envy-proof. This completes the first part of the proof.

($\Leftarrow$) Suppose that $\psi \in \mathcal{G}$ is strategy-proof and fenced. Consider an arbitrary type profile (P, τ), where there exists an agent $i \in I$ who envies another agent $j \in \tau_i$ at $\psi_j(P)$ under $\psi(P)$. Since ψ is strategy-proof, it holds that, for each $\tilde{P}_i \in \mathcal{P}_i$, $\psi_i(P)\; R_i\; \psi_i(\tilde{P}_i, P_{-i})$. Hence $\psi_j(P) \in \bar{X}_i(\psi, P_{-i})$. Now, since ψ is fenced, for all $\tilde{P}_i \in \mathcal{P}_i$, it holds that $\psi_j(P) = \psi_j(\tilde{P}_i, P_{-i})$. However, then i's envy towards j at $\psi_j(P)$ is inevitable. Since the instance of envy was selected arbitrarily, we conclude that reporting the truth P_i is an envy-neutral report for i. Next, note that also the type profile and agent i were selected arbitrarily and thus being truthful is envy-neutral for each agent and each type profile under ψ. We conclude that ψ is envy-proof. This completes the proof.

4.B.2 Proof of Theorem 4.1

Fix an agent $i \in I$ and a type profile (P, τ). Let agents report their true preferences P to $TTC^{\triangleright}$ and suppose that i envies an agent $j \in \tau_i$ at $TTC^{\triangleright}_j(P)$ under $TTC^{\triangleright}(P)$, where we denote $TTC^{\triangleright}_j(P) = x$.

Next, select an arbitrary $\tilde{P}_i \in \mathcal{P}$. We have to show that i's envy towards j at x under $TTC^{\triangleright}(\tilde{P}_i, P_{-i})$ is inevitable. Since $TTC^{\triangleright}$ is strategy-proof, we have $TTC^{\triangleright}_i(P)\ R_i\ TTC^{\triangleright}_i(\tilde{P}_i, P_{-i})$. We thus can distinguish the following cases:

Case 1: $TTC^{\triangleright}_i(P) = TTC^{\triangleright}_i(\tilde{P}_i, P_{-i})$. Non-bossiness of $TTC^{\triangleright}$ implies $TTC^{\triangleright}(P) = TTC^{\triangleright}(\tilde{P}_i, P_{-i})$ and hence $TTC^{\triangleright}_j(P) = TTC^{\triangleright}_j(\tilde{P}_i, P_{-i})$. Thus, i still envies j at x under $TTC^{\triangleright}(\tilde{P}_i, P_{-i})$.

Case 2: $TTC^{\triangleright}_i(P)\ P_i\ TTC^{\triangleright}_i(\tilde{P}_i, P_{-i})$. Under process $TTCA^{\triangleright}(P)$, let i be removed at step t_i and let j be removed with x at step t_x. Since $x\ P_i\ TTC^{\triangleright}_i(P)$, i does not point to $TTC^{\triangleright}_i(P)$, before x is removed. Thus, $t_x < t_i$.

Next, under $TTCA^{\triangleright}(\tilde{P}_i, P_{-i})$ let i be removed at step $\tilde{t}_i$. Let $t_i^{min} = \min\{t_i, \tilde{t}_i\}$. We have two more subcases.

Case 2.1. $\mathbf{\tilde{t}_i > t_x}$. Since $t_i > t_x$ and $\tilde{t}_i > t_x$, we have $t_i^{min} > t_x$. By Lemma 4.2, for each $t < t_i^{min}$ the cycles removed are identical in $TTCA^{\triangleright}(P)$ and $TTCA^{\triangleright}(\tilde{P}_i, P_{-i})$. Hence $t_i^{min} > t_x$ implies that $TTC^{\triangleright}_j(\tilde{P}_i, P_{-i}) = x$. Thus, since $TTC^{\triangleright}_i(P)\ P_i\ TTC^{\triangleright}_i(\tilde{P}_i, P_{-i})$, i still envies j at x under $TTC^{\triangleright}(\tilde{P}_i, P_{-i})$.

