Efficient and Essentially Stable Assignments

Andrew Kloosterman
Department of Economics
University of Virginia

Peter Troyan[*†]
Department of Economics
University of Virginia

This draft: April 26, 2017
First draft: July 18, 2016

Abstract

It is well-known in the matching literature that stability and Pareto efficiency cannot be jointly achieved. This paper alleviates this tension by defining an assignment to be essentially stable if any claim by an agent initiates a chain of reassignments that ultimately results in the initial claimant losing the initial object to a third agent. Our solution is practical, as explaining to agents why their claims are not valid is straightforward. We show that mechanisms based on Shapley and Scarf's TTC algorithm, while Pareto efficient, are not essentially stable; Kesten's EADA mechanism is both Pareto efficient and essentially stable.

JEL Classification: C78, D61, D63, I20

Keywords: matching, stability, efficiency, school choice, deferred acceptance

[*]We thank Mustafa Oğuz Afacan, Umut Dur, and Thayer Morrill for helpful discussions regarding this project.

[†]Mailing address: P.O. Box 400182, Charlottesville, VA, 22904. Email: ask5b@virginia.edu and troyan@virginia.edu.

1 Introduction

It is well known that there is a tradeoff between efficiency and equity in matching problems. The celebrated deferred acceptance (DA) algorithm of Gale and Shapley (1962) always produces an equitable match, and in fact produces the most efficient equitable match among all equitable matches. However, it does not always produce a Pareto efficient match: there may be non-equitable matches that Pareto dominate it. Are all non-equitable matches equally non-equitable? Here, we argue that the answer is no, and in fact there is a non-equitable and efficient match that is essentially just as equitable as classically defined equitable matches.

We consider one-sided matching problems of agents to objects where equity is a normative property, but it has a direct connection to the concept of *stability* that is familiar from the two-sided matching literature, and indeed has led the one-sided matching literature to directly adopt stability as a desideratum (albeit sometimes under the different names *fairness* or *no justified envy*). Agents are exogenously endowed with *priorities*. Given an assignment, an agent is said to have a (justified) *claim* to object A if she prefers A to her assignment, and she has higher priority than another agent who is assigned to A. Stability is the property that there should be no such claims.

At first glance, the classic definition of stability seems very reasonable. However it misses an important point: if an agent were to have a claim on an object, granting their claim displaces an agent currently assigned to the object. This agent will then have to be reassigned, and, using the same justification as the initial agent, she can claim her favorite object at which she has high enough priority. This will displace yet another agent, and so on. Eventually this chain of reassignments will end when some agent is reassigned to an object that is available (or determines that everything that is available is unacceptable to her and takes her outside option). It is possible that the agent who made the initial claim will be displaced by some agent further down the chain. In this case, as the initial agent will ultimately not receive the object she laid claim to, the claim is said to be *vacuous*.

We propose a new definition of stability that expands the set of stable matches by allowing some claims to remain: namely, those that are vacuous. While this is a weaker definition of stability, we argue that it captures the essential feature of the standard definition, in the sense that if an agent were to try to assert their claim, we can clearly explain to them why doing so will be for naught. For example, if some agent i were to claim object A, one could convince her to relinquish this claim by walking her through its ultimate effects in the following way:

> "Yes, I agree that your claim at A is valid because you have higher priority than agent j. But, if I assign you to A, then j will need to be reassigned, and she will claim B, and her claim at B is equally as valid as yours was at A. This will displace agent k, who will then claim A, and, because she has higher priority than you at A, I will need to give it to her. So, while you may claim A, ultimately, you will not receive it anyway."

For this reason, we call an assignment with only vacuous claims an *essentially stable* matching.

Stability is a constraint that limits the set of assignments that can be implemented, and indeed it is well-known that stability is in general incompatible with other desirable properties such as Pareto efficiency. The reason for imposing only essential stability is that it allows us to implement a larger set of assignments. The main contributions of this paper are (i) to introduce the concept of essential stability and (ii) to show that an essentially stable *and Pareto efficient* matching always exists; furthermore, whenever DA produces an inefficient assignment, there is an essentially stable and Pareto efficient assignment that Pareto dominates it.[1] Thus, by requiring only essential stability, we improve upon the DA outcome, entirely eliminate the tension with efficiency, and still adhere closely to the general principle behind imposing stability constraints in the first place.

[1] Stable matches themselves are also essentially stable, so the existence of an essentially stable match is trivial.

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.[2] Stability is known to be an important concern for school choice and, partially for this reason, DA has become a widely used student placement algorithm, including in large school districts such as Boston and New York City.[3] This suggests that stability constraints do indeed play an important role in many real-world markets, even at the price of efficiency. For example, using data from eighth grade assignment in New York City, Abdulkadiroğlu et al. (2009) show that (on average) over 4,000 students would receive a more preferred assignment under a Pareto efficient matching compared to the DA matching that was actually implemented. However, this comes at the price of stability, as they also show that implementing a Pareto efficient assignment also leads to a much larger increase in the number of blocking pairs. New York City decided to use DA, and so chose stability over efficiency; the goal of this paper is to provide a practical alternative definition of stability that will eliminate the need to make such a choice.[4]

Once we have established the existence of Pareto efficient and essentially stable matchings, the next natural question is the existence of essentially stable and Pareto efficient mechanisms. That is, how can we implement such a matching? The proof of our main result answers this immediately as it directly shows that the *efficiency-adjusted deferred acceptance* (EADA) mechanism introduced by Kesten (2010) produces an essentially stable match.[5] We

[2]For concreteness, we will henceforth use this terminology, where the agents are students and the objects are seats at schools; however, our results will apply to many other applications as well.

