Essentially Stable Matchings*

Peter Troyan
University of Virginia

David Delacrétaz
University of Oxford

Andrew Kloosterman†
University of Virginia

January 29, 2020

Abstract

We propose a solution to the conflict between fairness and efficiency in one-sided matching markets. A matching is *essentially stable* if any priority-based claim initiates a chain of reassignments that results in the initial claimant losing the object. We show that an essentially stable and Pareto efficient matching always exists and that Kesten's (2010) EADA mechanism always selects one while other common Pareto efficient mechanisms do not. Additionally, we show that there exists a student-pessimal essentially stable matching and that the Rural Hospital Theorem extends to essential stability. Finally, we analyze the incentive properties of essentially stable mechanisms.

JEL Classification: C78, D47, D61, D63, I20
Keywords: matching, stability, fairness, efficiency, school choice

*Authors' names appear in random order. We thank (also in random order) Meg Meyer, Umut Dur, Thayer Morrill, Alex Teytelboym, Paul Klemperer, Mustafa Oğuz Afacan, Martin van der Linden, Georgy Artemov, Bettina Klaus, and Samson Alva for helpful discussions regarding this project.

†Corresponding author: David Delacrétaz, Department of Economics and Nuffield College, University of Oxford, New Road, Oxford OX1 1NF. Emails: troyan@virginia.edu, david.delacretaz@economics.ox.ac.uk, and ask5b@virginia.edu. A large amount of the work on this project was conducted while Delacrétaz was at the University of Melbourne and received support through the Australian Research Council grant DP160101350. Declarations of interest: none.

1 Introduction

There exists a trade-off between efficiency and fairness in one-sided matching problems. The celebrated deferred acceptance (DA) mechanism of Gale and Shapley (1962) always produces a fair matching, and in fact produces the most efficient matching among all fair ones. However, it does not always produce a Pareto efficient matching: there may be unfair matchings that Pareto dominate it. Are all unfair matchings equally unfair? We argue that the answer is no, and that the formal fairness criterion typically used in the literature—stability—excludes many matchings unnecessarily. We propose a new fairness standard called *essential* stability, which takes these matchings into account and is not at odds with efficiency.

There are many real-world examples of problems that fit into our framework, but perhaps the largest and most important is public school choice as instituted in many cities across the United States and around the world. Fairness is a crucial concern for many school districts because they must be able to justify why one student is admitted to a school and another is rejected. This is typically done by assigning priorities to each student at each school according to some set criteria (which may vary across school districts) and then running a well-defined matching mechanism that takes these priorities and the student preferences as inputs.

In the one-sided matching framework, the standard approach in the literature is to analyze efficiency from the perspective of the students (the other side of the market consists of school seats to be "consumed") and to use the mathematical definition of *stability* as a formal fairness criterion. Given a matching, a student is said to have a (justified) *claim* to school A if she prefers A to her assignment and either she has higher priority than another student who is assigned to A or A has excess capacity. A matching is *stable* if there are no claims. Stability was introduced by Gale and Shapley (1962) as an equilibrium concept in a two-sided matching market with agents on both sides, where stability guarantees that no two agents prefer to be matched to each other than with their assigned partners. In one-sided markets, stability is a fairness criterion in the sense that stable matchings eliminate *justified envy* (Abdulkadiroğlu and Sönmez, 2003). This allows a school district to justify why some student j is not admitted to a school A (even though she prefers it) and another student i is: i has higher priority than j at A. In fact, some authors simply equate the terms fairness and stability (see, e.g., Balinski and Sönmez (1999)).

In this paper, we use the term "fair" in its normative sense and the term "stable" to describe the aforementioned mathematical property of a matching. At first glance, the use of stability as the standard for fairness seems very reasonable, because it ensures that there are no claims. However, this simple definition actually misses a subtle (and important) point: if a student were to have a claim to a seat at a school, granting her claim displaces a student currently assigned to that school. This student will then have to be reassigned, and, using the same justification as the initial student, she can claim her favorite school at which she has high enough priority. This will displace yet another student, and so on. This chain of reassignments ends in one of two ways, whichever comes first: either (i) the initial claimant is displaced from the school she claimed or (ii) a displaced student is reassigned to a school with an empty seat or takes her outside option. In the former case, the initial claimant ultimately does not receive the school to which she laid claim, and so the claim is *vacuous*.

We propose a new definition that expands the set of stable matchings by allowing vacuous claims. We argue that this weaker fairness criterion captures the essence of stability as a fairness standard. If a student has a vacuous claim, she does not have justified envy in the sense that she ultimately will not be matched to the school she claims. For this reason, we call a matching in which all claims are vacuous *essentially stable*.

Our first results show that this expansion of the set of admissible matchings is substantive. We show that Kesten's EADA mechanism always produces an essentially stable matching (Theorem 1). Since EADA also produces a Pareto efficient matching (when all students consent), the set of essentially stable matchings always contains at least one Pareto efficient matching (Corollary 1). This is in contrast to the stable set, which may not contain a Pareto efficient matching. Thus, essential stability provides a solution to the trade-off between fairness and efficiency.[1]

We then show that there may be multiple essentially stable and Pareto efficient matchings (Proposition 1). This suggests that other Pareto efficient mechanisms may produce different essentially stable matchings. However, we show that other classic Pareto efficient mechanisms such as Top Trading Cycles (TTC), a variant of TTC where DA assignments are used as the endowments (DA+TTC), and Immediate Acceptance (IA), do not always produce an essentially stable matching (Proposition 2).

Proposition 1 immediately implies that there may not exist a student-optimal essentially stable matching (this contrasts with the set of stable matchings which is known to form a

[1]The trade-off is consequential as the efficiency losses from imposing stability can be significant in practice. For example, using data from eighth-grade assignment in New York City, Abdulkadiroğlu et al. (2009) show that, by moving from the DA matching to a Pareto efficient matching, over 4,000 students could be made better off each year (on average) without making a single student worse off.

lattice). On the other hand, our second main result is that a student-pessimal essentially stable matching does exist, and is the same as the student-pessimal stable matching (Theorem 2). We also show that the Rural Hospital Theorem (Roth, 1986) holds for the set of essentially stable matchings (Proposition 3).

The existence of a student-pessimal essentially stable matching is critical for incentive results regarding essentially stable mechanisms. By combining this theorem with results from Alva and Manjunath (2019a,b), we show that DA is the only essentially stable and strategyproof mechanism (Proposition 4). This points to a trade-off among essential stability, Pareto efficiency, and strategyproofness: a mechanism exists to obtain any two of these properties, but no mechanism achieves all three. Nevertheless, strategyproofness is a demanding criterion, and recent work by Troyan and Morrill (2019) has investigated the severity of manipulations. Their results, when combined with our Theorem 2, imply that any essentially stable mechanism is *not obviously manipulable* (Proposition 5), and so the incentive problem may not be so severe. Overall, we believe essentially stable and Pareto efficient mechanisms constitute a potentially attractive alternative class of mechanisms that are worthy of further theoretical and empirical investigation.

Related Literature

Our paper is related to a growing literature that investigates weaker definitions of stability that are compatible with efficiency. Most papers in this literature are loosely based on the idea that a student with a claim must propose an alternative matching that is free of any counter-claims (and possibly some other conditions too) or else her initial claim can be disregarded. Work in this vein includes Morrill (2015), who introduces the concept of a *just* assignment, Alcalde and Romero-Medina (2015), who introduce the concept of $\tau-$*fairness*, and Cantala and Pápai (2014), who discuss the concepts of *reasonable stability* and *secure stability*.[2] Also very closely related is Ehlers and Morrill (2019), who define a *legal* set of assignments, where, in legal terminology, a student i's claim at a school c is not redressable (and thus can be disregarded) unless i can propose an alternative assignment (i.e., matching) that is "legal" and at which she is assigned to c.[3] They introduce an iterative procedure for finding the set of legal assignments, which is equivalent to the *von Neumann-Morgenstern stable set* (Von Neumann and Morgenstern, 1944).[4] Tang and Zhang (2016) introduce their

[2]Reasonable stability was first defined in Kesten (2004).
[3]See also Morrill (2016) for an earlier iteration of these results.
[4]Ehlers (2007) studies vNM stable sets in the context of marriage markets (see also Wako (2010)).

own new definition of weak stability for school choice problems that is also closely related to vNM stable sets. While sharing a similar motivation, all are independent properties, and in Appendix C, we show formally that these other concepts are distinct from ours.

A different strand of related literature focuses on mechanisms (rather than matchings), and in particular on a class of mechanisms that, besides simply asking students to report their preferences, also asks them if they "consent" to having their priority violated. The goal is then to use a mechanism that ensures students cannot gain from not consenting, or in other words the mechanism should be *no-consent-proof.* This is the original approach taken in Kesten's (2010) paper introducing the EADA mechanism, and was further expanded by Dur et al. (2015), who show that EADA is the unique constrained efficient mechanism that Pareto dominates DA and is no-consent-proof.[5] While related, there is an important conceptual distinction between the approaches. Essential stability is a fairness criterion that relaxes stability; no-consent-proofness is a justification for why, given a particular mechanism, students should affirmatively consent to violations of the classical definition of stability. In other words, essential stability is a property of matchings (for given preferences and priorities), while no-consent-proofness is a property of mechanisms, which are conceptually more complex than matchings.[6]

We believe that essential stability formalizes the normative idea of fairness in a way that is a particularly straightforward to explain to non-experts, which makes it well-suited for practical applications. This is because understanding essential stability only requires understanding the difference between a vacuous and a non-vacuous claim, which can be explained to policy makers by walking them through an example of a reassignment chain, step-by-step, to highlight why essential stability is a natural notion of fairness. Nevertheless, given the importance of reconciling efficiency and fairness, it is beneficial to have multiple ways to think about the issue, and the other ideas discussed above provide valuable insights and elegant theoretical justifications for what it means for a matching to be fair. As such, we view our more applied approach as complementary to this literature.

From a broader perspective, our paper also contributes to a growing literature on how to define stability when agents may anticipate more than one step of blocking, a question that

[5] Dur et al. (2015) also have a relaxation of stability called *partial stability* that exogenously takes a subset of priority violations as allowable. A partially stable matching is then constrained efficient if it is not Pareto dominated by any other partially stable matching.

[6] Of course, the definition of essential stability can (and will) be easily extended to mechanisms in the natural way by defining a mechanism as essentially stable if it always produces an essentially stable matching. While EADA is both no-consent-proof and essentially stable, there is no a priori logical relationship between the two concepts (see Appendix C for examples of the independence).

has received considerable attention in other game-theoretic contexts. The central concept in this literature is called farsightedness: an outcome is stable if there does not exist a series of blocks that culminate in better outcomes for every agent who participates in it. Farsightedness was first introduced by Harsanyi (1974) as a criticism of von Neumann-Morgenstern stable sets. In more recent work, Ray and Vohra (2015) and Dutta and Vohra (2017) carefully address farsightedness in coalition formation games. Page Jr et al. (2005) and Herings et al. (2009) consider related issues in network formation games. While similar in the sense that both look more than one step ahead, the ultimate effect of essential stability is actually opposite to that of farsightedness: farsightedness excludes myopically stable outcomes by providing a series of blocks that makes the initial outcome ultimately unstable, while essential stability includes myopically unstable outcomes by showing that a series of reassignments nullifies the original block. This allows expanding the set of admissible matchings in order to reach the Pareto frontier.

The remainder of the paper is organized as follows. We formally introduce essential stability in Section 2. Section 3 is devoted to exploring Pareto efficiency and essential stability. In Section 4, we show that there exists a student-pessimal essentially stable matching and investigate the consequences of that result. Section 5 concludes. All proofs not in the main text can be found in the appendix.

2 Preliminaries

2.1 Model

There is a set of **students** S who are to be assigned to a set of **schools** $C \cup \{\emptyset\}$. Each student is to be assigned to one school in $C \cup \{\emptyset\}$, where being assigned to $\emptyset$ is interpreted as remaining unmatched or taking some outside option. Each $c \in C \cup \{\emptyset\}$ has a **capacity** q_c, which is the number of students that can be assigned to it. We assume $q_\emptyset > |S|$, which captures that the outside option is not scarce, i.e., every student has the option to remain unmatched. Let $q = (q_c)_{c \in C \cup \{\emptyset\}}$ denote a profile of capacities.

Each $i \in S$ has a strict (complete, transitive, and antisymmetric) **preference relation** R_i over $C \cup \{\emptyset\}$. We use P_i to denote the asymmetric part of R_i, i.e., aP_ib if and only if aR_ib and $a \neq b$. We call school $c \in C$ **acceptable** to student $i \in S$ if $cP_i\emptyset$ and **unacceptable** otherwise. Similarly, each $c \in C \cup \{\emptyset\}$ has a strict **priority relation** $\succeq_c$ over S, and we analogously use $\succ_c$ to denote the asymmetric part of $\succeq_c$.[7]

For concreteness, we use the school choice terminology throughout as it is the best-known application; however, the model can be applied to many other priority-based matching problems, such as the military assigning cadets to branches (Sönmez, 2013), universities assigning students to dormitories (Chen and Sönmez, 2002), or cities assigning public housing units to tenants (Abdulkadiroğlu and Sönmez, 1999).

A **matching** is a correspondence $\mu : S \cup C \cup \{\emptyset\} \to S \cup C \cup \{\emptyset\}$ such that, for all $(i, c) \in S \times (C \cup \{\emptyset\})$, $\mu(i) \in C \cup \{\emptyset\}$, $\mu(c) \subseteq S$, $\mu(i) = c$ if and only if $i \in \mu(c)$, and $|\mu(c)| \leq q_c$. A matching ν **Pareto dominates** a matching μ if $\nu(i) R_i \mu(i)$ for all $i \in S$, and $\nu(i) P_i \mu(i)$ for at least one $i \in S$. A matching μ is **Pareto efficient** if it is not Pareto dominated by any other matching ν. Note that Pareto efficiency is evaluated only from the perspective of the students, and not the schools. This is a standard view in the mechanism design approach to school choice, beginning with the seminal papers of Balinski and Sönmez (1999) and Abdulkadiroğlu and Sönmez (2003).

In addition to Pareto efficiency, in many applications (particularly in school choice), market designers also care about fairness. Given a matching μ, we say student i **claims a seat at school** $c \in C \cup \{\emptyset\}$ if (i) $c P_i \mu(i)$ and (ii) either $|\mu(c)| < q_c$ or $i \succ_c j$ for some $j \in \mu(c)$.[8] We will sometimes use (i, c) to denote i's claim to c. If no student claims a seat at any school, then we say μ is **stable**. Stability is a fairness criterion in the sense that it ensures priorities are respected: a student only misses out on a school she wants if that school is filled to capacity with higher-priority students.