Case 2.2. $\mathbf{\tilde{t}_i \leq t_x}$. In this case, we first construct an auxiliary process, where i receives assignment $TTC^{\triangleright}_i(\tilde{P}_i, P_{-i})$ and envies j at x and then use the non-bossiness of $TTC^{\triangleright}$ to conclude that i envies j at x also under $TTC^{\triangleright}(\tilde{P}_i, P_{-i})$. Concretely, let i report P_i^* such that $x\ P_i^*\ TTC^{\triangleright}_i(\tilde{P}_i, P_{-i})$ and $TTC^{\triangleright}_i(\tilde{P}_i, P_{-i})\ R_i^*\ x'$ for all $x' \in X \cup \{\emptyset\} \setminus \{x\}$. Now note that under $TTCA^{\triangleright}(P_i^*, P_{-i})$, agent i

again does not point to $TTC_i^{\triangleright}(\tilde{P}_i, P_{-i})$ before x is removed. Thus, using Lemma 4.2, for each step $t \leq t_x$, the cycles removed are identical in $TTCA^{\triangleright}(P)$ and $TTCA^{\triangleright}(P_i^*, P_{-i})$. Hence j receives x under $TTC^{\triangleright}(P_i^*, P_{-i})$. Moreover, by strategy-proofness of $TTC^{\triangleright}$, $TTC_i^{\triangleright}(P_i^*, P_{-i}) = TTC_i^{\triangleright}(\tilde{P}_i, P_{-i})$. Since $TTC_i^{\triangleright}(P)\, P_i\, TTC_i^{\triangleright}(\tilde{P}_i, P_{-i})$, we obtain that i envies j at x under $TTC^{\triangleright}(P_i^*, P_{-i})$. Finally, given $TTC_i^{\triangleright}(P_i^*, P_{-i}) = TTC_i^{\triangleright}(\tilde{P}_i, P_{-i})$, we have $TTC^{\triangleright}(\tilde{P}_i, P_{-i}) = TTC^{\triangleright}(P_i^*, P_{-i})$ by non-bossiness of $TTC^{\triangleright}$ and therefore, i must also envy j at x under $TTC^{\triangleright}(\tilde{P}_i, P_{-i})$.

Note that all choices made were arbitrary. That is, being truthful is envy-neutral for each agent under each type profile. Hence $TTC^{\triangleright}$ is envy-proof. This completes the proof.

Bibliography

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.

Atila Abdulkadiroğlu, Parag A. Pathak, and Alvin E. Roth. Strategy-proofness versus efficiency in matching with indifferences: Redesigning the nyc high school match. *American Economic Review*, 99(5):1954–78, 2009.

Atila Abdulkadiroğlu, Yeon-Koo Che, Parag A. Pathak, Alvin E. Roth, and Olivier Tercieux. Efficiency, justified envy, and incentives in priority-based matching. *American Economic Review: Insights*, 2(4):425–42, December 2020.

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

José Alcalde and Salvador Barberà. Top dominance and the possibility of strategy-proof stable solutions to matching problems. *Economic theory*, 4(3):417–435, 1994.

José Alcalde and Antonio Romero-Medina. Strategy-proof fair school placement. *Available at SSRN 1743082*, 2015.

Samson Alva and Vikram Manjunath. Stable-dominating rules. Working paper, University of Ottawa, 2019.

Itai Ashlagi and Yannai A Gonczarowski. Stable matching mechanisms are not obviously strategy-proof. *Journal of Economic Theory*, 177:405–425, 2018.

Eduardo M Azevedo and Jacob D Leshno. A supply and demand framework for two-sided matching markets. *Journal of Political Economy*, 124(5):1235–1268, 2016.

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.

Keisuke Bando. On the existence of a strictly strong nash equilibrium under the student-optimal deferred acceptance algorithm. *Games and Economic Behavior*, 87:269 – 287, 2014.

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.

Anna Bogomolnaia and Hervé Moulin. A new solution to the random assignment problem. *Journal of Economic theory*, 100(2):295–328, 2001.

Gary E Bolton. A comparative model of bargaining: Theory and evidence. *The American Economic Review*, pages 1096–1136, 1991.