[3]See Abdulkadiroğlu et al. (2005a) and Abdulkadiroğlu et al. (2005b) for discussions of redesign of school choice procedures (and the role of economists) in New York City and Boston, respectively.

[4]In New Orleans, the first mechanism used was an unstable, yet Pareto efficient, mechanism (the top trading cycles mechanism of Shapley and Scarf, 1974). However, this was only in place for a short time, and was soon abandoned in favor of DA, providing further evidence that stability-type constraints are important to real-world practitioners, even at the cost of efficiency.

[5]Our formal proof actually uses the outcome-equivalent *simplified efficiency-adjusted deferred acceptance* (SEADA) mechanism introduced by Tang and Yu (2014).

also analyze the two other most commonly considered Pareto efficient mechanisms, the *top trading cycles* (TTC) mechanism and a variant of it that first runs DA and follows this by running top trading cycles (DA+TTC).[6] We show that, while efficient, both of these mechanisms are not essentially stable and so we call them strongly unstable.

Previous work on the incompatibility of Pareto efficiency and stability falls into two categories. One line of research has focused on asking students if they will "consent" to have their priority violated and then ensures the mechanism is designed in such a way that students cannot gain from choosing not to consent, a property that Dur et al. (2015) call *no-consent-proofness*. The basic idea behind no-consent-proofness is first considered in Kesten (2010)'s original paper introducing EADA. Dur et al. (2015) show that EADA is the unique constrained efficient mechanism that Pareto dominates DA and is no-consent-proof. The reason that EADA is no-consent-proof is that a student's own assignment is unaffected by her consent decision, and so all students end up being indifferent between consenting and not consenting.

While related, there is an important conceptual distinction between the no-consent-proofness approach and the essential stability approach. Essential stability is a novel definition of what it means for a matching to be stable; no-consent-proofness is a procedural justification for why, given a mechanism, students should affirmatively consent to violations of the classical definition of stability.[7,8] The second line of research to which our paper belongs has attempted to follow the former approach of introducing weaker definitions of

[6]TTC was first proposed by Shapley and Scarf (1974), and extended to school choice settings by Abdulkadiroğlu and Sönmez (2003). Erdil and Ergin (2008) introduce DA+TTC, which is also analyzed by Alcalde and Romero-Medina (2015).

[7]Note also that no-consent-proofness is a property of a mechanism, while essential stability is a property of a matching (see also footnote 8).

[8]It turns out that EADA is both no-consent-proof and essentially stable, but the concepts of no-consent-proofness and essential stability themselves are distinct. For example, consider the following mechanism, which is no-consent-proof, but not essentially stable: First, 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, the DA+TTC assignment may have non-vacuous claims, and so this mechanism may produce assignments that are not essentially stable.

stability. Most of these papers 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.[9] Work in this vein includes Alcalde and Romero-Medina (2015), who introduce the concept of $\tau-$fairness, and Cantala and Pápai (2014), who introduce the concepts of reasonable stability and secure stability. Morrill (2016) introduces the concept of a *legal assignment,* 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 that is legal and at which she is assigned to c. He introduces an iterative procedure for finding the set of legal assignments, which he shows is equivalent to the von Neumann-Morgenstern stable set (Von Neumann and Morgenstern, 2007).[10] Tang and Zhang (2016) introduce their own new definition of weak stability for school choice problems that is also closely related to vNM stable sets.[11,12]

While all of these concepts are theoretically very appealing, we think a main advantage of essential stability is the ease with which it can be explained, which is important for practical applications. There are several reasons for this. First, essential stability is a property of a matching, independent of the mechanism that determines matchings, which can be very abstract, especially

[9]This idea is similar in spirit to the idea of "bargaining sets" introduced by Zhou (1994) for cooperative games.

[10]Ehlers (2007) studies vNM stable sets in the context of marriage markets (see also Wako (2010)). Outside of the matching market literature, Ray and Vohra (2015) investigate how the stable set changes when agents are more forward-looking than von Neumann and Morgenstern suggest.

[11]Alcalde and Romero-Medina (2015) and Cantala and Pápai (2014) show that the DA+TTC mechanism satisfies their respective definitions of stability, while EADA satisfies our definition as well as those of Morrill (2016) and Tang and Zhang (2016). However, they are all independent properties and in the appendix, we show formally that these other concepts are distinct from ours.