Let $\mathcal{M}$ denote the set of all possible matchings, and $\mathcal{P}$ denote the set of all possible preference relations. A **mechanism** $\psi : \mathcal{P}^{|S|} \to \mathcal{M}$ is a function that assigns a matching to each possible preference profile that can be submitted by the students. That is, for any mechanism ψ and profile of preferences $P = (P_i)_{i \in S} \in \mathcal{P}^{|S|}$, $\psi(P)$ is the matching determined by ψ when the submitted preferences are P. We write $\psi_i(P)$ to denote the school to which ψ assigns i after the reports P.[9]

The properties of matchings are adapted to mechanisms by saying the property holds for

[7] When comparing sets of students, we assume the priority relation satisfies *responsiveness* (Roth, 1985): for any $I \subset S$, any $i, j \in S \setminus I$, and any $c \in C \cup \{\emptyset\}$, $I \cup \{i\} \succ_c I \cup \{j\}$ whenever $i \succ_c j$.

[8] Note that if $\emptyset P_i \mu(i)$, and so if i is assigned to an unacceptable school at μ, then $|\mu(\emptyset)| < q_\emptyset < |S|$ and i will claim a seat at $\emptyset$. In this way, our stability concepts below implicitly incorporate the standard notion of *individual rationality*. Also, the $|\mu(c)| < q_c$ case is often called *non-wastefulness* and so our definition also incorporates this standard notion.

[9] Since we assume that students have (and report) strict preferences over $C \cup \{\emptyset\}$, we write mechanisms as a function of P rather than R.

the mechanism if it holds for each possible report. That is, mechanism ψ is **Pareto efficient** if $\psi(P)$ is a Pareto efficient matching for all P. Mechanism ψ is **stable** if $\psi(P)$ is a stable matching for all P. And, later, we will also address incentives in the reporting game, where we say ψ is **strategyproof** if $\psi_i(P) R_i \psi_i(P'_i, P_{-i})$ for all i, $P = (P_i, P_{-i})$, and P'_i. In words, reporting true preferences is a dominant strategy.

In this paper, we consider five mechanisms that have been proposed in the matching and school choice literature. Each of these mechanisms is defined as the outcome of a certain algorithm, i.e., for any priority profile, each algorithm defines a mechanism (mapping from preference profiles to matchings). The mechanisms we consider are:

- Student-proposing deferred acceptance (DA; Gale and Shapley, 1962; Abdulkadiroğlu and Sönmez, 2003)
- Efficiency-adjusted deferred acceptance (EADA; Kesten, 2010)
- Top trading cycles (TTC; Shapley and Scarf, 1974; Abdulkadiroğlu and Sönmez, 2003)
- DA followed by TTC (DA+TTC; Alcalde and Romero-Medina, 2015)
- Immediate acceptance (IA, also called the 'Boston' mechanism; Abdulkadiroğlu and Sönmez, 2003).

DA is the benchmark mechanism in school choice as well as several other applications; in fact, it is used in practice in many cities including New York, Boston, and New Orleans. The reason for its popularity is that DA is stable, strategyproof, and Pareto dominates every other stable mechanism. However, the shortcoming of DA is that it is not Pareto efficient. The other four mechanisms in the list are Pareto efficient; hence, they are not stable. We will define stability more broadly and show that only one, EADA, is stable under our broader definition. As these mechanisms are standard in the literature, we relegate formal definitions to Appendix A (except for EADA, which we will define in Section 3).

2.2 Motivating Example

We now present an example that both shows that there may be unstable matchings that Pareto dominate the DA outcome, as well as highlights the key insight behind the new definition of essential stability which will be introduced in the next section.

Example 1. Let there be 5 students, $S = \{i_1, i_2, i_3, i_4, i_5\}$, and 5 schools with capacity 1, $C = \{A, B, C, D, E\}$.[10] The priorities and preferences are given in the following tables.

$\succ_A$	$\succ_B$	$\succ_C$	$\succ_D$	$\succ_E$
i_1	i_2	i_3	i_4	i_5
i_2	i_3	i_4	i_5	$\vdots$
i_4	i_1	i_2	i_3	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	

P_{i_1}	P_{i_2}	P_{i_3}	P_{i_4}	P_{i_5}
$\dagger B$	$\dagger C*$	$B*$	$\dagger A$	D
$\boxed{A*}$	A	$\dagger D$	C	$\boxed{\dagger E*}$
$\vdots$	$\boxed{B}$	$\boxed{C}$	$\boxed{D*}$	$\vdots$
	$\vdots$	$\vdots$	$\vdots$	

The table on the right indicates three different potential matchings, a matching $\mu^\square$ (denoted by boxes $\square$), and two Pareto efficient matchings μ^* (denoted by stars $*$) and $\mu^\dagger$ (denoted by daggers $\dagger$). The DA matching in this example is $\mu^\square$, which therefore can readily be shown to be stable. It is, however, not Pareto efficient: it is easy to see that it is Pareto dominated by both μ^* and $\mu^\dagger$. This directly implies that μ^* and $\mu^\dagger$ are both unstable. In fact, at μ^*, student i_4 claims the seat at C (because $i_4 \succ_C i_2 = \mu^*(C)$) and at $\mu^\dagger$, student i_3 claims the seat at B (because $i_3 \succ_B i_1 = \mu^\dagger(B)$) and student i_5 claims the seat at D (because $i_5 \succ_D i_3 = \mu^\dagger(D)$).

There are simpler examples to show that the DA matching may not be Pareto efficient. We present this one to illustrate the main point of our paper, which is that not all instability is the same. We argue that μ^* is truly unstable while $\mu^\dagger$ is not.

To understand our argument, consider $\mu^\dagger$ first. Suppose student i_3 claims the seat at school B. If we grant i_3's claim and assign her to B, then student i_1 becomes unmatched. Student i_1 must be assigned somewhere, and (using the same logic as i_3), she can ask to be assigned to A, her next most-preferred school where she has higher priority than the student who is matched to it (student i_4). Granting i_1's claim just as we did i_3's, she is assigned to A and now student i_4 is unmatched. Student i_4 then asks for C, which is her most preferred school where she has high enough priority to be assigned. Student i_2 is now unmatched, and asks for B,[11] which means student i_3 is removed from B. In summary, student i_3 starts by claiming B. If her request is granted based on the fact that $i_3 \succ_B i_1$, then we must also grant the next request of i_1, since she has the same justification for claiming A as i_3 did for claiming B. Continuing, we see that ultimately another student with higher priority than

[10] In all our examples, we do not use the outside option $\emptyset$ and so we omit it for simplicity. To formally match the model, it could always be added in at the bottom of each student's preference relation.

[11] Note that her next most preferred school is A, but school A is now assigned to i_1 and $i_1 \succ_A i_2$, so she cannot get A and must go to B.

i_3 at B (in this case i_2) ends up claiming it, and so i_3's initial claim is unfounded, or as we will call it, vacuous. Student i_5's claim to D also begins a chain of reassignments where eventually i_4 takes D away, so i_5's claim is also vacuous.

Now let us contrast this with the instability found in matching μ^*. Assume that student i_4 claims the seat at C, and this request is granted. Following similar logic to the above, i_2 then asks for B, and i_3 asks for D. This is the end of our reassignments because D is the school that i_4 gave up to claim C. In this case, the original claimant's (student i_4) request does *not* result in her ultimately losing the school she claimed to a higher-priority student, and therefore this claim is not vacuous in the manner that the claims at $\mu^\dagger$ were.

Thus, both μ^* and $\mu^\dagger$ are unstable, but in different ways (which will be made more precise in the next section). What is more, the way in which they are different is straightforward. It would be easy to explain to a non-expert why the claims of i_3 and i_5 at $\mu^\dagger$ are vacuous by showing them the chain of reassignments as we have just done. Our new definition of stability is designed to capture this idea and, in the process, recover inefficiencies by expanding the set of permissible matchings to include those like $\mu^\dagger$, but still exclude those like μ^*.

2.3 Essentially Stable Matchings

We now formalize the intuition from the previous example. Recall that, fixing a matching μ, we use the notation (i, c) to denote i's claim to a seat at c.

Definition 1. Consider a matching μ and a claim (i, c). The **reassignment chain** Γ **initiated by claim** (i, c) is the list

$$i^0 \to c^0 \to i^1 \to c^1 \to \cdots$$

where,

- $i^0 = i$, $\mu^0 = \mu$, $c^0 = c$ and, for each $k \geq 1$,
- $i^k \in S$ is the lowest-priority student in $\mu^{k-1}(c^{k-1})$ according to $\succ_{c^{k-1}}$,
- μ^k is defined as: $\mu^k(i^k) = \emptyset$, $\mu^k(i^{k-1}) = c^{k-1}$, and $\mu^k(j) = \mu^{k-1}(j)$ for all $j \in S \setminus \{i^{k-1}, i^k\}$,
- if $i^k = i$, the chain terminates,

- if $i^k \neq i$, then $c^k \in C$ is i^k's most preferred school to which she has a claim at μ^k if such a school exists; otherwise, $c^k = \emptyset$. If $|\mu^k(c^k)| < q_{c^k}$, the chain terminates.

For a reassignment chain Γ initiated by claim (i, c) at matching μ (that terminates at step K), the **final matching** μ^Γ is defined as: $\mu^\Gamma(i^K) = \mu(i)$ if $i^K = i$ or $\mu^\Gamma(i^K) = c^K$ if $i^K \neq i$ and $\mu^\Gamma(j) = \mu^K(j)$ for all $j \in S \setminus \{i^K\}$.

A reassignment chain ends in one of two ways: either a student $j \neq i$ claims a seat at a school with excess capacity (which could be $\emptyset$), leaving i matched to c, or c rejects i. In the first case, i's claim is valid in the sense that she is still matched to c at the end of the chain. In the second case, i's claim is not valid in the sense that i is removed from c and so i's claim will not be implemented. Formally, we say that claim (i, c) at matching μ is **vacuous** if $i^K = i$ in the reassignment chain initiated by (i, c) at μ that terminates at step K.

Definition 2. Matching μ is **essentially stable** if all claims at μ are vacuous. If there exists at least one claim at μ that is not vacuous, μ is **strongly unstable**.

We also define these terms for mechanisms analagously to above. A mechanism ψ is **essentially stable** if $\psi(P)$ is an essentially stable matching for all P. If ψ is not essentially stable, then we say that it is **strongly unstable**.

Returning to Example 1, we can check that $\mu^\dagger$ is essentially stable, while μ^* is strongly unstable. As we showed above, at $\mu^\dagger$, the claims (i_3, B) and (i_5, D) are vacuous, because they ultimately result in the initial claimant losing the seat she claimed to a higher-priority student. At μ^*, on the other hand, the reassignment chain initiated by (i_4, C) ends with i_4 assigned to C. Thus, i_4's claim is not vacuous.

Two aspects of our definition for reassignment chains may be of concern. First, if $i^K = i$, then the reassignment chain immediately ends. But i may still have a claim to school c^K that she prefers to $\mu(i)$. Instead of ending the chain and matching i to $\mu(i)$, we could match i to c^K and let the reassignment chain continue. At the end of that reassignment chain, i may still be matched to c^K (or to another school that she prefers to $\mu(i)$), arguably making her initial claim not entirely vacuous. Second, some claims in the reassignment chain may themselves be vacuous and therefore should not be allowable as the reassignment chain progresses. In Appendix E, we propose two alternative definitions of reassignment chains to show that allowing i to make additional claims or removing vacuous claims within the reassignment chain does not affect essential stability. We chose Definition 1 over other possible definitions of reassignment chains that would lead to an equivalent definition of essential stability because it straightforwardly states the essence of the issue; the original claimant either keeps the seat she has claimed or she loses it, in which case her claim is vacuous.

3 Essential Stability and Pareto Efficiency

Essential stability relaxes stability in a natural way, thereby increasing the set of matchings that are classified as fair. Given that stability and Pareto efficiency are mutually incompatible, the main question we investigate is whether there exists an **essentially stable and Pareto efficient (ESPE)** matching and, if so, whether there exists a mechanism that always finds one. In this section, we answer both questions in the affirmative by showing that Kesten's (2010) efficiency-adjusted deferred acceptance (EADA) mechanism is essentially stable. We then show that, while there may be multiple ESPE matchings, all of the other common Pareto efficient mechanisms introduced in Section 2 are strongly unstable.

We first show that the EADA mechanism is essentially stable. To prove this, we use the **Simplified Efficiency-Adjusted Deferred Acceptance (SEADA)** mechanism. SEADA was introduced by Tang and Yu (2014) as a simplification of the original EADA mechanism, and they show that the two mechanisms are outcome-equivalent.[12] Following their terminology, we say that a school c is **underdemanded** at matching μ if $\mu(i) R_i c$ for all i. That is, all students who are not matched to c strictly prefer their own assignment.

Simplified Efficiency-Adjusted Deferred Acceptance (SEADA)

Round t Compute the DA matching on the submarket (the whole market in the first round) at the beginning of round t. Identify the schools that are underdemanded, and for each student at these schools, permanently assign them to their DA matching. Create the submarket for round $t+1$ as a market with all the schools in $C \cup \{\emptyset\}$ that are not underdemanded in round t and the students who have not been permanently assigned.

Termination The algorithm terminates at the first round T where all students are permanently assigned.

Theorem 1. *The matching produced by the SEADA mechanism is essentially stable.*

[12] Kesten (2010) and Tang and Yu (2014) define slightly more general classes of mechanisms that ask students whether they "consent" to having their priority violated. The definition given below is the version of SEADA in which all students consent which ensures that the resulting matching is always Pareto efficient.

The proof proceeds by considering an arbitrary claim (i, c) and then defining two alternative preference profiles, one that gives i her SEADA assignment when DA is run and one with a DA rejection chain that coincides with the reassignment chain initiated by the claim (i, c) until the reassignment chain terminates. Using the fact that DA is weakly Maskin monotonic (Kojima and Manea, 2010), we show that both profiles lead to the same DA assignment for i (i.e., i's SEADA assignment which is not c) and so the reassignment chain must terminate with i removed from c. The full details can be found in the appendix.

As the EADA or equivalent SEADA mechanism (hereafter we use the notation (S)EADA to reference them) produces a Pareto efficient matching, the existence of an ESPE matching is an immediate corollary.

Corollary 1. *There exists an ESPE matching.*

Theorem 1 provides a clear justification for using the (S)EADA mechanism in practice when Pareto efficiency is an important concern for the school district, because it achieves Pareto efficiency while only allowing vacuous claims. At the same time, (S)EADA finds just one possible ESPE matching, and a natural question is whether there are others. The next proposition shows that, while existence is guaranteed, ESPE matchings are not in general unique.