Peter Chen, Michael Egesdal, Marek Pycia, and M Bumin Yenmez. Quantile stable mechanisms. *Available at SSRN 2526505*, 2015.

Yiqiu Chen and Markus Möller. Regret-free truth-telling in school choice with consent. Working paper, University of Cologne, 2021.

Jan Crusius, Manuel F Gonzalez, Jens Lange, and Yochi Cohen-Charash. Envy: An adversarial review and comparison of two competing views. *Emotion Review*, 12(1):3–21, 2020.

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

Battal Dogan and Lars Ehlers. Robust minimal instability of the top trading cycles mechanism. *American Economic Journal: Microeconomics*, January 2021.

Bnaya Dreyfuss, Ori Heffetz, and Matthew Rabin. Expectations-based loss aversion may help explain seemingly dominated choices in strategy-proof mechanisms. Technical report, National Bureau of Economic Research, 2019.

Lester E Dubins and David A Freedman. Machiavelli and the gale-shapley algorithm. *The American Mathematical Monthly*, 88(7):485–494, 1981.

Umut Dur, A Arda Gitmez, and Özgür Yılmaz. School choice under partial fairness. *Theoretical Economics*, 14(4):1309–1346, 2019.

Lars Ehlers. Coalitional strategy-proof house allocation. *Journal of Economic Theory*, 105(2):298–317, 2002.

Lars Ehlers. Truncation strategies in matching markets. *Mathematics of Operations Research*, 33(2): 327–335, 2008.

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 and Thayer Morrill. (il) legal assignments in school choice. *The Review of Economic Studies*, 87(4):1837–1875, 2020.

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.

Gerald Eisenkopf and Sabrina Teyssier. Envy and loss aversion in tournaments. *Journal of Economic Psychology*, 34:240–255, 2013.

Richard Engelbrecht-Wiggans. The effect of regret on optimal bidding in auctions. *Management Science*, 35(6):685–692, 1989.

Aytek Erdil and Haluk Ergin. What's the matter with tie-breaking? improving efficiency in school choice. *American Economic Review*, 98(3):669–89, 2008.

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

MA Fernandez. Deferred acceptance and regret-free truthtelling. Working paper, John Hopkins University, 2020.

Emel Filiz-Ozbay and Erkut Y Ozbay. Auctions with anticipated regret: Theory and experiment. *American Economic Review*, 97(4):1407–1418, 2007.

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

Thomas Gilovich and Victoria Husted Medvec. The experience of regret: what, when, and why. *Psychological review*, 102(2):379, 1995.

Christian Grund and Dirk Sliwka. Envy and compassion in tournaments. *Journal of Economics & Management Strategy*, 14(1):187–207, 2005.

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. Transparency in centralised allocation: Theory and experiment. *Available at SSRN 3572020*, 2020.

Patrick Harless. A School Choice Compromise: Between Immediate and Deferred Acceptance. MPRA Paper 61417, University Library of Munich, Germany, June 2014.

Aanund Hylland and Richard Zeckhauser. The efficient allocation of individuals to positions. *Journal of Political economy*, 87(2):293–314, 1979.

Paula Jaramillo and Vikram Manjunath. The difference indifference makes in strategy-proof allocation of objects. *Journal of Economic Theory*, 147(5):1913–1946, 2012.

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

Onur Kesten. School choice with consent. *The Quarterly Journal of Economics*, 125(3):1297–1348, 2010.

Georg Kirchsteiger. The role of envy in ultimatum games. *Journal of economic behavior & organization*, 25(3):373–389, 1994.

Bettina Klaus and Flip Klijn. Median stable matching for college admissions. *International Journal of Game Theory*, 34(1):1, 2006.

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

Fuhito Kojima and Mihai Manea. Axioms for deferred acceptance. *Econometrica*, 78(2):633–653, 2010.

Aleksei Y Kondratev and Alexander S Nesterov. Minimal envy and popular matchings. *European Journal of Operational Research*, 296(3):776–787, 2022.

Hyukjun Kwon and Ran I Shorrer. Justified-envy-minimal efficient mechanisms for priority-based matching. *Available at SSRN 3495266*, 2020.