[12]Afacan et al. (2015) show that a variation of EADA always produces a so-called "sticky stable" assignment, which is based on the idea that students will only invest in a costly appeal if they can greatly improve their assignment. Other papers that study EADA in different contexts are Bando (2014), who shows that the EADA outcome is equivalent to a strictly strong Nash equilibrium outcome in the preference revelation game induced by standard DA, and Dogan (2014), who proposes a mechanism similar to EADA in the context of affirmative action.

to non-experts.[13] Second, suppose a student raised an objection to a particular matching. Using other weakenings of stability, we would have to show them that every possible match where they get the school they want is not feasible (either because of a counter-objection or possibly for more complicated reasons). Given the size of these markets, this alone would be an enormous task. Even then, the counter-objection may involve a completely separate set of agents, and, since it does not directly impact the student making the claim, it may be very unclear to them why the counter-objection takes precedence over their original objection. On the other hand, following our approach, we simply walk them through the chain of reassignments in the manner highlighted above to show them that their objection is actually vacuous, and thus there is no reason to make the claim in the first place. Still, achieving both efficiency and stability is a very important problem, and there may be more than one way to do so. All of the aforementioned papers have useful and interesting results, and we view all the papers in this literature as complementary, as they each provide a different way to reach Pareto efficiency while still achieving some reasonable definition of stability.

2 Preliminaries

2.1 Model

There is a set of agents S who are to be assigned to a set of objects C. Each $i \in S$ has a strict preference relation P_i over C and each $c \in C$ has a strict priority relation $\succ_c$ over S. Let $P = (P_j)_{j \in S}$ denote a profile of preference relations, one for each agent, and $\succ = (\succ)_{c \in C}$ the profile of priority relations. Each $c \in C$ has a capacity q_c which is the number of agents in S that can

[13]For example, under EADA, convincing a student that consenting to a priority violation is not harmful requires explaining the inner workings of the EADA mechanism in detail. This is very nontrivial, and thus students may default to the status quo of not consenting (especially since consenting also does not benefit them). Matchings, on the other hand, are more concrete objects, and so explaining to a student why her claim is vacuous for a specific matching is simple, as we can just walk her through the steps noted above.

be assigned to it. Let $q = (q_c)_{c \in C}$ denote a profile of capacities. We assume $S, C, \succ$, and q are fixed throughout the paper, and associate a **market** with its preference profile P. Probably the best-known application of this model is when S is a set of *students* and C is a set of *schools* (or "colleges"). In this case, the priorities $\succ$ are determined by state/local laws and/or criteria determined by each school district. While we will generally use the student/school terminology for concreteness, it should be noted that the model can be applied to many other real-world assignment problems. Examples include the military assigning cadets to branches, business schools assigning students to projects, universities assigning students to dormitories, or cities assigning public housing units to tenants.[14]

A **matching** is a correspondence $\mu : S \cup C \to S \cup C$ such that, for all $(i, c) \in S \times C$, $\mu(i) \in C$, $\mu(c) \subseteq S$, and $\mu(s) = c$ if and only if $s \in \mu(c)$.[15] 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 $i \in S$.[16] 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 S, and not the schools C. This is a standard view in the mechanism design approach to object assignment and 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 (also called no justified envy). Typically, a matching is deemed unfair if student i desires a school c and has higher priority at c than some other student who is assigned there. Fairness is the natural counterpart of the more familiar concept of "stability" from the

[14]For more detail on these markets, see Sönmez (2013), Sönmez and Switzer (2013), Fragiadakis and Troyan (2016), Chen and Sönmez (2002), Chen and Sönmez (2004), Sönmez and Ünver (2005), Sönmez and Ünver (2010), and Thakral (2015).

[15]We assume all students are assigned to a school and vice-versa. While in practice some students may prefer taking an "outside option" to some schools, our model is without loss of generality, as we could simply model the outside option as a particular school $o \in C$ with capacity $q_o \geq |S|$.

[16]R_i denotes the weak part of i's preference relation P_i. Given the assumption that preferences are strict, aR_ib but not aP_ib if and only if $a = b$.

two-sided matching literature. Formally, given a matching μ, we say student i **claims a seat at school** c if (i) $cP_i\mu(i)$ and (ii) either $|\mu(c)| < q_c$ or $i \succ_c j$ for some $j \in \mu(c)$. We will sometimes use (i, c) to denote i's claim of c. If no student claims a seat at any school, then we say μ is **stable**.[17]

Let $\mathcal{M}$ denote the space of all possible matchings, and $\mathcal{P}$ denote the space 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.[18] For any mechanism ψ and preferences P, $\psi(P)$ is the matching output by ψ when the submitted preferences are P. Mechanism ψ is said to be Pareto efficient if $\psi(P)$ is a Pareto efficient matching for all P. Mechanism ψ is said to be stable if $\psi(P)$ is a stable matching for all P.

2.2 Motivating example

Deferred acceptance (DA) is one of the benchmark algorithms that form the foundation for both the theory and practical applications of a myriad of matching markets.[19] DA is an enormously successful mechanism in the field because it produces the so-called *student-optimal stable match:* that is, any other stable match ν is Pareto dominated by the DA match μ^{DA}. The prevalence of

[17] Some papers break this definition into two parts, *nonwastefulness* and *fairness*. We use "stability" because we think this terminology is more familiar, and will serve to highlight the connection of our results with the broader literature. An additional important property is *individual rationality*, which says that no student is assigned to a school that she disprefers to taking an outside option. However, as noted in footnote 15, we can model the outside option as a particular school $o \in C$ that has enough capacity for all students. Then, individual rationality is implied by nonwastefulness.