Proposition 1. *There may exist multiple ESPE matchings.*

This proposition is proved with the following example.

Example 2. Let there be 4 students, $S = \{i_1, i_2, i_3, i_4\}$, and 4 schools with capacity 1, $C = \{A, B, C, D\}$. The priorities and preferences are given in the following tables.

$\succ_A$	$\succ_B$	$\succ_C$	$\succ_D$
i_3	i_1	i_2	i_4
i_1	i_4	i_3	i_3
i_2	$\vdots$	$\vdots$	$\vdots$
$\vdots$			

P_{i_1}	P_{i_2}	P_{i_3}	P_{i_4}
$\boxed{A}$	$\dagger A$	$\boxed{D}$	$\boxed{B}$
$\dagger B$	$\boxed{C}$	$\dagger C$	$\dagger D$
$\vdots$	$\vdots$	A	$\vdots$
		$\vdots$	

$\mu^\square$ and $\mu^\dagger$ are two ESPE matchings. $\mu^\square$ is the DA matching, which in this example happens to be Pareto efficient. $\mu^\dagger$ is not stable as i_1 has a claim to A. However, it is the only claim and the reassignment chain it initiates is

$$i_1 \to A \to i_2 \to C \to i_3 \to A \to i_1;$$

therefore the claim is vacuous and $\mu^\dagger$ is essentially stable. □

The multiplicity of ESPE matchings naturally raises the question as to whether other well-known Pareto efficient mechanisms are essentially stable. We answer this question negatively for the three commonly-proposed Pareto efficient mechanisms that were introduced in Section 2: TTC, DA+TTC, and IA.

Proposition 2. *The TTC, DA+TTC, and IA mechanisms are each strongly unstable.*

We show this proposition by returning to Example 1 and showing that each of these mechanisms produces a strongly unstable matching. It is straightforward to calculate that the TTC outcome for Example 1 is the matching μ^*, which we showed earlier is strongly unstable. Next, consider DA+TTC. Since each student has top priority at their DA school, the TTC stage of DA+TTC is equivalent to the standard TTC mechanism (with no DA stage); hence DA+TTC also produces μ^*.

We finally consider IA. In the first round, all students except i_1 are accepted by their first preference. In the second round, i_1 can only apply to E as all other schools are full. IA terminates and produces

$$\mu^{IA} = \begin{pmatrix} A & B & C & D & E \\ i_4 & i_3 & i_2 & i_5 & i_1 \end{pmatrix}.$$

The reassignment chain initiated by the claim (i_1, A) is

$$i_1 \to A \to i_4 \to C \to i_2 \to B \to i_3 \to C \to i_4 \to D \to i_5 \to E;$$

therefore, i_1's claim to A is not vacuous and μ^{IA} is strongly unstable. □

The possible multiplicity of ESPE matchings directly implies that (S)EADA is not the only ESPE mechanism; however, as Proposition 2 shows, all of the well-known Pareto efficient mechanisms except (S)EADA are strongly unstable. Finding other ESPE mechanisms and defining criteria to select the "best" one constitutes an interesting open problem, though one that is beyond the scope of this paper.

4 Structural and Incentive Properties

A well-known result about the set of stable matchings is that it forms a lattice, which implies the existence of a student-optimal and a student-pessimal stable matching. Moreover, the matching produced by DA, which we denote by μ^{DA}, is the **student-optimal stable matching**; that is, μ^{DA} Pareto dominates all other stable matchings. We denote by μ^p the **student-pessimal stable matching**, which is the matching that is Pareto dominated by all other stable matchings. Formally, for any stable matching μ and any student i, $\mu^{DA}(i) R_i \mu(i) R_i \mu^p(i)$. This begs the questions of whether there exist analogously defined student-optimal and student-pessimal essentially stable matchings. The possible multiplicity of ESPE matchings directly implies a negative answer to the first question but our next result provides an affirmative answer to the second question.

Theorem 2. *μ^p is the student-pessimal essentially stable matching.*

As it turns out, Theorem 2 has important implications for the incentive properties of essentially stable mechanisms, which we present at the end of this section (Propositions 4 and 5).

Theorem 2 relies on three properties of a reassignment chain started by a (vacuous) claim in an essentially stable matching. Let μ be an essentially stable matching and consider the reassignment chain Γ initiated by the claim (i, c) at μ. As μ is essentially stable, (i, c) is a vacuous claim and so Γ ends when c rejects i. Let K be the number of steps of that reassignment chain and recall that μ^Γ denotes the final matching obtained after Γ is carried out. We say that the reassignment chain Γ **affects** student $j \in S$ if $\mu(j) \neq \mu^\Gamma(j)$ and school $d \in C \cup \{\emptyset\}$ if $\mu(d) \neq \mu^\Gamma(d)$. Our first result states that Γ affects students and schools in a monotonic way.

Lemma 1. *For every student j and every school d that is affected by Γ:*

$$\mu(j) P_j \mu^\Gamma(j) \quad \text{and} \quad \mu^\Gamma(d) \succ_d \mu(d).$$

The statement related to schools is straightforward. At each step, a school replaces a student by another with a higher priority; therefore, by the end of the reassignment chain, it is assigned a set of students with a higher priority overall.

The statement related to students is not as obvious. A student is assigned to her favorite school to which she has a claim and may prefer that school to the one that just rejected her. However, as we formally show in the appendix, that school (and any others she subsequently claims that she prefers to her original assignment) always rejects her before the end of the reassignment chain.

Throughout a reassignment chain, students get matched to the school they prefer among those to which they have a claim; therefore any student who is affected by Γ does not have any claim at μ^Γ. As students who are not affected by Γ remain matched to the same school and schools have weakly higher-priority students, the reassignment chain does not create any new claim, as we next formalize:

Lemma 2. *If student $j \in S$ has a claim to school $d \in C \cup \{\emptyset\}$ at μ^Γ, then j has a claim to d at μ.*

Every claim at μ^Γ is a claim at μ and, because μ is essentially stable, any such claim is vacuous at μ. A natural question at this point is whether such a claim can become non-vacuous as a result of a reassignment chain. As we show in the appendix, this is not the case, which implies that reassignment chains preserve essential stability:

Lemma 3. *μ^Γ is essentially stable.*

The proofs Lemmas 1-3 can be found in the appendix. With these lemmas in hand, we can now prove Theorem 2.

Proof of Theorem 2: Consider any essentially stable matching $\mu \neq \mu^p$. Either μ is stable, or there exists a claim (i, c) at μ. In the latter case, it is possible to carry out the reassignment chain Γ induced by that claim in order to obtain μ^Γ. By Lemmas 1 and 3, μ^Γ is Pareto dominated by μ and is essentially stable. By Lemma 2, any claim at μ^Γ is a claim at μ. In addition, (i, c) is a claim at μ by assumption but, as c rejects i at the end of Γ, it is not a claim at μ^Γ. Combining the last two statements implies that μ^Γ has strictly fewer claims than μ. So, starting from μ, it is possible to carry out reassignment chains, one at a time, until a stable matching ν that is Pareto dominated by μ is found. By definition, ν Pareto dominates μ^p; therefore μ Pareto dominates μ^p. □

Reassignment chains initiated by a claim at an essentially stable matching are similar in spirit to *rotations* (Irving, 1985; Irving and Leather, 1986). Starting from any stable matching other than μ^p, it is possible to carry out a rotation to obtain a stable matching that makes all affected students worse off and assigns higher-priority students to all affected schools. As Lemmas 1 and 3 show, reassignment chains achieve the same for essentially stable matchings.

Our analysis reveals an asymmetric structure since an extreme matching exists on one end but not on the other. Clearly, our negative result implies that the set of essentially stable matchings does not form a full lattice (as the set of stable matchings does); however, given the existence of a student-pessimal essentially stable matching, it seems natural to think it may form a semilattice (more precisely, a meet-semilattice with respect to the partial order R). We provide a counterexample in Appendix D to show that, perhaps surprisingly, this is in fact not the case.

Another important property of stable matchings is that the schools that are not filled to capacity are assigned the same set of students at all stable matchings, a property often referred to as the *Rural Hospital Theorem* (Roth, 1986). We can also use Lemmas 1-3 to show that this property extends to essentially stable matchings.[13]

Proposition 3. *(Rural Hospital) For any two essentially stable matchings μ and μ' and for every school $d \in C \cup \{\emptyset\}$, $|\mu(d)| < q_d$ implies $\mu(d) = \mu'(d)$.*

Proof: Consider an essentially stable matching μ. Let d be a school such that $|\mu(d)| < q_d$. That school is not affected by any reassignment chain initiated by a claim (i, c), as otherwise the reassignment chain would end immediately and i's claim would not be vacuous. As in the proof of Theorem 2, starting from μ, carry out the reassignment chains until a stable matching ν is found. As d is not affected by any of these reassignment chains, $\mu(d) = \nu(d)$. As shown by Roth (1986), $\nu(d) = \nu'(d)$ for any two stable matchings ν and ν'. Combining the last two statements yields the desired result. □

A common interpretation of the Rural Hospital Theorem is that the same students are unmatched at all stable matchings. This statement is implied by Proposition 3 as $q_\emptyset > |S|$, and therefore $\emptyset$ is not filled to capacity.

Incentive properties

Theorem 2 allows us to answer important questions regarding the incentive properties of essentially stable mechanisms. Alva and Manjunath (2019a,b) call a mechanism **stable-dominating** if it always produces a matching that weakly Pareto dominates at least one stable matching. By Theorem 2, every essentially stable matching weakly Pareto dominates the student-pessimal stable matching; therefore, every essentially stable mechanism is stable-dominating. Combining this result with Corollary 5 of Alva and Manjunath (2019b) which shows that DA is the only strategyproof and stable-dominating mechanism,[14] we conclude the following:

[13] An alternative proof of this result can be obtained by combining our Theorem 2 with Lemma 2 and Proposition 2 of Alva and Manjunath (2019a), who show that the Rural Hospital Theorem holds for all stable-dominating matchings. We discuss stable-dominating matchings below.

[14] Proposition 2 of Alva and Manjunath (2019a) can also be used to obtain this result.

Proposition 4. *DA is the only essentially stable and strategyproof mechanism.*

By Proposition 4, all other strategyproof mechanisms besides DA are strongly unstable. More broadly, our results shed a new light on the trade-off between fairness, efficiency, and strategyproofness. Proposition 4 shows that no mechanism can achieve all three properties. Among strategyproof mechanisms, the trade-off between efficiency and fairness is very well understood and effectively comes down to choosing between DA and TTC. Our definition opens up a third option: combining efficiency and fairness by using an ESPE mechanism. This trade-off is illustrated in Figure 1.

We believe there are good reasons to at least seriously consider using an ESPE mechanism. First, as discussed in footnote 1, the inefficiency of DA is sizeable. Second, just because a mechanism is not strategyproof does not necessarily imply that it will be manipulated in practice. Building on the recent work of Li (2017) on obviousness in mechanism design, Troyan and Morrill (2019) argue that the existence of some manipulations may be tolerable, so long as these manipulations are not *obvious* manipulations. More formally, given a student i with true preferences P_i, Troyan and Morrill (2019) define a manipulation P_i' as **obvious** if either (i) the worst possible outcome from reporting P_i' is strictly better than the worst possible outcome from P_i or (ii) the best possible outcome from P_i' is strictly better than the best possible outcome from P_i. Analogously, a mechanism is **obviously manipulable** if there is an obvious manipulation. They show that no stable-dominating mechanism is obviously manipulable, which by Theorem 2 implies the following:

Proposition 5. *No essentially stable mechanism is obviously manipulable.*

Given the restrictive nature of strategyproofness, real-world markets often make use of non-strategyproof mechanisms, and many do so quite successfully. While this does not mean that incentives should be ignored, Proposition 5 suggests that ESPE mechanisms may provide a satisfying alternative in practice. Ultimately, the choice between stronger incentive properties and greater efficiency is up to policy makers and more theoretical, experimental, and empirical investigations are needed to inform this trade-off.

5 Conclusion

This paper introduces the concept of essential stability, a weakening of classical stability that allows a matching to have some priority-based claims to seats at schools as long as those

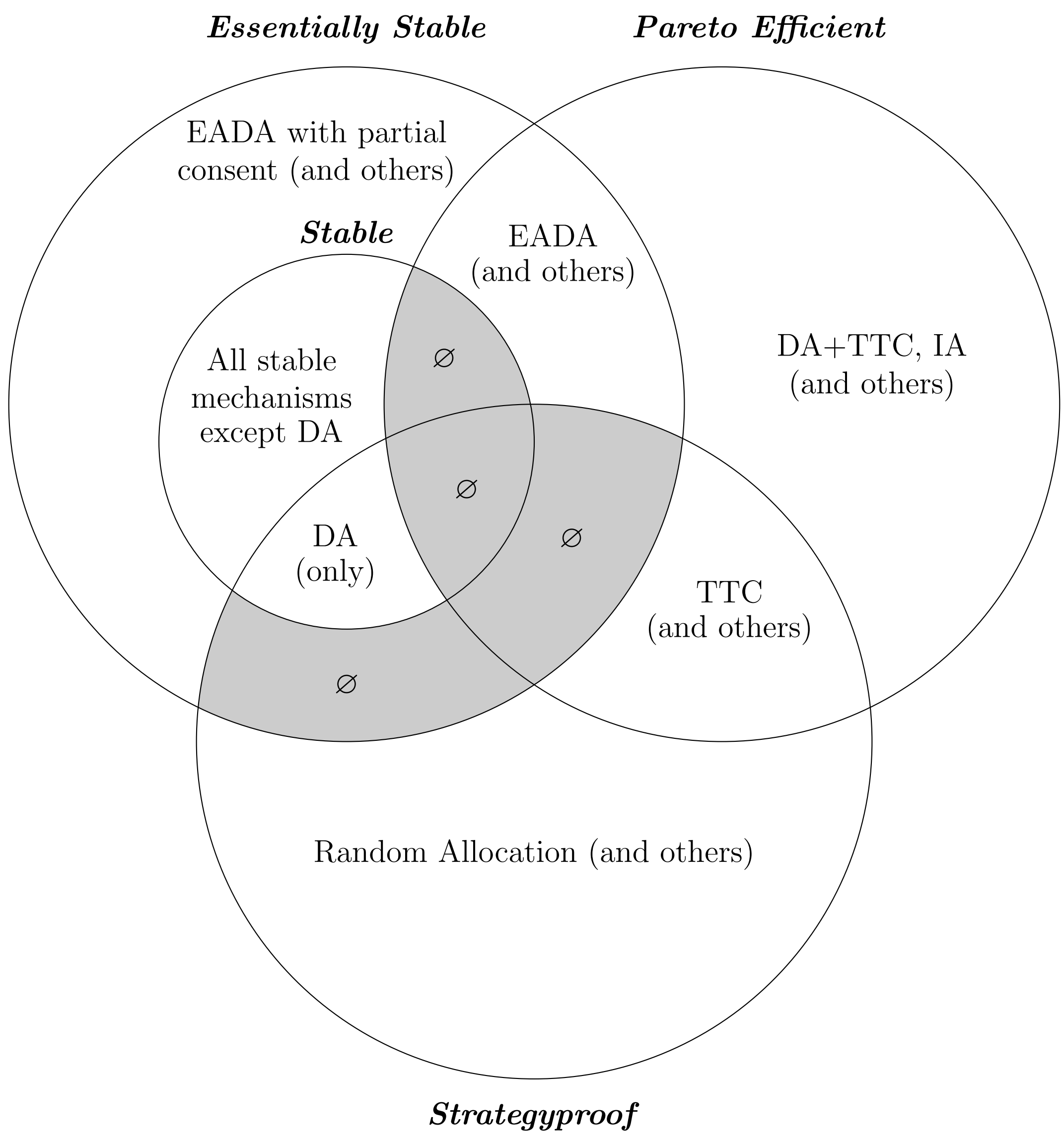

Figure 1: Classification of mechanisms. The gray areas represent combinations of properties that no mechanism can achieve.

claims are vacuous. The motivation for this definition is twofold. First, it is compatible with Pareto efficiency, which can significantly improve the welfare of participants. Second, it still adheres to the principle behind imposing stability as a fairness criterion in the first place: students should not have valid claims. The definition is simple enough that it can easily be explained to non-experts as a reasonable standard of fairness, which we believe constitutes a key advantage for the purpose of practical implementation.