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.

Vikram Manjunath and Alexander Westkamp. Strategy-proof exchange under trichotomous preferences. *Journal of Economic Theory*, 193:105197, 2021.

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

Vincent Meisner and Jonas von Wangenheim. School choice and loss aversion. *Available at SSRN 3985777*, 2021.

Markus Möller. Transparent matching mechanisms. Working paper, University of Cologne, 2021a.

Markus Möller. Envy and strategic choice. Working paper, University of Cologne, 2021b.

Thayer Morrill. Making just school assignments. *Games and Economic Behavior*, 92:18–27, 2015.

Vai-Lam Mui. The economics of envy. *Journal of Economic Behavior & Organization*, 26(3):311–336, 1995.

Alexander S Nesterov. Fairness and efficiency in strategy-proof object allocation mechanisms. *Journal of Economic Theory*, 170:145–168, 2017.

Siqi Pan. The instability of matching with overconfident agents. *Games and Economic Behavior*, 113:396–415, 2019.

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.

W Gerrod Parrott and Richard H Smith. Distinguishing the experiences of envy and jealousy. *Journal of Personality and social psychology*, 64(6):906, 1993.

Parag A Pathak and Tayfun Sönmez. Leveling the playing field: Sincere and sophisticated players in the boston mechanism. *American Economic Review*, 98(4):1636–52, 2008.

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

Alvin E Roth. The economics of matching: Stability and incentives. *Mathematics of operations research*, 7(4):617–628, 1982a.

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

Alvin E Roth. The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica*, 70(4):1341–1378, 2002.

Alvin E. Roth. Deferred acceptance algorithms: History, theory, practice, and open questions. *International Journal of Game Theory*, 36(3-4):537–569, 2008.

Alvin E Roth and Uriel G Rothblum. Truncation strategies in matching markets—in search of advice for participants. *Econometrica*, 67(1):21–43, 1999.

Alvin E Roth, Tayfun Sönmez, and M Utku Ünver. Kidney exchange. *The Quarterly Journal of Economics*, 119(2):457–488, 2004.

Peter Salovey and Judith Rodin. Some antecedents and consequences of social-comparison jealousy. *Journal of Personality and Social Psychology*, 47(4):780, 1984.

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

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

Richard H Smith and Sung Hee Kim. Comprehending envy. *Psychological bulletin*, 133(1):46, 2007.

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

Qianfeng Tang and Jingsheng Yu. A new perspective on kesten's school choice with consent idea. *Journal of Economic Theory*, 154:543–561, 2014.

Chung-Piaw Teo and Jay Sethuraman. The geometry of fractional stable matchings and its applications. *Mathematics of Operations Research*, 23(4):874–891, 1998.

Clayton Thomas. Classification of priorities such that deferred acceptance is obviously strategyproof. *arXiv preprint arXiv:2011.12367*, 2020.

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

Peter Troyan and Thayer Morrill. Obvious manipulations. *Journal of Economic Theory*, 185:104970, 2020.

Peter Troyan, David Delacrétaz, and Andrew Kloosterman. Essentially stable matchings. *Games and Economic Behavior*, 120:370–390, 2020.

Helena Wenninger, Christy MK Cheung, and Hanna Krasnova. College-aged users behavioral strategies to reduce envy on social networking sites: A cross-cultural investigation. *Computers in Human Behavior*, 97:10–23, 2019.

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

Robert Wilson. Game-theoretic analyses of trading processes. In *Advances in Economic Theory, Fifth World Congress*, pages 33–70, 1987.

Kyle Woodward. Self-auditable auctions. Working paper, University of North Carolina, 2020.

Marcel Zeelenberg and Rik Pieters. A theory of regret regulation 1.0. *Journal of Consumer psychology*, 17(1):3–18, 2007.

Lin Zhou. On a conjecture by gale about one-sided matching problems. *Journal of Economic Theory*, 52(1):123–135, 1990.

Daniel John Zizzo and Andrew J Oswald. Are people willing to pay to reduce others' incomes? *Annales d'Economie et de Statistique*, pages 39–65, 2001.