[18] A mechanism can also depend on $S, C, \succ$, and q but we have assumed those are fixed for the remainder of the paper and so suppress the dependence of ψ on them.

[19] In particular, we consider the student-proposing version of the DA mechanism, which works as follows. To start, all students apply to their favorite school. Each school c tentatively accepts the q_c students with the highest priority (or all of them if less than q_c apply). Then, rejected students apply to their favorite school that has not yet rejected them, and each school tentatively accepts the highest priority students from those that are currently tentatively accepted and new applicants. The algorithm continues with rejected students making new applications and schools tentative acceptances until no students make new applications. Since this mechanism is very standard and well-known, we do not provide a complete formal definition. Such a definition can be found in many other papers (for example, Gale and Shapley (1962) or Abdulkadiroğlu and Sönmez (2003)).

DA in the field suggests that such stability constraints do play an important role in many settings. However, as discussed in the introduction, ensuring stability comes at the price of efficiency: DA is in general not a Pareto efficient mechanism. The example below illustrates this point, and serves to motivate our new definition of stability. Throughout the paper, given preferences P, the matching produced by the DA algorithm is denoted as either $DA(P)$, or, if the preferences are understood, as μ^{DA}. Student i's assigned school under the DA matching is denoted as either $DA_i(P)$, or, if the preferences are understood, as $\mu^{DA}(i)$.

Example 1. Let there be 5 students $S = \{i_1, i_2, i_3, i_4, i_5\}$ and 5 schools $C = \{A, B, C, D, E\}$, each with capacity 1. 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	$\vdots$
i_2	i_3	i_4	i_5	
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 μ^{DA} 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 of course implies that μ^* and $\mu^\dagger$ are both unstable. For instance, under μ^*, student i_4 claims the seat at C (because $i_4 \succ_C i_2 = \mu^*(C)$) and under $\mu^\dagger$, 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 $i_5 \succ_B i_3 = \mu^\dagger(D)$).

There are simpler examples to show that the DA match 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 will 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,[20] 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 should also grant the next request of i_1, since he 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 i_3 at B (in this case i_2) ends up claiming it, and so i_3's initial claim is vacuous. Student i_5 claiming the seat at 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^*. Similarly, 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 under $\mu^\dagger$ were.

Thus, both μ^* and $\mu^\dagger$ are "unstable" according to the classical definition, but they are unstable in different ways (which will be made more precise in a moment). While student i_3 can protest $\mu^\dagger$ and request B, if she does so, she will ultimately be rejected from B. What is more, while in practice students may not fully understand the workings of the mechanism, if a student were to

[20]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.

protest, it would be very easy to walk her through the above chain of reassignments to show her that granting her request would ultimately not be beneficial to her, and therefore convince her to relinquish her claim. Thus, to the extent that the goal of imposing stability is to prevent students from protesting an assignment, matchings like $\mu^\dagger$, while not fully stable in the classical sense, are "essentially" stable.

On the other hand, the instability of matching μ^* is a much stronger type of instability, because i_4 claiming C will ultimately benefit her, and so we would not be able to convince her to relinquish her claim. Our new definition of stability is designed to capture this idea and, in the process, recover inefficiencies by expanding the set of "feasible" matchings to include those like $\mu^\dagger$, but 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 of a seat at c.

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

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

where,

- $i^0 = i$, $\mu^0 = \mu$, $c^0 = c$ and for each $k \geq 1$:
- i^k is the lowest priority student in $\mu^{k-1}(c^{k-1})$.
- μ^k is the matching defined as: $\mu^k(j) = \mu^{k-1}(j)$ for all $j \neq i^{k-1}, i^k$, $\mu^k(i^{k-1}) = c^{k-1}$, and student i^k is unassigned
- c^k is student i^k's most preferred school at which she can claim a seat under μ^k
- and terminates at the first K such that $|\mu^K(c^K)| < q_{c^K}$.

We should note a few features of assignment chains. First, a student i may claim a seat at a school c either because c is not filled to capacity, or because she has higher priority than someone currently assigned there. Whenever the former occurs, that is the end of the reassignment chain (if this occurs immediately, then the original matching μ was not nonwasteful). Second, the definition of a reassignment is closely related, but slightly different from, the common definition of a rejection chain. The main difference is that in a rejection chain, when a student is rejected from a school, she applies to the next school on her preference list. In a reassignment chain, she may in principle apply to a school that is more preferred than the school she is rejected from. Note that a school or a student can appear multiple times in a reassignment chain.[21]

For a reassignment chain Γstarted by a claim (i, c), if there exists $j \neq 0$ such that $i^j = i$, we say that the reassignment chain **returns to** i. If the reassignment chain returns to i, then i will ultimately be removed from the school c that she claimed initially by some student with higher priority. When this is the case, we say that claim (i, c) is **vacuous**.