Our paper opens several avenues for future research. First, the existence of multiple ESPE matchings raises the question of whether some can be argued to be more desirable than others. If these matchings could be compared in a meaningful way, it may be possible to improve upon the EADA mechanism by selecting the "best" ESPE matching in each market. Second, essential stability could constitute a useful concept beyond the model studied in this paper; it could prove particularly valuable, for example, in settings where a stable matching is not guaranteed to exist, such as "roommate" matching markets or matching markets with couples.[15] While the right formal definition will likely depend on the particular setting, we hope that the ideas in this paper provide inspiration for thinking about how to appropriately define a fairness criterion that is not only compatible with efficiency, but is also intuitive and convincing to policymakers and market participants.

References

Abdulkadiroğlu, A., P. A. Pathak, and A. E. Roth (2009): "Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match," *American Economic Review*, 99, 1954–1978.

Abdulkadiroğlu, A. and T. Sönmez (1999): "House Allocation with Existing Tenants," *Journal of Economic Theory*, 88, 233–260.

——— (2003): "School Choice: A Mechanism Design Approach," *American Economic Review*, 93, 729–747.

Alcalde, J. and A. Romero-Medina (2015): "Strategy-proof fair school placement," *Available at SSRN 1743082*.

Alva, S. and V. Manjunath (2019a): "Stable-dominating Rules," Working paper, University of Ottawa.

[15]See Hirata et al. (2019) for a related solution concept in the roommate problem.

——— (2019b): "Strategy-proof Pareto-improvement," *Journal of Economic Theory*, 181, 121–142.

Balinski, M. and T. Sönmez (1999): "A Tale of Two Mechanisms: Student Placement," *Journal of Economic Theory*, 84, 73–94.

Cantala, D. and S. Pápai (2014): "Reasonably and Securely Stable Matching," Tech. rep., mimeo.

Chen, Y. and T. Sönmez (2002): "Improving efficiency of on-campus housing: An experimental study," *American Economic Review*, 1669–1686.

Dubins, L. E. and D. A. Freedman (1981): "Machiavelli and the Gale-Shapley algorithm," *American Mathematical Monthly*, 88, 485–494.

Dur, U., A. Gitmez, and O. Yilmaz (2015): "School Choice Under Partial Fairness," *Theoretical Economics*, forthcoming.

Dutta, B. and R. Vohra (2017): "Rational expectations and farsighted stability," *Theoretical Economics*, 12, 1191–1227.

Ehlers, L. (2007): "von Neumann–Morgenstern stable sets in matching problems," *Journal of Economic Theory*, 134, 537–547.

Ehlers, L. and T. Morrill (2019): "(Il)legal Assignments in School Choice," *Review of Economic Studies*, forthcoming.

Gale, D. and L. S. Shapley (1962): "College Admissions and the Stability of Marriage," *The American Mathematical Monthly*, 69, 9–15.

Harsanyi, J. C. (1974): "An equilibrium-point interpretation of stable sets and a proposed alternative definition," *Management science*, 20, 1472–1495.

Herings, P. J.-J., A. Mauleon, and V. Vannetelbosch (2009): "Farsightedly stable networks," *Games and Economic Behavior*, 67, 526–541.

Hirata, D., Y. Kasuya, and K. Tomoeda (2019): "Stability against Robust Deviations in the Roommate Problem," .

Irving, R. W. (1985): "An efficient algorithm for the "stable roommates" problem," *Journal of Algorithms*, 6, 577–595.

Irving, R. W. and P. Leather (1986): "The complexity of counting stable marriages," *SIAM Journal on Computing*, 15, 655–667.

Kesten, O. (2004): "Student placement to public schools in the US: Two new solutions," *Unpublished mimeo, University of Rochester.*

——— (2010): "School Choice with Consent," *Quarterly Journal of Economics*, 125, 1297–1348.

Kojima, F. and M. Manea (2010): "Axioms for Deferred Acceptance," *Econometrica*, 78, 633–653.

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

McVitie, D. G. and L. B. Wilson (1971): "The stable marriage problem," *Communications of the ACM*, 14, 486–490.

Morrill, T. (2015): "Making just school assignments," *Games and Economic Behavior*, 92, 18–27.

——— (2016): "Which School Assignments Are Legal?" Working paper, North Carolina State University.

Page Jr, F. H., M. H. Wooders, and S. Kamat (2005): "Networks and farsighted stability," *Journal of Economic Theory*, 120, 257–269.

Ray, D. and R. Vohra (2015): "The farsighted stable set," *Econometrica*, 83, 977–1011.

Roth, A. E. (1985): "The College Admissions Problem is Not Equivalent to the Marriage Problem," *Journal of Economic Theory*, 36, 277–288.

——— (1986): "On the Allocation of Residents to Rural Hospitals: A General Property of Two-Sided Matching Markets," *Econometrica*, 54, 425–427.

Shapley, L. S. and H. E. Scarf (1974): "On Cores and Indivisibility," *Journal of Mathematical Economics*, 1, 23–28.

Sönmez, T. (2013): "Bidding for Army Career Specialties: Improving the ROTC Branching Mechanism," *Journal of Political Economy*, 121, 186–219.

Tang, Q. and J. Yu (2014): "A new perspective on Kesten's school choice with consent idea," *Journal of Economic Theory*, 154, 543–561.

Tang, Q. and Y. Zhang (2016): "Weak Stability and Pareto Efficiency in School Choice," Working paper, Shanghai University of Finance and Economics.

Troyan, P. and T. Morrill (2019): "Obvious Manipulations," Working paper.

Von Neumann, J. and O. Morgenstern (1944): *Theory of Games and Economic Behavior*, Princeton University Press.

Wako, J. (2010): "A polynomial-time algorithm to find von Neumann-Morgenstern stable matchings in marriage games," *Algorithmica*, 58, 188–220.

A Mechanism definitions

In this section, we provide brief definitions of the algorithms that determine the matchings for the following mechanisms: DA, TTC, DA+TTC, and IA.

Deferred acceptance (DA)

Round t All students apply to their most-preferred school $c \in C \cup \{\emptyset\}$ that has not rejected them. Each school c tentatively accepts the q_c highest-priority students among those who have applied to it (or all, if fewer than q_c apply), and rejects the rest.

Termination The algorithm terminates at the first round T where no students are rejected and then students are permanently assigned to the schools they are tentatively accepted by in this round.

Top Trading Cycles (TTC)

Round t Each student i in the submarket at the beginning of round t (the whole market for $t = 1$) points to i's most-preferred school in the submarket and each school c in the submarket points to c's highest-priority student in the submarket. Find all the cycles: ordered lists where i_1 points to c_1 who points to i_2 etc. ending in c_k pointing to i_1 for some k. All students in cycles are permanently assigned to the school they point to. Create the submarket for round $t+1$ as a market with all the unassigned students and all the schools with capacity $q_{c_{t+1}} > 0$ where $q_{c_{t+1}}$ is calculated as q_{c_t} less the number of students assigned to c in round t (where $q_{c_1} = q_c$).

Termination The algorithm terminates at the first round T where all students are permanently assigned.

Deferred Acceptance + Top Trading Cycles (DA + TTC)

DA round Run the deferred acceptance algorithm. Let μ^{DA} be the resulting matching.

TTC round For each school c, create a new priority relation $\succ'_c$ from $\succ_c$ by raising all students in $\mu^{DA}(c)$ to the top of $\succ'_c$ (and otherwise keeping the order of students unchanged). Run the TTC algorithm using the student's preferences, and the new priority relations.

Immediate Acceptance (IA)

Round t Each student i that is not assigned at the beginning of round t applies to the t^{th}-ranked school on their preference list. For each school c, the q_{c_t}-highest priority students among those who applied to c in round t are permanently assigned to c (or all if fewer than q_{c_t} apply), where $q_{c_{t+1}}$ is calculated as q_{c_t} less the number of students assigned to c in round t (where $q_{c_1} = q_c$).

Termination The algorithm terminates at the first round T where all students are permanently assigned.

B Omitted proofs

In this appendix, we provide proofs of all results that were not proved in the main text. We first present the proof of Theorem 1 and then the proofs of all remaining lemmas (including those from the main text and those introduced in the proof of Theorem 1).

B.1 Proof of Theorem 1

In addition to the notation in the main text, we use two more pieces of notation in this proof. First, we denote by μ^t the matching after round t of SEADA consisting of the permanent matching for all students who have been removed at a round before t and the round t DA matching for all other students. Let T be the number of steps in the SEADA algorithm so μ^T is the matching produced by SEADA. Second, we denote by $DA(P)$, the DA matching when students report preferences P with $DA_i(P)$ being student i's matching.

Consider some arbitrary claim (i, c) at μ^T, and let Γ denote the reassignment chain initiated by this claim.[16] We will show that student i must be rejected from c at some point in Γ, and hence the claim (i, c) is vacuous, and μ^T is essentially stable.[17]

We start with the following monotonicity lemma, part (i) of which is due to Kojima and Manea (2010). To state it, say that a preference relation P'_i is a **monotonic transformation** of P_i at $c \in C \cup \{\emptyset\}$ if $bP'_ic \implies bP_ic$. Preference profile P' is a monotonic transformation of P at a matching μ if P'_i is a monotonic transformation of P_i at $\mu(i)$ for all i.

[16]If there are no claims, then the matching is stable, and so is also essentially stable trivially. Also, μ^T is nonwasteful, and so any claim (i, c) must be because there exists some $j \in \mu^T(c)$ such that $i \succ_c j$.

[17]In an earlier version of this paper, we also prove that every round t matching μ^t is essentially stable. For simplicity, we focus on the most important one, μ^T, here.

Lemma 4. *(i) If P' is a monotonic transformation of P at $DA(P)$, then $DA_i(P')R'_iDA_i(P)$ for all students $i \in S$.*

(ii) If P' is a monotonic transformation of P at $DA(P)$, then $DA_i(P')R_iDA_i(P)$ for all students $i \in S$.

Now, consider again the claim (i, c) at μ^T. Because DA on the round T submarket is stable, only students who were removed in a round strictly earlier than T (the final round) can have a claim. That is, i must have been removed in some round $\hat{t} < T$. Define an alternative preference profile $P^{\hat{t}}$ as follows: for any student j removed before round $\hat{t}$, $P^{\hat{t}}_j$ ranks her assignment $\mu^{\hat{t}}(j)$ first, and the remaining schools in the same order as the true P_j; for all j not removed before round $\hat{t}$, $P^{\hat{t}}_j = P_j$. Note that this is a simple way to describe preferences so that $DA(P^{\hat{t}}) = \mu^{\hat{t}}$.[18]

Define a second preference profile $\bar{P}$ as follows: for each $j \neq i$, $\bar{P}_j$ ranks $\mu^T(j)$ first, and every other school is listed in the same order as the true P_j, while for student i, $\bar{P}_i$ ranks c first and the remaining schools in the order of the true P_i.

Lemma 5. $DA_i(\bar{P}) = DA_i(P^{\hat{t}}) = \mu^T(i)$.

The lemma is formally proved in the "Proofs of lemmas" subsection that follows the proof of this theorem, but the main step is that $\bar{P}$ is a monotonic transformation of $P^{\hat{t}}$ at $DA(P^{\hat{t}})$. Now, it is well-known that the following is an alternative description of the DA mechanism (McVitie and Wilson, 1971; Dubins and Freedman, 1981):

DA At each step t, arbitrarily choose one student among those who are currently unmatched, and allow her to apply to her most preferred a school that has not yet rejected him. All schools other than a tentatively hold the same students as the last step. School a holds the highest-priority students up to their capacity among those held from last step combined with the new applicant and reject the (at most one) other.

In this new method, the choice of the applicant at each step is arbitrary, in the sense that the order in which they are chosen does not affect the final outcome. So, for any fixed preference profile, one way to find the DA outcome is to have i be the last student chosen to enter the market. That is, as long as there is some other student besides i who is tentatively unmatched, we always choose one of these students to make the next application. Once all of

[18] Raising school $\mu^{\hat{t}}(j)$ for all j removed prior to round $\hat{t}$ to the top of her preferences is a way to effectively "remove" student j from the market, because no student who has not been removed prior to round $\hat{t}$ will ever apply to such a school because it is underdemanded.

these students have been (tentatively) assigned to a school, we allow i to enter by applying to the first school on her preference list. Student i's application then initiates a rejection chain, where i applies to some school a, a rejects its lowest-priority student i^1, i^1 applies to her most preferred school that has not yet rejected her, and so on, until we reach a school (which could be $\emptyset$) with an empty seat, at which point the rejection chain (and the entire DA mechanism) end, and all tentative matchings are made final.