Definition 2. At matching μ, if all claims (i, c) are vacuous, then μ is **essentially stable**. If there exists at least one claim that is not vacuous, μ 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, under $\mu^\dagger$, the claims (i_3, B) and (i_5, D) are vacuous, because they ultimately result in the initial claimant losing the seat he claimed to a higher priority student. Under 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.

[21] Note also that reassignment chains are well-defined (i.e., they must end in finite steps). This is because, if a student i^k is rejected from a school c^{k-1}, then $\mu^k(c^{k-1})$ must be filled to capacity with higher priority students. For all $k' > k$, the lowest priority student in $\mu^{k'}(c^{k-1})$ only increases, and thus, if i^k is ever rejected again later in the chain, the best school at which she can claim a seat is ranked lower than c^{k-1}. Since students only apply to worse and worse schools as the chain progresses and preference lists are finite in length, the chain must end.

Similarly to classical stability, a mechanism ψ is said to be essentially stable if $\psi(P)$ is an essentially stable matching for all P. If ψ is not essentially stable, then we say it is strongly unstable.

3 Results

It is obvious that any stable match is also essentially stable and so the existence of an essentially stable match is trivial. The more interesting question is whether other, and specifically Pareto efficient, essentially stable matches exist. The first theorem shows that the answer to this question is "yes."

Theorem 1. *An essentially stable and Pareto efficient matching exists for every market* $(S, C, P, \succ, q)$.

Proof. The proof proceeds by providing an algorithm that always produces an essentially stable and Pareto efficient match; namely, the simplified efficiency adjusted deferred acceptance (SEADA) mechanism of Tang and Yu (2014).[22] Say school c is **underdemanded** at matching μ if $\mu(i) R_i c$ for all i. For the DA matching, the set of underdemanded schools are the schools that never reject a student in the running of the algorithm.

SEADA

Round 0 Compute the deferred acceptance outcome $DA(P)$. Identify the schools that are underdemanded, and for each student at these schools, make their assignments permanent. Remove these students and their assigned schools from the market.

Round $r \geq 1$ Compute the DA outcome on the submarket consisting of those students who still remain at the beginning of round r. Identify the schools that are underdemanded, and for each student at these schools,

[22] SEADA is a simplification of Kesten's EADA algorithm that is outcome equivalent. More formally, the definition of SEADA that we use is one in which it is assumed all students "consent."

make their assignments permanent. Remove these students and their assigned schools from the market.

Let $\mu^0 = DA(P)$, and, for $r \geq 1$, let μ^r denote the matching at the end of round r, defined as follows: if i was removed from the market prior to the beginning of round r, then $\mu^r(i) = \mu^{r-1}(i)$; if i remains in the market at the beginning of round r, then $\mu^r(i)$ is the school she is assigned at the end of DA on the round r submarket. The final output of the mechanism is $\mu^{SEADA} = \mu^R$, where round R is the final round of the above algorithm.

Tang and Yu (2014) show that μ^R is Pareto efficient. What we show is that μ^R is also essentially stable. Consider some arbitrary claim (i, c), and let Γ denote the reassignment chain initiated by this claim.[23] 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^R is essentially stable.[24]

The proof proceeds by defining two alternative preference profiles, one that gives i her SEADA assignment when DA is run and one for which the rejection chain i starts when DA is run that is ultimately identical to the reassignment chain Γ. A well-known monotonicity property of DA is applied to show that both profiles lead to the same assignment for i and so the reassignment chain Γ must end with i back at her SEADA assignment. The formal proof follows, with technical details relegated to the appendix.

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 $c \in C$ if $bR'_ic \implies bR_ic$. 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. In words, P' is a monotonic transformation of another preference profile P at a matching μ if, for all i, the ranking of $\mu(i)$ only increases in moving from P_i to P'_i.

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

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

Lemma 1. *(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) under μ^R. Because DA on the round R submarket is stable, only students who were removed in a round strictly earlier than R (the final round) can have a claim. That is, i must have been removed in some round $\hat{r} < R$. Define an alternative preference profile $P^{\hat{r}}$ as follows: for any student j removed before round $\hat{r}$, $P^{\hat{r}}_j$ ranks their assignment $\mu^{\hat{r}}(j)$ first, and the remaining schools in the same order as the true P_j; for all j not removed before round $\hat{r}$, $P^{\hat{r}}_j = P_j$. Note that this is a simple way to describe preferences so that $DA(P^{\hat{r}}) = \mu^{\hat{r}}$.[25]

Define a second preference profile $\bar{P}_j$ as follows: for each $j \neq i$, $\bar{P}_j$ ranks $\mu^R(j)$ 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 2. *Student i's DA assignment at the end of round $\hat{r}$ is the same as her DA assignment under $\bar{P}$, which is the same as her assignment at the end of round R: $DA_i(\bar{P}) = DA_i(P^{\hat{r}}) = \mu^R(i)$.*

The lemma is formally proved in the appendix, but the main step is that $\bar{P}$ is a monotonic transformation of $P^{\hat{r}}$ at $DA(P^{\hat{r}})$. 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 him to apply to his 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

[25] Raising school $\mu^{\hat{r}}(j)$ for all j removed prior to round $\hat{r}$ 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 $\hat{r}$ will ever apply to such a school because it is underdemanded.

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 these students have been (tentatively) assigned to a school, we allow i to enter by applying to the first school on his 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 his most preferred school that has not yet rejected him, and so on, until we reach a school a^K with an empty seat, at which point the rejection chain (and the entire DA algorithm) end, and all tentative matchings are made final.

Consider running DA on the preference profile $\bar{P}$ in this manner where i enters the market last. All students j other than i are tentatively matched to $\mu^R(j)$, and then i starts a rejection chain which, as the next lemma proves, turns out be identical to the reassignment chain Γ.

Lemma 3. *The final matching at the end of* Γ *is* $DA(\bar{P})$.

By Lemma 3, the outcome of the reassignment chain Γ is the same as the outcome of $DA(\bar{P})$. By Lemma 2, $DA_i(\bar{P}) = \mu^R(i)$, and so i's assignment at the end Γ is also $\mu^R(i)$. The only way this is possible is if i is rejected from c at some point in Γ; that is, the claim (i, c) is vacuous. □

Tang and Yu (2014) also show that, when DA is not efficient, the SEADA outcome Pareto dominates DA. This leads to the following immediate corollary.

Corollary 1. *If* μ^{DA} *is not Pareto efficient, then* μ^{SEADA} *is Pareto efficient, essentially stable, and Pareto dominates* μ^{DA}.

One additional remark is worth making. We defined a claim (i, c) as vacuous if the induced reassignment chain returns to i and ultimately rejects her from c. However, one might argue that maybe i claims a seat at a very good school c and, while she is rejected from c in the reassignment chain, perhaps i's final match at the end of the chain is still better than i's initial match. However, the proof Theorem 1 actually showed that not only does the reassignment chain return to reject i from c, but also i ends up with their initial match. Hence, we equally could have defined a vacuous claim as the property that i's match at the end of the reassignment chain is the same as i's initial match and still obtained the same result.

4 Which mechanisms are essentially stable?

Theorem 1 shows that an essentially stable and Pareto efficient matching always exists. Furthermore, the proof directly shows that the SEADA and, equivalently, EADA, mechanisms are essentially stable mechanisms. This raises the question of whether other Pareto efficient mechanisms are essentially stable or not. In this section, we look at two Pareto efficient mechanisms that are popular in the literature, top trading cycles (TTC), and DA+TTC, and show that both are strongly unstable.

Top Trading Cycles

One of the most commonly proposed alternative mechanisms to DA is the top trading cycles (TTC) mechanism of Shapley and Scarf (1974), extended to school choice mechanisms by Abdulkadiroğlu and Sönmez (2003). The intuition behind TTC is that "priority" at a school is interpreted as an ownership right to a seat at that school, which can be traded away: if i has high priority at j's first choice, and j has high priority at i's first choice, then TTC allows i and j to trade, even though this may violate the priority of some third student k who is not involved in the trade. By continually making all mutually beneficial trades, we eventually end up at a final assignment that is Pareto efficient.[26] However, as TTC ignores the priorities of students not involved in a cycle, it is well-known to be unstable. It turns out that TTC is not just unstable, but is in fact strongly unstable.

Theorem 2. *The top trading cycles mechanism is strongly unstable.*

Proof. The proof is by example. Let there be 5 students $S = \{i_1, i_2, i_3, i_4, i_5\}$ and 5 schools $C = \{A, B, C, D, E\}$ with one seat each. The preferences and priorities are given in the table below.

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

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

TTC proceeds with each student pointing to her favorite school and each school to its top-priority student. In the initial round, there is one cycle: (i_1, B, i_3, C, i_1). We implement this trade between i_1 and i_3 and remove them from the market with their assignments. Then, in the next round, i_2 forms a self-cycle with A and is therefore assigned to it. In the final round, i_4 and i_5 are assigned to D and E, respectively, and the final TTC outcome is

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

Now, consider student i_2, who claims a seat at school C. The reassignment chain initiated by this claim is:

$$i_2 \to C \to i_3 \to B \to i_1 \to A.$$

We see that the i_2's claim is not vacuous, and so μ^{TTC} is strongly unstable. □

[26]See, for example, Abdulkadiroğlu and Sönmez (2003) for a formal definition of TTC as applied to school choice problems.

DA + TTC

One drawback of using TTC is that it completely ignores the priorities of students not involved in a trade, and so can produce very unstable assignments. Another alternative often proposed is to first calculate DA, and then allow students to trade by running TTC, using the initial DA assignments as the "ownership rights" in the TTC mechanism.[27] The advantage of this is that it will guarantee that students always receive an assignment that is no worse than their DA school, which is not guaranteed by using TTC alone. However, the next theorem shows that this, too, is a strongly unstable mechanism.

Theorem 3. *The DA+TTC mechanism is strongly unstable.*

Proof. Consider again the motivating Example 1. The DA outcome is

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

Now, TTC is applied by first giving each student her DA assignment as her initial endowment, and then creating a graph in which each student points to her favorite school, and the school points to the student who is endowed with it. Carrying this out, we see that there is only one cycle, (i_2, C, i_3, B, i_2) and so we implement this trade between i_2 and i_3 and remove them from the market with their assignments. After this, all remaining cycles are self-cycles, and so the final allocation is

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

This is the matching μ^* introduced in Example 1 which we found had a non-vacuous claim and so we have that μ^{DA+TTC} is strongly unstable. □

[27]Similar "improvement cycles" algorithms are proposed by Erdil and Ergin (2008) to recover inefficiencies in school choice caused by ties in priority among students.