Run $DA(\bar{P})$ with the alternative method by letting each student $j \neq i$ make applications in any arbitrary order. By construction of $\bar{P}$, each j applies to $\mu^T(j)$ and is tentatively matched to $\mu^T(j)$. No rejections occur because each $j \neq i$ is assigned to the unique seat to which she is assigned at μ^T. Now, again by construction of $\bar{P}$, when i enters, she begins by applying to c. We can index the rest of the steps of DA as a chain of rejections, which we denote Ξ, where

$$\text{Step}\,\Xi(l): \qquad \text{“student } i^\ell \text{ applies to school } a^\ell \text{ which rejects student } i^{\ell+1}\text{”}.$$

This chain of rejections eventually terminates at some L when a student applies to a school with a vacant seat (perhaps $\emptyset$). When a student $i^{\ell+1}$ is rejected, she goes to the next school on her list and applies. It may be the case that when a student applies to a school, she is rejected immediately, and must continue down her list. Formally, if $i^\ell \neq i^{\ell+1}$ we say step $\Xi(\ell)$ is **effective**. If a step is ineffective ($i^\ell = i^{\ell+1}$), then the same student who applied is also the one rejected, and nothing would change if i^ℓ simply skipped her application to a^ℓ. Let Ξ' be an alternative rejection chain that deletes all of the ineffective steps of Ξ. Deleting ineffective steps has no effect on the final outcome, and so the final matching at the end of Ξ and Ξ' is the same, and by construction, is $DA(\bar{P})$.

The key now is that the (initial) steps of Ξ' are the same as the steps of the reassignment chain Γ. Recall from above that all students $j \neq i$ are tentatively matched to the same school when i enters under $DA(\bar{P})$ as they are matched to when Γ begins (namely, school $\mu^T(j)$). Consider step 1. In the former case, a student j is rejected from her initial school $c = \mu^T(j)$. The rest of her preference list $\bar{P}_j$ coincides with her true preferences P_j so she goes down her true list P_j until she reaches a school where she has higher priority than some tentatively matched student. This is the same as step 1 of the reassignment chain Γ. We now have a new tentative matching for DA that is the same as the $\ell = 1$ matching for Γ, and the same student i^1 who is tentatively unassigned and will make the next application. Using the same argument, the second step of Ξ' leads to the same tentative matching as the $\ell = 2$

matching. By Lemma 5, $DA_i(\bar{P}) = \mu^T(i)$ so some step of Ξ' corresponds to c rejecting i. Given our argument that the steps are the same as in Γ, this corresponds to i being removed from c in Γ, and thus Γ ends at this step and the claim is vacuous.[19] $\square$

Proofs of lemmas

Proof of Lemma 1.

As we argued in the main text, schools receive higher-priority students throughout a reassignment chain, which directly implies the second part of the statement. We therefore focus on the first part of the statement, that is we show that students who are affected by Γ are worse off after Γ is carried out.

Let $j \in S$ be a student who is affected by Γ (i.e., $\mu^\Gamma(j) \neq \mu(j)$). For ease of notation, let $d = \mu^\Gamma(j)$. In addition, for any matching $\nu \in \mathcal{M}$ and any school $e \in C$ such that $|\nu(e)| = q_e$, we denote by $\underline{\nu}(e)$ the lowest-priority student in $\nu(e)$.

As $j \in \mu^\Gamma(d) \setminus \mu(d)$, d is affected by Γ. Then, d is filled to capacity throughout Γ as, otherwise, Γ ends whenever a student moves to d, which contradicts the assumption that the claim (i, c) is vacuous at μ. In particular, $|\mu(d)| = |\mu^\Gamma(d)| = q_d$ and $j \in \mu^\Gamma(d)$ implies that $j \succeq_d \underline{\mu}^\Gamma(d)$. Moreover, as d is affected by Γ and schools receive higher-priority students throughout a reassignment chain, we have $\underline{\mu}^\Gamma(d) \succ_d \underline{\mu}(d)$. Combining our last two statements yields $j \succ_d \underline{\mu}(d)$.

Towards a contradiction, suppose that $dP_j\mu(j)$, that is j is better off after Γ is carried out. Then, $j \succ_d \underline{\mu}(d)$ implies that j has a claim to d at μ. That claim is vacuous since μ is essentially stable. Let

$$j = j^0 \to d = d^0 \to j^1 \to d^1 \cdots \to j^{L-1} \to d = d^{L-1} \to j = j^L$$

be the reassignment chain initiated by (j, d) at μ, which we denote by Δ. Because (j, d) is a vacuous claim, Δ ends at some step L, where j is removed from d. Analogously to Definition 1, for every $\ell = 0, 1, \ldots, L$, let ν^ℓ be the matching obtained at step ℓ of Δ. For

[19] Technically, the reassignment chain Γ goes back to the top of P_j every time j needs an assignment while the rejection chain goes to the next school in $\bar{P}_j$, but they are equivalent here. This is because, as the reassignment chain progresses, the lowest priority of all the students matched to any school only increases, and so, even though j keeps going back to the top of the list in the reassignment chain, once j has been rejected from a school, she will continue to be rejected, and it is equivalent for her to just start with the next school down the list. Since all schools other than the top school under $\bar{P}_j$ are in the same order as P_j, the next (effective) school that j applies to will be equivalent under both scenarios.

every school $e \in C$, we define $\phi^\Gamma(e)$ as follows. If $|\mu^\Gamma(e)| < q_e$, then $\phi^\Gamma(e) = \mu^\Gamma(e)$. If $|\mu^\Gamma(e)| = q_e$, then $\phi^\Gamma(e) = \mu^\Gamma(e) \cup \{h \in S : \underline{\mu}^\Gamma(e) \succ_e h\}$. In words, $\phi^\Gamma(e)$ contains all those students who are either matched to e at μ^Γ or such that e is filled to capacity with higher-priority students.

Our argument proceeds by induction to prove the following statement:

For all $\ell = 1, \ldots, L$ and for every school $e \in C$, $\nu^\ell(e) \subseteq \phi^\Gamma(e)$.

We begin by showing that our statement is satisfied for $\ell = 1$. Consider any school $e \in C$ and recall that, by definition, $\nu^0 = \mu$. If $\mu(e) \subseteq \mu^\Gamma(e)$, it follows immediately that $\nu^0(e) = \mu(e) \subseteq \phi^\Gamma(e)$. Otherwise, all of the students in $\mu(e) \setminus \mu^\Gamma(e)$ are removed from e along Γ. Because schools receive higher-priority students throughout a reassignment chain, it follows that $|\mu^\Gamma(e)| = q_e$ and all of the students in $\mu(e) \setminus \mu^\Gamma(e)$ have a lower priority than $\underline{\mu}^\Gamma(e)$ at e. We conclude that $\nu^0(e) = \mu(e) \subseteq \phi^\Gamma(e)$. For all $e \in C \setminus \{d, \mu(j)\}$, $\nu^0(e) = \nu^1(e)$ while $\nu^1(\mu(j)) = (\nu^0(\mu(j))\setminus\{j\}) \subseteq \phi^\Gamma(\mu(j))$; therefore it remains to show that $\nu^1(d) \subseteq \phi^\Gamma(d)$. By construction, $\nu^1(d) \subset (\nu^0(d) \cup \{j\})$. As $j \in \mu^\Gamma(d) \subseteq \phi^\Gamma(d)$, $\nu^1(d) \subseteq \phi^\Gamma(d)$, as required.

We next suppose that our statement is satisfied for some $\ell = 1, \ldots, L-1$ (induction hypothesis) and show that it is then also satisfied for $\ell + 1$. For every $e \in C \setminus \{d^\ell\}$, $\nu^\ell(e) = \nu^{\ell+1}(e)$; therefore the induction hypothesis directly implies that $\nu^{\ell+1}(e) \subseteq \phi^\Gamma(e)$. Moreover, by construction, $\nu^{\ell+1}(d^\ell) \subset (\nu^\ell(d^\ell) \cup \{j^\ell\})$; therefore it remains to show that $j^\ell \in \phi^\Gamma(d^\ell)$. (Note that $d^\ell \neq \emptyset$ as, otherwise, Δ would end after $\ell < L$ steps.)

We first show that, at ν^ℓ, $\mu^\Gamma(j^\ell)$ is not filled to capacity with students who all have a higher priority than j^ℓ. This is trivially the case if $|\nu^\ell(\mu^\Gamma(j^\ell))| < q_{\mu^\Gamma(j^\ell)}$. If $|\nu^\ell(\mu^\Gamma(j^\ell))| = q_{\mu^\Gamma(j^\ell)}$, we need to show that $j^\ell \succeq_{\mu^\Gamma(j^\ell)} \underline{\nu}^\ell(\mu^\Gamma(j^\ell))$. By the induction hypothesis, $\nu^\ell(\mu^\Gamma(j^\ell)) \subseteq \phi^\Gamma(\mu^\Gamma(j^\ell))$; therefore $\underline{\nu}(\mu^\Gamma(j^\ell))$ has at best the $q^{th}_{\mu^\Gamma(j^\ell)}$ highest priority among students in $\phi^\Gamma(\mu^\Gamma(j^\ell))$. By construction, any student in $\mu^\Gamma(\mu^\Gamma(j^\ell))$ (including j^ℓ) has at least the $q^{th}_{\mu^\Gamma(j^\ell)}$ highest priority among students in $\phi^\Gamma(\mu^\Gamma(j^\ell))$. Therefore, $j^\ell \succeq_{\mu^\Gamma(j^\ell)} \underline{\nu}^\ell(\mu^\Gamma(j^\ell))$.

We now conclude our inductive argument by showing that $j^\ell \in \phi^\Gamma(d^\ell)$. By definition, our previous result that $\mu^\Gamma(j^\ell)$ is not filled to capacity with students who all have a higher priority than j^ℓ implies that $d^\ell R_{j^\ell} \mu^\Gamma(j^\ell)$; otherwise j^ℓ would be matched to $\mu^\Gamma(j^\ell)$ (or a more preferred school) rather than d^ℓ. If $d^\ell = \mu^\Gamma(j^\ell)$, then $j \in \mu^\Gamma(d^\ell) \subseteq \phi^\Gamma(d^\ell)$ and the inductive argument is complete. We devote the remainder of our argument to the case where $d^\ell P_{j^\ell} \mu^\Gamma(j^\ell)$. By construction, j^ℓ has been removed from $\mu(j^\ell)$ before step ℓ of Δ; hence $|\nu^\ell(\mu(j^\ell))| = q_{\mu(j^\ell)}$ and $\underline{\nu}^\ell(\mu(j^\ell)) \succ_{\mu(j^\ell)} j^\ell$. As $\emptyset$ is never filled to capacity, it directly follows that $\mu(j^\ell) \neq \emptyset$; therefore our induction hypothesis applies and yields $\nu^\ell(\mu(j^\ell)) \subseteq \phi^\Gamma(\mu(j^\ell))$.

Then, by construction, $j^\ell \in \mu^\Gamma(\mu(j^\ell))$ would imply $j^\ell \succeq_{\mu(j^\ell)} \underline{\nu}^\ell(\mu(j^\ell))$, a contradiction. It follows that $j^\ell \notin \mu^\Gamma(\mu(j^\ell))$, or equivalently $\mu(j^\ell) \neq \mu^\Gamma(j^\ell)$. Consequently, j^ℓ is matched to $\mu^\Gamma(j^\ell)$ at some step of Γ. At that point, d^ℓ is filled to capacity with students who all have a higher priority than j^ℓ; otherwise our assumption that $d^\ell P_{j^\ell} \mu^\Gamma(j^\ell)$ would imply that j^ℓ is matched to d^ℓ (or a more preferred school) rather than $\mu^\Gamma(j^\ell)$. As schools get higher-priority students throughout a reassignment chain, it follows that that $|\mu^\Gamma(d^\ell)| = q_{d^\ell}$ and $\underline{\mu}^\Gamma(d^\ell) \succ_{d^\ell} j^\ell$. Then, by definition, we have $j^\ell \in \phi^\Gamma(d^\ell)$, which concludes our inductive argument.

On the one hand, j is removed from d at step L of Δ, which means that $|\nu^L(d)| = q_d$ and $\underline{\nu}^L(d) \succ_d j$. On the other hand, we have established through our inductive argument that $\nu^L(d) \subseteq \phi^\Gamma(d)$. Then, by construction, $j \in \mu^\Gamma(d)$ implies $j \succeq_d \underline{\nu}^L(d)$, a contradiction. □

Proof of Lemma 2.

Suppose that j has a claim to d at μ^Γ; we need to show that j has a claim to d at μ. By assumption, $dP_j\mu^\Gamma(j)$ and $\mu^\Gamma(d)$ contains at most $q_d - 1$ students who have a higher priority than j at d.

Suppose first, towards a contradiction, that j is affected by Γ, i.e., $\mu^\Gamma(j) \neq \mu(j)$. In that case, j is matched to $\mu^\Gamma(j)$ somewhere along Γ, i.e., $i^k = j$ and $c^k = \mu^\Gamma(j)$ for some $k = 1, \ldots, K-1$. By definition, $\mu^\Gamma(j)$ is j's most preferred school to which she has a claim at μ^k; therefore, as $dP_j\mu^\Gamma(j)$, j does not have a claim to d at μ^k. Then, $\mu^k(d)$ contains q_d students, all of whom have a higher priority than j at d. As schools receive higher-priority students throughout a reassignment chain, the same is true of $\mu^\Gamma(d)$, a contradiction.

We have established that $\mu^\Gamma(j) = \mu(j)$, which directly implies $dP_j\mu(j)$. As $\mu^\Gamma(d)$ contains at most $q_d - 1$ students who have a higher priority than j at d and schools receive higher-priority students throughout a reassignment chain, $\mu(d)$ contains at most $q_d - 1$ students who have a higher priority than j at d. Combining our last two findings implies that j has a claim to d at μ. □

Proof of Lemma 3.

Consider a claim (j, d) at μ^Γ; we need to show that this claim is vacuous. By Lemma 2, (j, d) is a claim at μ and, as μ is essentially stable, it is vacuous. The remainder of the proof follows a similar inductive argument to the one presented in the proof of Lemma 1. Let

$$j = j^0 \to d = d^0 \to j^1 \to d^1 \to \cdots \to j^{L-1} \to d = d^{L-1} \to j = j^L$$

be the reassignment chain initiated by (j, d) at μ, which we denote by Δ. Because (j, d) is a vacuous claim at μ, Δ ends at some step L, where j is removed from d. Analogously to Definition 1, for every $\ell = 0, 1, \ldots, L$, let ν^ℓ be the matching obtained at step ℓ of Δ. We denote by Δ^* the reassignment chain initiated by (j, d) at μ^Γ and by $\mu^* = (\mu^\Gamma)^{\Delta^*}$, the final matching obtained after Δ^* is carried out. For any matching $\nu \in \mathcal{M}$ and any school $e \in C$ such that $|\nu(e)| = q_e$, we denote by $\underline{\nu}(e)$ the lowest-priority student in $\nu(e)$. For every school $e \in C$, we define $\phi^*(e)$ as follows. If $|\mu^*(e)| < q_e$, then $\phi^*(e) = \mu^*(e)$. If $|\mu^*(e)| = q_e$, then $\phi^*(e) = \mu^*(e) \cup \{h \in S : \underline{\mu}^*(e) \succ_e h\}$. In words, $\phi^*(e)$ contains all those students who are either matched to e at μ^* or such that e is filled to capacity with higher-priority students. Our argument proceeds by induction to prove the following statement:

For all $\ell = 1, \ldots, L$ and for every school $e \in C$, $\nu^\ell(e) \subseteq \phi^*(e)$.

We begin by showing that our statement is satisfied for $\ell = 1$. Consider any school $e \in C$ and recall that, by definition, $\nu^0 = \mu$. If $\mu(e) \subseteq \mu^*(e)$, it follows immediately that $\nu^0(e) = \mu(e) \subseteq \phi^*(e)$. Otherwise, all of the students in $\mu(e) \setminus \mu^*(e)$ are removed from e along either Γ or Δ^*. Because schools receive higher-priority students throughout a reassignment chain, it follows that $|\mu^*(e)| = q_e$ and all of the students in $\mu(e) \setminus \mu^*(e)$ have a lower priority than $\underline{\mu}^*(e)$ at e. We conclude that $\nu^0(e) = \mu(e) \subseteq \phi^*(e)$. For all $e \in C \setminus \{d, \mu(j)\}$, $\nu^0(e) = \nu^1(e)$ while $\nu^1(\mu(j)) = (\nu^0(\mu(j)) \setminus \{j\}) \subseteq \phi^*(\mu(j))$; therefore it remains to show that $\nu^1(d) \subseteq \phi^*(d)$. By construction, $\nu^1(d) \subset (\nu^0(d) \cup \{j\})$ and j is matched to d at the beginning of Δ^*; therefore either $j \in \mu^*(d)$ or j is removed from d somewhere along Δ^*, in which case $|\mu^*(d)| = q_d$ and $\underline{\mu}^*(d) \succ_d j$ since schools receive higher-priority students throughout a reassignment chain. We conclude that $j \in \phi^*(d)$, which implies that $\nu^1(d) \subseteq \phi^*(d)$, as required.

We next suppose that our statement is satisfied for some $\ell = 1, \ldots, L-1$ (induction hypothesis) and show that it is then also satisfied for $\ell + 1$. For every $e \in C \setminus \{d^\ell\}$, $\nu^\ell(e) = \nu^{\ell+1}(e)$; therefore the induction hypothesis directly implies that $\nu^{\ell+1}(e) \subseteq \phi^*(e)$. By construction, $\nu^{\ell+1}(d^\ell) \subset (\nu^\ell(d^\ell) \cup \{j^\ell\})$; therefore it remains to show that $j^\ell \in \phi^*(d^\ell)$. (Note that $d^\ell \neq \emptyset$ as, otherwise, Δ would end after $\ell < L$ steps.)

We first show that, at ν^ℓ, $\mu^*(j^\ell)$ is not filled to capacity with students who all have a higher priority than j^ℓ. This is trivially the case if $|\nu^\ell(\mu^*(j^\ell))| < q_{\mu^*(j^\ell)}$. If $|\nu^\ell(\mu^*(j^\ell))| = q_{\mu^*(j^\ell)}$, we need to show that $j^\ell \succeq_{\mu^*(j^\ell)} \underline{\nu}^\ell(\mu^*(j^\ell))$. By the induction hypothesis, $\nu^\ell(\mu^*(j^\ell)) \subseteq \phi^*(\mu^*(j^\ell))$; therefore $\underline{\nu}(\mu^*(j^\ell))$ has at best the $q^{th}_{\mu^*(j^\ell)}$ highest priority among students in $\phi^*(\mu^*(j^\ell))$. By construction, any student in $\mu^*(\mu^*(j^\ell))$ (including j^ℓ) has at least the $q^{th}_{\mu^*(j^\ell)}$ highest priority among students in $\phi^*(\mu^*(j^\ell))$. Therefore, $j^\ell \succeq_{\mu^*(j^\ell)} \underline{\nu}^\ell(\mu^*(j^\ell))$.

We now conclude our inductive argument by showing that $j^\ell \in \phi^*(d^\ell)$. By definition, our previous result that $\mu^*(j^\ell)$ is not filled to capacity with students who all have a higher priority than j^ℓ implies that $d^\ell R_{j^\ell} \mu^*(j^\ell)$; otherwise j^ℓ would be matched to $\mu^*(j^\ell)$ (or a more preferred school) rather than d^ℓ. If $d^\ell = \mu^*(j^\ell)$, then $j \in \mu^*(d^\ell) \subseteq \phi^*(d^\ell)$ and the inductive argument is complete. We devote the remainder of our argument to the case where $d^\ell P_{j^\ell} \mu^*(j^\ell)$. By construction, j^ℓ has been removed from $\mu(j^\ell)$ before step ℓ of Δ; hence $|\nu^\ell(\mu(j^\ell))| = q_{\mu(j^\ell)}$ and $\underline{\nu}^\ell(\mu(j^\ell)) \succ_{\mu(j^\ell)} j^\ell$. As $\emptyset$ is never filled to capacity, it directly follows that $\mu(j^\ell) \neq \emptyset$; therefore our induction hypothesis applies and yields $\nu^\ell(\mu(j^\ell)) \subseteq \phi^*(\mu(j^\ell))$. Then, by construction, $j^\ell \in \mu^*(\mu(j^\ell))$ would imply $j^\ell \succeq_{\mu(j^\ell)} \underline{\nu}^\ell(\mu(j^\ell))$, a contradiction. It follows that $j^\ell \notin \mu^*(\mu(j^\ell))$, or equivalently $\mu(j^\ell) \neq \mu^*(j^\ell)$. Consequently, j^ℓ is matched to $\mu^*(j^\ell)$ at some step of either Γ or Δ^*. At that point, d^ℓ is filled to capacity with students who all have a higher priority than j^ℓ; otherwise our assumption that $d^\ell P_{j^\ell} \mu^*(j^\ell)$ would imply that j^ℓ is matched to d^ℓ (or a more preferred school) rather than $\mu^*(j^\ell)$. As schools get higher-priority students throughout a reassignment chain, it follows that that $|\mu^*(d^\ell)| = q_{d^\ell}$ and $\underline{\mu}^*(d^\ell) \succ_{d^\ell} j^\ell$. Then, by definition, we have $j^\ell \in \phi^*(d^\ell)$, which concludes our inductive argument.

On the one hand, j is removed from d at step L of Δ, which means that $|\nu^L(d)| = q_d$ and $\underline{\nu}^L(d) \succ_d j$. On the other hand, we have established through our inductive argument that $\nu^L(d) \subseteq \phi^*(d)$. If j's claim to d at μ^Γ is not vacuous, then $j \in \mu^*(d)$. As $|\nu^L(d)| = q_d$ and $\nu^L(d) \subseteq \phi^*(d)$, it follows by construction that $j \succeq_d \underline{\nu}^L(d)$, a contradiction. □

Proof of Lemma 4.

Part (i) is shown in Kojima and Manea (2010), and they refer to this property as *weak Maskin monotonicity*. For part (ii), consider a student i, and let $DA_i(P) = a$ and $DA_i(P') = a'$. If $a = a'$, then it is immediate. Otherwise, by part (i), we have $a' P'_i a$. Since P'_i is a monotonic transformation of P_i at a, $a' P'_i a$ implies $a' P_i a$. □

Proof of Lemma 5.

We start by showing that $\bar{P}$ is a monotonic transformation of $P^{\hat{t}}$ at $DA(P^{\hat{t}})$. For each $j \in S$, let $DA_j(P^{\hat{t}}) = a_j$. For all j removed from the market at some round $t < \hat{t}$, $\mu^t(j) = \mu^{\hat{t}}(j) = a_j$. Thus, both $P^{\hat{t}}_j$ and $\bar{P}_j$ rank school a_j first, and $\bar{P}_j$ is trivially a monotonic transformation of $P^{\hat{t}}_j$ at a_j for these students.

Next, consider the students who are still in the market at the beginning of round $\hat{t}$, and note that for all such students, $P_j^{\hat{t}} = P_j$. Consider some such $j \neq i$. By Lemma 2 of Tang and Yu (2014), $\mu^T(j) R_j a_j$ for all j. Since $P_j^{\hat{t}} = P_j$, this further implies that $\mu^T(j) R_j^{\hat{t}} a_j$. Now, consider preference profile $\bar{P}_j$. $\bar{P}_j$ simply raises $\mu^T(j)$ to the top of the ordering, without altering the relative rankings of any other seats (in particular, no schools "jump" over student j's round $\hat{t}$ assignment a_j in the move from $P_j^{\hat{t}}$ to $\bar{P}_j$), and so $\bar{P}_j$ is a monotonic transformation of $P_j^{\hat{t}}$ at a_j for all $j \neq i$.

Last, consider student i. She is removed in round $\hat{t}$, and so $c P_i^{\hat{t}} a_i$ (otherwise, student i would not claim a seat at c at μ^T).[20] By similar logic (no school a' "jumps" over a_i in going from $P_i^{\hat{t}}$ to $\bar{P}_i$), $\bar{P}_i$ is a monotonic transformation of $P_i^{\hat{t}}$ at a_i. Thus, we have shown that $\bar{P}_j$ is a monotonic transformation of $P_j^{\hat{t}}$ at $DA_j(P^{\hat{t}})$ for all $j \in S$, and so preference profile $\bar{P}$ is a monotonic transformation of preference profile $P^{\hat{t}}$ at $DA(P^{\hat{t}})$.

Finally, given a matching μ, say student j is **not Pareto improvable** if, for every ν that Pareto dominates μ, $\nu(j) = \mu(j)$. Since $\bar{P}$ is a monotonic transformation of $P^{\hat{t}}$ at $DA(P^{\hat{t}})$, Lemma 4, part (ii) gives $DA_j(\bar{P}) R_j^{\hat{t}} DA_j(P^{\hat{t}})$ for all $j \in S$, i.e., the matching $DA(\bar{P})$ Pareto dominates the matching $DA(P^{\hat{t}})$ with respect to $P^{\hat{t}}$. Since i is removed in round $\hat{t}$, she must be matched with an underdemanded school at $DA(P^{\hat{t}})$ which, by Lemma 1 of Tang and Yu (2014), implies that she is not Pareto improvable (relative to preferences $P^{\hat{t}}$). Since $DA(\bar{P})$ Pareto dominates $DA(P^{\hat{t}})$ and i is not Pareto improvable, her matching does not change: $DA_i(\bar{P}) = DA_i(P^{\hat{t}})$. Since i is removed at round $\hat{t}$, her assignment at $T > \hat{t}$ is the same as her assignment at the end of round $\hat{t}$: $\mu^T(i) = DA_i(P^{\hat{t}})$. $\square$

C Comparison to other definitions in the literature

In this appendix, first we show formally that our definition of essential stability is distinct from other approaches to weakening stability that have been proposed in the literature by finding matchings that satisfy each of the other definitions but are strongly unstable under our definition. Both Alcalde and Romero-Medina (2015) and Cantala and Pápai (2014) show that the DA+TTC mechanism satisfies their respective definitions of stability, while we showed in Section 3 that DA+TTC is not essentially stable. Therefore, the matching μ^*

[20] Because i is removed in round $\hat{t}$, we have $\mu^T(i) = \mu^{\hat{t}}(i) = a_i$; because she claims a seat at c at μ^T, we have $c P_i \mu^T(i)$; again because i is still in the market at round $\hat{t}$, we have $P_i^{\hat{t}} = P_i$. This all implies that $c P_i^{\hat{t}} a_i$.

from Example 1 is τ-fair, reasonably stable, and securely stable according to their respective definitions, but is strongly unstable according to the definition used in this paper.

The definitions of Tang and Zhang (2016) and Ehlers and Morrill (2019) are satisfied by the EADA mechanism and so it is less obvious that they are formally distinct. However, as we show here, they are not equivalent.[21] We first consider Ehlers and Morrill (2019), who define the concept of a **legal** set of assignments. In contrast to stability, which is defined on a matching itself, legality is defined on a set of matchings; i.e., an individual matching μ cannot be deemed "legal" or "illegal" independently, but is only legal in relation to other matchings. Also, note that the model of Ehlers and Morrill (2019) allows for more general school "choice functions"; the definitions presented here are simplified to apply to our model.[22]

A matching μ **blocks** a matching ν if there exists some i such that $\mu(i) = aP_i\nu(i)$ and $i \succ_a j$ for some $j \in \nu(a)$. Then, a set of matchings L is a **legal set** if

1. For all $\mu, \nu \in L$, μ is not blocked by ν
2. For all $\mu \notin L$, μ is blocked by some $\nu \in L$.

Example 1 can be used to show that essential stability is different from legality. More precisely, we exhibit a matching μ that must be included in any legal set of matchings L, but is not essentially stable. To shorten notation, we refer to a matching by a string of letters representing the school assigned to each student in order of their indices. For example, $\mu = ABCDE$ means that i_1 is assigned to A, i_2 to B, i_3 to C, and so forth.

Consider the matching $\mu = BACDE$. We claim that $\mu \in L$ for any legal set L, but μ is not essentially stable. Showing μ is not essentially stable is simple. Note that i_3 claims the seat at school B, and the reassignment chain that follows is $(i_3 \to B \to i_1 \to A \to i_2 \to C)$. Since i_3 remains matched to B at the end of the chain, the claim (i_3, B) is non-vacuous and so μ is not essentially stable.

Let L be a legal set of matchings. First, note that the DA outcome is $\mu^{DA} = ABCDE$, and $\mu^{DA} \in L$ for any L (because it is not blocked by anything). Next, observe that each student i has the highest priority at her DA school. So, i can use the DA matching to block

[21]In an earlier draft of this paper, we used the arguments below to show that essential stability is different from a related definition of Morrill (2016). Ehlers and Morrill (2019) supersedes Morrill (2016), though the same arugment works for both papers.

[22]Further, in defining legality, Ehlers and Morrill (2019) restrict attention only to individually rational assignments. In the example we construct below to show the distinction between the two concepts, all assignments will be individiually rational, and so to avoid introducing unnecessary notation we omit this requirement from the formal definition.

any other matching ν that gives her a school she disprefers to her DA school. This implies that for all $\nu \in L$, ν Pareto dominates μ^{DA}.