5 Conclusion

This paper introduces the concept of essential stability, which is a weakening of classical stability that allows a matching to have some priority-based claims to objects as long as those claims are vacuous. The motivation for this definition is two-fold. First, it makes Pareto efficiency achievable by expanding the set of "stable" assignments, and second, it still adheres to the important principle behind the definition of classical stability: agents should not want to claim an object based on having a high priority for it. Indeed, essential stability makes evidently clear why there is no reason for an agent to claim an object even if she desires it and has high priority. The definition is simple enough that it can easily be explained and implemented by non-experts, which we believe is one of its main practical advantages that will allow it to be implemented in applications, thereby achieving more efficient assignments in practice.

Our paper focuses on resolving the mutual incompatibility of two desirable properties of matchings, Pareto efficiency and stability. Another property often discussed in the literature is strategyproofness, which is the property that all students have a dominant strategy to truthfully report their preferences. Abdulkadiroğlu et al. (2009) show an impossibility result that states that any mechanism that Pareto improves upon DA is not strategyproof.[28] At the same time, strategyproofness is neither necessary nor sufficient for successful market design in practice. For example, the mechanism used by the National Resident Matching Program to match doctors with hospitals performs well in practice even though it is not strategyproof (Kojima and Pathak, 2009; Kojima et al., 2013), while the New Orleans school district provides evidence that strongly unstable mechanisms like TTC (which is strategyproof) may not succeed.[29]

[28]This result applies to ex-post Pareto efficiency. From the viewpoint of ex-ante efficiency, a recent strand of literature started by Abdulkadiroğlu et al. (2011) and Miralles (2008) shows that highly manipulable algorithms such as the Boston mechanism may outperform strategyproof algorithms such as DA in some environments (see also Troyan, 2012; Featherstone and Niederle, 2014; Akyol, 2016).

[29]In the context of auctions, Ausubel et al. (2006) argue that while the VCG mechanism is the only efficient and strategyproof mechanism, it has a number of other drawbacks that limit its usefulness in practice.

We hope that the good stability properties we have highlighted here inspire future research to address this important issue, perhaps through laboratory experiments and ultimately field implementations of essentially stable and Pareto efficient mechanisms.

References

Abdulkadiroğlu, A., Y.-K. Che, and Y. Yasuda (2011): "Resolving Conflicting Preferences in School Choice: The "Boston" Mechanism Reconsidered," *American Economic Review*, 101, 399–410.

Abdulkadiroğlu, A., P. A. Pathak, and A. E. Roth (2005a): "The New York City high school match," *American Economic Review, Papers and Proceedings*, 95, 364–367.

——— (2009): "Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match," *American Economic Review*, 99, 1954–1978.

Abdulkadiroğlu, A., P. A. Pathak, A. E. Roth, and T. Sönmez (2005b): "The Boston Public School Match," *American Economic Review, Papers and Proceedings*, 95, 364–367.

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

Afacan, M. O., Z. H. Aliogullari, and M. Barlo (2015): "Sticky Matching in School Choice," *Available at SSRN 2596637*.

Akyol, E. (2016): "Welfare Comparison of Allocation Mechanisms under Incomplete Information," Working paper, Pennsylvania State University.

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

Ausubel, L. M., P. Milgrom, et al. (2006): "The Lovely but Lonely Vickrey auction," in *Combinatorial Auctions*.

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

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

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.

——— (2004): "An experimental study of house allocation mechanisms," *Economics Letters*, 83, 137–140.

Dogan, B. (2014): "Responsive affirmative action in school choice," *Available at SSRN 2468320*.

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," Tech. rep., Working paper, North Carolina State University, 2015.[19].

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

Erdil, A. and H. Ergin (2008): "What's the Matter with Tie-Breaking? Improving Efficiency in School Choice," *American Economic Review*, 98, 669–689.

Featherstone, C. and M. Niederle (2014): "Improving on Strategyproof School Choice Mechanisms: An Experimental Investigation," Working paper, Stanford University.

Fragiadakis, D. and P. Troyan (2016): "Designing Mechanisms to Focalize Welfare-Improving Strategies," Working paper, Texas A&M University.

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

Kesten, O. (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.

Kojima, F. and P. A. Pathak (2009): "Incentives and Stability in Large

Two-Sided Matching Markets," *American Economic Review*, 99, 608–627.

Kojima, F., P. A. Pathak, and A. E. Roth (2013): "Matching with Couples: Stability and Incentives in Large Markets," *The Quarterly Journal of Economics*, 128, 1585–1632.

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

Miralles, A. (2008): "School Choice: The Case for the Boston Mechanism," Boston University, unpublished mimeo.

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

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

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.

Sönmez, T. and T. Switzer (2013): "Matching with (Branch-of-Choice) Contracts at United States Military Academy," *Econometrica*, 81, 451–488.