We now show that $\mu = BACDE \in L$. Assume not, i.e., $\mu \notin L$. By part (2) of the definition of legality, there exists some $\nu \in L$ that blocks it. The only potential student who can block μ is i_3, who can block with B. Let ν be some $\nu \in L$ at which $\nu(i_3) = B$. Since ν must Pareto dominate μ^{DA}, there is only one possibility: $\nu = ACBDE$.[23] Thus, $\nu = ACBDE \in L$.

Since $\nu \in L$, there is no $\rho \in L$ that blocks it. Since ν can be blocked by any matching ρ such that $\rho(i_4) = C$, we have $\rho(i_4) = C$ implies that $\rho \notin L$; in particular, $\rho = ABDCE \notin L$.

Since $\rho \notin L$, there must be some $\sigma \in L$ that blocks ρ. The only student who can block ρ is i_5, who can block with any σ such that $\sigma(i_5) = D$. However, any such σ has some student who is assigned to a school worse than her DA assignment,[24] which contradicts that every $\sigma \in L$ Pareto dominates μ^{DA}.

The above shows that essential stability is different from legality of Ehlers and Morrill (2019). In fact, it also shows that essential stability is different from weak stability of Tang and Zhang (2016) as well. This follows because legal sets are vNM stable sets (Ehlers and Morrill, 2019), and Tang and Zhang (2016) show that every matching that is in the vNM stable set is weakly stable in their sense.

Finally, we show that essential stability is independent of the no-consent-proofness property of Dur et al. (2015) by constructing two mechanisms, each of which satisfies exactly one of the two properties. First, the following mechanism is no-consent-proof, but not essentially stable: Ask everyone if they are willing to consent to having all of their priorities violated. If everyone consents, then run the DA+TTC mechanism. If anyone does not, run the standard DA mechanism. Since DA+TTC Pareto dominates DA, this mechanism is no-consent-proof; however, as we show in the paper, DA+TTC may produce assignments with non-vacuous claims, and so this mechanism is strongly unstable.

To construct a mechanism that is essentially stable but not no-consent-proof, we use Example 2 (this can easily be embedded in larger markets). Again, first ask everyone if they are willing to consent to having all of their priorities violated. If everyone consents and the reported preference profile is the one from the example, output the matching $\mu^{\dagger}$ from the

[23]Since ν must Pareto dominate μ^{DA}, i_1 must get A (because i_3 is assigned B). Then, since A and B are taken, $\nu(i_2) = C$, which further implies that $\nu(i_4) = D$. The only school left is E, and so $\nu(i_5) = E$.

[24]For each student i_1, i_2, i_3, and i_4, the schools weakly preferred to her DA assignment are some subset of $\{A, B, C\}$. Since there are only 3 seats at these schools and 4 students, some student must be assigned to a school worse than her DA assignment.

example; otherwise, output the DA matching for the reported preferences. This mechanism is essentially stable, as any DA matching is essentially stable, and so is $\mu^\dagger$. However, it is not no-consent-proof: when the preferences are those from the example, i_1 (or i_3 or i_4) are better off not consenting, since by doing so they get their DA matching, which they prefer.

D Semilattice

As mentioned in Section 4, the existence of a student-pessimal essentially stable matching may intuitively suggest that the set of essentially stable matchings forms a semilattice. We show in this appendix that this is in fact not the case.

For any two matchings $\mu, \nu \in \mathcal{M}$, we write $\mu R \nu$ if μ weakly Pareto dominates ν, i.e., if $\mu(i) R_i \nu(i)$ for all $i \in S$. The set of essentially stable matchings forms a **meet-semilattice** with respect to the partial order R if for any two essentially stable matchings μ_1 and μ_2, there exists a *greatest lower bound* (also called *infimum* or *meet*) $\overline{\mu}$ such that (i) $\overline{\mu}$ is an essentially stable matching, (ii) $\mu_1 R \overline{\mu}$ and $\mu_2 R \overline{\mu}$, and (iii) for any essentially stable matching μ: $\mu_1 R \mu$ and $\mu_2 R \mu$ imply $\overline{\mu} R \mu$.

Proposition 6. *The set of essentially stable matchings may not form a meet-semilattice with respect to the partial order R.*

Proof. The proof is by counterexample, which we present below.

Example 3. Let there be 7 students, $S = \{i_1, i_2, i_3, i_4, i_5, i_6, i_7\}$, and 7 schools with capacity 1, $C = \{A, B, C, D, E, F, G\}$. The priorities and preferences are given in the following tables.

$\succ_A$	$\succ_B$	$\succ_C$	$\succ_D$	$\succ_E$	$\succ_F$	$\succ_G$
i_7	i_1	i_2	i_6	i_4	i_5	i_3
i_5	i_2	i_3	i_3	i_5	i_6	i_7
i_3	$\vdots$	$\vdots$	i_4	$\vdots$	$\vdots$	$\vdots$
i_1			i_1			
$\vdots$			$\vdots$			

P_{i_1}	P_{i_2}	P_{i_3}	P_{i_4}	P_{i_5}	P_{i_6}	P_{i_7}
A	B	C	D	E	F	G
D	C	A	E	A	D	A
B	$\vdots$	D	$\vdots$	F	$\vdots$	$\vdots$
$\vdots$		G		$\vdots$		
		$\vdots$				

The following matchings are essentially stable:[25]

$$\mu_1 = \begin{pmatrix} A & B & C & D & E & F & G \\ i_3 & i_1 & i_2 & i_4 & i_5 & i_6 & i_7 \end{pmatrix} \qquad \mu_2 = \begin{pmatrix} A & B & C & D & E & F & G \\ i_5 & i_2 & i_3 & i_1 & i_4 & i_6 & i_7 \end{pmatrix}$$

$$\mu_3 = \begin{pmatrix} A & B & C & D & E & F & G \\ i_5 & i_1 & i_2 & i_3 & i_4 & i_6 & i_7 \end{pmatrix} \qquad \mu_4 = \begin{pmatrix} A & B & C & D & E & F & G \\ i_3 & i_1 & i_2 & i_6 & i_4 & i_5 & i_7 \end{pmatrix}$$

It is easy to verify that μ_1 and μ_3 are stable. At μ_2, the only claim is i_4's claim to D. The reassignment chain initiated by that claim is

$$i_4 \to D \to i_1 \to B \to i_2 \to C \to i_3 \to D \to i_4;$$

therefore the claim is vacuous and μ_2 is essentially stable. At μ_4, the only claim i_5's claim to A. The reassignment chain initiated by that claim is

$$i_5 \to A \to i_3 \to G \to i_7 \to A \to i_5;$$

therefore the claim is vacuous and μ_4 is essentially stable.

It is easy to verify that neither one of μ_1 and μ_2 Pareto dominates the other and that the same holds for μ_3 and μ_4; however, μ_1 and μ_2 both Pareto dominate μ_3 as well as μ_4. To conclude the proof, suppose towards a contradiction that the set of essentially stable matchings forms a meet-semilattice with respect to the partial order R. Then, μ_1 and μ_2 have a greatest lower bound $\overline{\mu}$. By definition, $\mu_1 R \overline{\mu}$, $\mu_2 R \overline{\mu}$, $\overline{\mu} R \mu_3$, and $\overline{\mu} R \mu_4$; therefore

$$A = \mu_1(i_3) R_{i_3} \overline{\mu}(i_3) R_{i_3} \mu_4(i_3) = A \quad \text{and} \quad A = \mu_2(i_5) R_{i_5} \overline{\mu}(i_5) R_{i_5} \mu_3(i_5) = A.$$

It follows that $\overline{\mu}(i_3) = \overline{\mu}(i_5) = A$, a contradiction since each school has capacity 1. □

E Alternative Definitions

In this appendix, we consider two alternative definitions of essential stability and show that both are equivalent to our original definition.

[25] There are also a student-optimal (essentially) stable matching that assigns each student to her favorite school and a student-pessimal (essentially) stable matching that assigns to each school the student with the top priority.

Long-chain Essential Stability

We have defined reassignment chains to end as soon as the initial claimant is removed from the school she claimed. One possible concern is that this student could still have a claim to a school she prefers to her original match. We propose an alternative definition such that reassignment chains only end when a student is matched to a school that has an free seat (or takes her outside option). We show that essential stability is not affected as a result.

Definition 3. Consider a matching μ and a claim (i, c). The **long reassignment chain** Γ **initiated by claim** (i, c) is the list

$$i^0 \to c^0 \to i^1 \to c^1 \to \cdots$$

where,

- $i^0 = i$, $\mu^0 = \mu$, $c^0 = c$ and, for each $k \geq 1$,
- $i^k \in S$ is the lowest-priority student in $\mu^{k-1}(c^{k-1})$ according to $\succ_{c^{k-1}}$,
- μ^k is defined as: $\mu^k(i^k) = \emptyset$, $\mu^k(i^{k-1}) = c^{k-1}$, and $\mu^k(j) = \mu^{k-1}(j)$ for all $j \in S \setminus \{i^{k-1}, i^k\}$,
- $c^k \in C$ is i^k's most preferred school to which she has a claim at μ^k if such a school exists; otherwise, $c^k = \emptyset$. If $|\mu^k(c^k)| < q_{c^k}$, the chain terminates.

There is one difference between Definitions 1 and 3. The reassignment chain (Definition 1) initiated by the claim (i, c) ends as soon c rejects i. In contrast, if c rejects i in the long reassignment chain (Definition 3) initiated by the claim (i, c), then the chain continues and i is matched to her most preferred school among those to which she has a claim. The long chain only ends when a student is matched to a school that is not filled to capacity, at which point i may still be matched to a school she prefers to $\mu(i)$. In that case, while i has been rejected by c, her claiming c has still proved valuable so one may dispute the vacuity of the claim (i, c). We say that i's claim to c at μ is **long-chain vacuous** if the *long* reassignment chain it initiates ends at some step K with i matched to $\mu(i)$, i.e., $i^K = i$ and $c^K = \mu(i)$. Matching μ is **long-chain essentially stable** if all claims at μ are long-chain vacuous.

Proposition 7. *A matching is long-chain essentially stable if and only if it is essentially stable.*

Robust-chain Essential Stability

Another possible concern with our definition is that some of the claims that are satified alongside a reassignment chain may be vacuous. If we argue that vacuous claims are not as serious as non-vacuous one, then perhaps they should be discarded when constructing reassignment chains. We show, however, that this does not affect our results.

Definition 4. Consider a matching μ and a claim (i, c). The **robust reassignment chain** Γ **initiated by claim** (i, c) is the list

$$i^0 \to c^0 \to i^1 \to c^1 \to \cdots$$

where,

- $i^0 = i$, $\mu^0 = \mu$, $c^0 = c$ and, for each $k \geq 1$,
- $i^k \in S$ is the lowest-priority student in $\mu^{k-1}(c^{k-1})$ according to $\succ_{c^{k-1}}$,
- μ^k is defined as: $\mu^k(i^k) = \emptyset$, $\mu^k(i^{k-1}) = c^{k-1}$, and $\mu^k(j) = \mu^{k-1}(j)$ for all $j \in S \setminus \{i^{k-1}, i^k\}$,
- if $i^k = i$, the chain terminates,
- if $i^k \neq i$, then $c^k \in C$ is i^k's most preferred school to which she has a *non-vacuous* claim at μ^k if such a school exists; otherwise, $c^k = \emptyset$. If $|\mu^k(c^k)| < q_{c^k}$, the chain terminates.

A robust reassignment chain is identical to a reassignment chain except for one difference: at any step k, student i^k is matched to the school she prefers among those to which she has a *non-vacuous* claim. We say that claim (i, c) at matching μ is **robust-chain vacuous** if $i^K = i$ in the reassignment chain initiated by (i, c) at μ that terminates at step K. Matching μ is **robust-chain essentially stable** if all claims at μ are robust-chain vacuous. As our next result shows, essential stability is unaffected by this alternative definition.

Proposition 8. *A matching is robust-chain essentially stable if and only if it is essentially stable.*

Proofs

Proof of Proposition 7.

Our definitions directly imply that all long-chain vacuous claims are vacuous; therefore, all long-chain essentially stable matchings are essentially stable. To show the converse, consider an essentially stable matching μ and a (vacuous) claim (i, c) at μ; we need to show that (i, c) is long-chain vacuous at μ.

Denote by Γ_1 the reassignment chain initiated by (i, c) at μ and by $\mu_1 = \mu^{\Gamma_1}$ the final matching obtained after, starting from μ, Γ_1 is carried out. Let $c_1, c_2, \ldots, c_N$ be a sequence of schools, $\Gamma_2, \ldots, \Gamma_N$ be a sequence of reassignment chains, and $\mu_2, \ldots, \mu_N$ be a sequence of matchings such that $c_1 = c$ and, for each $n = 2, \ldots, N$,

- c_n is i's most preferred school to which she has a claim at μ_{n-1},
- Γ_n is the reassignment chain initiated by (i, c_n) at μ_{n-1},
- $\mu_n = \mu_{n-1}^{\Gamma_n}$ is the final matching obtained after, starting from μ_{n-1}, Γ_n is carried out,
- and i does not have any claim at μ_N.

For every $n = 1, \ldots, N$, μ_n is essentially stable by Lemma 3, which means that Γ_n ends at some step K_n where i is removed from c_n. Lemmas 1 and 2 imply that i has striclty fewer claims at the end of each reassignment chain; therefore N is finite, i.e., a matching in which i does not have any claim is reached after a finite number of reassignment chains are caried out. For every $n = 1, \ldots, N$, it will prove useful to define $K_{(n)} = \sum_{m=1}^{n} K_m$ to be the sum of the steps of the first n reassignment chains.

Let

$$i = j^0 \to c = d^0 \to j^1 \to d^1 \to \cdots \to j^L \to d^L$$

be the *long* reassignment chain initiated by (i, c) at μ, which we denote by Δ. Analogously to Definition 3, for every $\ell = 0, 1, \ldots, L$, let ν^ℓ be the matching obtained at step ℓ of Δ. We need to show that $j^L = i$ and $d^L = \mu(i)$.

Our argument proceeds by induction to prove the following statement:

For all $n = 1, \ldots, N$, $j^{K_{(n)}} = i$ and $\nu^{K_{(n)}}(j) = \mu_n(j)$ for all $j \in S \setminus \{i\}$.

We begin by showing that our statement is satisfied for $n = 1$. By definition, Γ_1 and Δ are identical up to step $K_{(1)} = K_1$ where i is removed from c; therefore, $j^{K_{(1)}} = i$ and $\nu^{K_{(1)}}(j) = \mu_1(j)$ for all $j \in S \setminus \{i\}$.

We next suppose that our statement is satisfied for some $n = 1, \ldots, N-1$ (induction hypothesis) and show that it is then also satisfied for $n+1$. Observe first that $\mu(i) R_i \emptyset$, as otherwise, because $|\mu(\emptyset)| \leq |S| < q_\emptyset$, i has a non-vacuous claim to $\emptyset$ at μ, which contradicts the assumption that μ is essentially stable. By definition, c_{n+1} is i's most preferred school to which she has a claim at μ_n, which directly implies that $c_{n+1} P_i \mu(i)$, hence $c_{n+1} P_i \emptyset$. By the induction hypothesis, $\nu^{K_{(n)}}(i) = \emptyset$ and $\nu^{K_{(n)}}(j) = \mu_n(j)$ for all $j \in S \setminus \{i\}$ so c_{n+1} is also i's most preferred school to which she has a claim at $\nu^{K_{(n)}}$, which by definition means that $d^{K_{(n)}} = c_{n+1}$. Then, $\nu^{K_{(n)}+1}$ is identical to the matching obtained after the first step of Γ_{n+1}, where the lowest-priority student in $\mu^{K_{(n)}}(d^{K_{(n)}})$ is removed from $d^{K_{(n)}}$. Therefore, the next K_{n+1} steps of Δ are identical to the K_{n+1} steps of Γ_{n+1}. As Γ_{n+1} ends when i is removed from c_{n+1}, $j^{K_{(n+1)}} = i$ and $\nu^{K_{(n+1)}}(j) = \mu_{n+1}(j)$ for all $j \in S \setminus \{i\}$, which concludes our inductive argument.

Through our inductive argument, we have established that $j^{K_{(N)}} = i$ and $\nu^{K_{(N)}}(j) = \mu_N(j)$ for all $j \in S \setminus \{i\}$. By construction, i is removed from $\mu(i)$ in the first step of Δ so $|\nu^1(\mu(i))| < q_{\mu(i)}$. No student moves to $\mu(i)$ in any step $\ell < K_{(N)}$ as this would end Δ; therefore $|\nu^{K_{(N)}}(\mu(i))| < q_{\mu(i)}$. As $\mu(i) R_i \emptyset$, it follows that either $\mu(i) = \emptyset$, or $\mu(i) P_i \emptyset$ in which case $\nu^{K_{(N)}}(i) = \emptyset$ and $|\nu^{K_{(N)}}(\mu(i))| < q_{\mu(i)}$ imply that i has a claim to $\mu(i)$ at $\nu^{K_{(N)}}$. Then, by definition, $d^{K_{(N)}} R_i \mu(i)$ as, otherwise, i would be matched to $\mu(i)$ (or a more preferred school) rather than $d^{K_{(N)}}$. If $d^{K_{(N)}} P_i \mu(i)$, then i has a claim to $d^{K_{(N)}}$ at $\nu^{K_{(N)}}$. As $\mu_N(i) = \mu(i)$ and $\nu^{K_{(N)}}(j) = \mu_N(j)$ for all $j \in S \setminus \{i\}$, it follows that i has a claim to $d^{K_{(N)}}$ at μ_N, which contradicts the assumption that i does not have any claim at μ_N. We conclude that $d^{K_{(N)}} = \mu(i)$.

We have established that $j^{K_{(N)}} = i$, $d^{K_{(N)}} = \mu(i)$, and $|\nu^{K_{(N)}}(\mu(i))| < q_{\mu(i)}$; therefore, at step $K_{(N)}$ of Δ, i moves to $\mu(i)$ and the chain ends. We conclude that $K_{(N)} = L$; hence $j^L = i$ and $d^L = \mu(i)$. $\square$

Proof of Proposition 8. Consider a matching μ and a claim (i, c) at μ. Let

$$i = i^0 \to c = c^0 \to i^1 \to c^1 \to \cdots$$

be the reassignment chain initiated by (i, c) at μ, which we denote by Γ and let K be the number of steps after which Γ ends. Analogously to Definition 1, for every $k = 0, 1, \ldots, K$, let μ^k be the matching obtained at step k of Γ. We denote by μ^Γ the final matching obtained after, starting from μ, Γ has been carried out. Similarly, let

$$i = j^0 \to c = d^0 \to j^1 \to d^1 \to \cdots$$

be the *robust* reassignment chain initiated by (i, c) at μ, which we denote by Δ and let L be the number of steps after which Δ ends. Analogously to Definition 4, for every $\ell = 0, 1, \ldots, L$, let ν^ℓ be the matching obtained at step ℓ of Δ. We denote by μ^Δ the final matching obtained after, starting from μ, Δ has been carried out.

For every school $e \in C$, we define $\phi^\Gamma(e)$ as follows. If $|\mu^\Gamma(e)| < q_e$, then $\phi^\Gamma(e) = \mu^\Gamma(e)$. If $|\mu^\Gamma(e)| = q_e$, then $\phi^\Gamma(e) = \mu^\Gamma(e) \cup \{h \in S : \underline{\mu}^\Gamma(e) \succ_e h\}$. In words, $\phi^\Gamma(e)$ contains all those students who are either matched to e at μ^Γ or such that e is filled to capacity with higher-priority students. Throughout the proof, we make use of the following lemma, which is proved separately.

Lemma 6. *For all $\ell = 1, \ldots, L$, if $\mu^\Gamma(j^\ell) P_{j^\ell} \emptyset$ and $\nu^\ell(e) \subseteq \phi^\Gamma(e)$ for all $e \in C$, then j^ℓ has a non-vacuous claim to $\mu^\Gamma(j^\ell)$ at ν^ℓ.*

We proceed with an inductive argument to prove the following statement:

For all $\ell = 1, \ldots, L$ and for every school $e \in C$, $\nu^\ell(e) \subseteq \phi^\Gamma(e)$.

We begin by showing that our statement is satisfied for $\ell = 1$. By an analogous reasoning to the one developed in the proof of Lemma 1, it is sufficient to show that $i \in \phi^\Gamma(c)$. If (i, c) is vacuous at μ, then i is removed from c in step K of Γ so $|\mu^\Gamma(c)| = q_c$ and $\underline{\mu}^\Gamma(c) \succ_c i$, which by definition implies that $i \in \phi^\Gamma(c)$. If (i, c) is not vacuous, then by definition $i \in \mu^\Gamma(c) \subseteq \phi^\Gamma(c)$.

We next suppose that our statement is satisfied for some $\ell = 1, \ldots, L-1$ (induction hypothesis) and show that it is then also satisfied for $\ell + 1$. Again by a reasoning analogous to the one developed in the proof of Lemma 1, it is sufficient to show that $j^\ell \in \phi^\Gamma(d^\ell)$. Suppose towards a contradiction that $\mu^\Gamma(j^\ell) P_{j^\ell} d^\ell$. By definition, $d^\ell R_{j^\ell} \emptyset$ so $\mu^\Gamma(j^\ell) P_{j^\ell} \emptyset$ which, combined with our induction hypothesis, means that Lemma 6 applies. It follows that j^ℓ has a non-vacuous claim to $\mu^\Gamma(j^\ell)$ at ν^ℓ, a contradiction since j^ℓ would then be matched to $\mu^\Gamma(j^\ell)$ (or a more preferred school) rather than d^ℓ. We conclude that $d^\ell R_{j^\ell} \mu^\Gamma(j^\ell)$. The remainder of the argument to show that $d^\ell R_{j^\ell} \mu^\Gamma(j^\ell)$ implies $j^\ell \in \phi^\Gamma(d^\ell)$ is once again analogous to the one developed in Lemma 1.

We have established through our inductive argument that $\nu^L(e) \subseteq \phi^\Gamma(e)$ for every $e \in C$. The last part of the proof makes use of that result to prove the equivalence between essential stability and robust-chain essential stability.

(RCES ⇒ ES) Suppose towards a contradiction that (i, c) is robust-chain vacuous but not vacuous at μ. On the one hand, (i, c) is robust-chain vacuous; therefore i is removed from c at step L of Δ, which means that $\nu^L(c)$ contains q_c students who have a higher priority than i. As $\nu^L(c) \subseteq \phi^\Gamma(c)$, it follows that $\phi^\Gamma(c)$ contains at least q_c students who have a higher priority than i. On the other hand, (i, c) is not vacuous so $i \in \mu^\Gamma(c)$. As $|\mu^\Gamma(\mu(j^L))| \leq q_{\mu(j^L)}$ and, by definition, the students in $\mu^\Gamma(\mu(j^L))$ have the highest priorities among the students in $\phi^\Gamma(\mu(j^L))$, it follows that $\phi^\Gamma(c)$ contains at most $q_c - 1$ students who have a higher priority than i, a contradiction.

(ES ⇒ RCES) Suppose towards a contradiction that (i, c) is vacuous but not robust-chain vacuous at μ. Then, Γ ends when $i = i^K$ is removed from $c = c^{K-1}$ and Δ ends when some student $j^L \neq i$ is matched to some school d^L such that $|\mu^L(d^L)| < q_{d_L}$. By construction, $|\mu^L(d^L)| < q_{d_L}$ implies that either $d^L = \mu(i)$ or $|\mu(d^L)| < q_{d^L}$. In both cases, $|\mu^1(d^L)| < q_{d^L}$; hence, for all $k = 0, \ldots, K-1$, $c^k \neq d^L$ as otherwise Γ would end before i is removed from c. A direct consequence is that $|\mu^k(d^L)| < q_{d^L}$ for all $k = 1, \ldots, K$.

By definition, j^L is removed from $\mu(j^L)$ at some step of Δ. At that point, $\mu(j^L)$ is filled to capacity with students who all have a higher priority than j^L. As schools get higher-priority students throughout a (robust) reassignment chain, this is still the case at step L; therefore $|\nu^L(\mu(j^L))| = q_{\mu(j^L)}$ and $\underline{\nu}^L(\mu(j^L)) \succ_{\mu(j^L)} j^L$. As $\emptyset$ is never filled to capacity, $\mu(j^L) \neq \emptyset$ so our inductive argument applies and yields $\nu^L(\mu(j^L)) \subseteq \phi^\Gamma(\mu(j^L))$. Therefore, $\phi^\Gamma(\mu(j^L))$ contains at least $q_{\mu(j^L)}$ students who have a higher priority than j^L. As $|\mu^\Gamma(\mu(j^L))| \leq q_{\mu(j^L)}$ and, by definition, the students in $\mu^\Gamma(\mu(j^L))$ have the highest priorities among the students in $\phi^\Gamma(\mu(j^L))$, it follows that all of the students in $\mu^\Gamma(\mu(j^L))$ have a higher priority than j^L. Therefore, $j^L \notin \mu^\Gamma(\mu(j^L))$ or, equivalently, $\mu(j^L) \neq \mu^\Gamma(j^L)$.

On the one hand, $\mu(j^L) \neq \mu^\Gamma(j^L)$ implies that j^L is matched to $\mu^\Gamma(j^L)$ at some step k of Γ, i.e, there exists $k = 1, \ldots, K-1$ such that $i^k = j^L$ and $c^k = \mu^\Gamma(j^L)$. Recall that $c^{k'} \neq d^L$ and $|\mu^{k'}(d^L)| < q_{d^L}$ for all $k' = 1, \ldots, K-1$; therefore, $d^L \neq \mu^\Gamma(j^L)$ and $|\mu^k(d^L)| < q_{d^L}$. By definition, it follows that $\mu^\Gamma(j^L) P_{j^L} d^L$, as otherwise j^L would be matched to d^L (or a more preferred school) rather than $\mu^\Gamma(j^L)$. On the other hand, $\mu^\Gamma(j^L) P_{j^L} \emptyset$ as otherwise j^L would not be matched to μ^Γ in step k of Γ. Moreover, our inductive argument implies that $\nu^L(e) \subseteq \phi^\Gamma(e)$ for all $e \in C$. Therefore, Lemma 6 applies and j^L has a non-vacuous claim to $\mu^\Gamma(j^L)$ at ν^L. By definition, it follows that $d^L R_{j^L} \mu^\Gamma(j^L)$, as otherwise j^L would be matched to $\mu^\Gamma(j^L)$ (or a more preferred school) rather than d^L. We conclude that $\mu^\Gamma(j^L) P_{j^L} d^L R_{j^L} \mu^\Gamma(j^L)$, a contradiction. □

Proof of Lemma 6.

As $|\mu^\Gamma(\mu^\Gamma(j^L))| \leq q_{\mu^\Gamma(j^L)}$ and, by definition, the students in $\mu^\Gamma(\mu^\Gamma(j^\ell))$ have the highest priorities among the students in $\phi^\Gamma(\mu^\Gamma(j^\ell))$, $\phi^\Gamma(\mu^\Gamma(j^\ell))$ contains at most $q_{\mu^\Gamma(j^\ell)} - 1$ students who have a higher priority than j^ℓ. By assumption, $\mu^\Gamma(j^\ell) P_{j^\ell} \emptyset = \nu^\ell(j^\ell)$ and $\nu^\ell(\mu^\Gamma(j^\ell)) \subseteq \phi^\Gamma(\mu^\Gamma(j^\ell))$; therefore j^ℓ has a claim to $\mu^\Gamma(j^\ell)$ at ν^ℓ. It remains to show that this claim is not vacuous.

Denote by Θ the reassignment chain initiated by $(j^\ell, \mu^\Gamma(j^\ell))$ at ν^ℓ and let M be the number of steps after which Θ ends. Analogously to Definition 1, for every $m = 0, 1, \ldots, M$, let ρ^m be the matching obtained at step m of Θ. By assumption, $\rho^0 = \nu^\ell \subseteq \phi^\Gamma(e)$ for all $e \in C$. Then, an analogous inductive argument to the one developed in the proof of Lemma 1 implies the following statement:

For all $m = 1, \ldots, M$ and for every school $e \in C$, $\rho^m(e) \subseteq \phi^\Gamma(e)$.

A direct consequence is that $\rho^M(\mu^\Gamma(j^\ell)) \subseteq \phi^\Gamma(\mu^\Gamma(j^\ell))$; therefore $\rho^M(\mu^\Gamma(j^\ell))$ contains at most $q_{\mu^\Gamma(j^\ell)} - 1$ students who have a higher priority than j^ℓ (as we have shown above that $\phi^\Gamma(\mu^\Gamma(j^\ell))$ does). Suppose towards a contradiction that j^ℓ's claim to $\mu^\Gamma(j^\ell)$ at ν^ℓ is vacuous. Then, by construction, j^ℓ is removed from $\mu^\Gamma(j^\ell)$ in step M of Θ. Therefore, $\rho^M(\mu^\Gamma(j^\ell))$ contains $q_{\mu^\Gamma(j^\ell)}$ students who have a higher priority than j^ℓ, a contradiction. □