Sönmez, T. and M. U. Ünver (2010): "House allocation with existing tenants: A characterization," *Games and Economic Behavior*, 69, 425–445.

Sönmez, T. and U. M. Ünver (2005): "House allocation with existing tenants: an equivalence," *Games and Economic Behavior*, 52, 153–185.

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.

Thakral, N. (2015): "Matching with Stochastic Arrival," Working paper, Harvard University.

Troyan, P. (2012): "Comparing School Choice Mechanisms by Interim and Ex-Ante Welfare," *Games and Economic Behavior*, 75, 936–947.

Von Neumann, J. and O. Morgenstern (2007): *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.

Zhou, L. (1994): "A new bargaining set of an n-person game and endogenous coalition formation," *Games and Economic Behavior*, 6, 512–526.

A Omitted proofs

Proof of Lemma 1. 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 $DA_i(P') = a'$. By part (i), we have $a'R'_i a$. Since P'_i is a monotonic transformation of P_i at a, $a'R'_i a$ implies $a'R_i a$. □

Proof of Lemma 2. We start by showing that $\bar{P}$ is a monotonic transformation of $P^{\hat{r}}$ at $DA(P^{\hat{r}})$. For each $j \in S$, let $DA_j(P^{\hat{r}}) = a_j$. For all j removed from the market at some round $r < \hat{r}$, $\mu^r(j) = \mu^{\hat{r}}(j) = a_j$. Thus, both $P_j^{\hat{r}}$ and $\bar{P}_j$ rank school a_j first, and $\bar{P}_j$ is trivially a monotonic transformation of $P_j^{\hat{r}}$ at a_j for these students.

Next, consider the students who are still in the market at the beginning of round $\hat{r}$. We first note the following result.

Lemma 4. *(Tang and Yu, 2014, Lemma 2) For each $r \geq 1$, μ^r Pareto dominates μ^{r-1}.*

Now, for all students who are still in the market at round $\hat{r}$, $P_j^{\hat{r}} = P_j$. Consider some such $j \neq i$. By the above lemma, $\mu^R(j) R_j a_j$ for all j. Since $P_j^{\hat{r}} = P_j$, this further implies that $\mu^R(j) R_j^{\hat{r}} a_j$. Now, consider preference profile $\bar{P}_j$. $\bar{P}_j$ simply raises $\mu^R(j)$ to the top of the ordering, without altering the relative rankings of any other objects (in particular, no objects "jump" over student j's round $\hat{r}$ assignment a_j in the move from $P_j^{\hat{r}}$ to $\bar{P}_j$), and so $\bar{P}_j$ is a monotonic transformation of $P_j^{\hat{r}}$ at a_j for all $j \neq i$.

Last, consider student i. She is removed in round $\hat{r}$, and so $cP_i^{\hat{r}}a_i$ (otherwise, student i would not claim a seat at c under μ^R).[30] By similar logic (no school a' "jumps" over a_i in going from $P_i^{\hat{r}}$ to $\bar{P}_i$), $\bar{P}_i$ is a monotonic transformation of $P_i^{\hat{r}}$ at a_i. Thus, we have shown that $\bar{P}_j$ is a monotonic transformation of $P_j^{\hat{r}}$ at $DA_j(P^{\hat{r}})$ for all $j \in S$, and so preference profile $\bar{P}$ is a monotonic transformation of preference profile $P^{\hat{r}}$ at $DA(P^{\hat{r}})$.

Next, given a matching μ, say student i is **not Pareto improvable** if, for every ν that Pareto dominates μ, $\nu(i) = \mu(i)$.

Lemma 5. *(Tang and Yu, 2014, Lemma 1) All students matched to underdemanded schools at $DA(P)$ are not Pareto improvable (with respect to P).*

Since $\bar{P}$ is a monotonic transformation of $P^{\hat{r}}$ at $DA(P^{\hat{r}})$, Lemma 1, part (ii) gives $DA_j(\bar{P})R_j^{\hat{r}}DA_j(P^{\hat{r}})$ for all $j \in S$, i.e., the matching $DA(\bar{P})$ Pareto dominates the matching $DA(P^{\hat{r}})$ with respect to $P^{\hat{r}}$. Since i is removed in round $\hat{r}$, she must be matched with an underdemanded school at $DA(P^{\hat{r}})$ which, by Lemma 5, implies that she is not Pareto improvable relative to $P^{\hat{r}}$. Since $DA(\bar{P})$ Pareto dominates $DA(P^{\hat{r}})$ and i is not Pareto improvable, her matching does not change: $DA_i(\bar{P}) = DA_i(P^{\hat{r}})$. Since i is removed at round $\hat{r}$, her assignment at $R > \hat{r}$ is the same as her assignment at the end of round $\hat{r}$: $\mu^R(i) = DA_i(P^{\hat{r}})$. $\square$

Proof of Lemma 3. 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 $\mu^R(j)$ and is tentatively matched to $\mu^R(j)$. No rejections occur because each $j \neq i$ is assigned to the unique seat they are assigned to under μ^R. 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

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

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

which eventually terminates in some K when a student applies a school with a vacant seat. When a student i^{k+1} is rejected, they go to the next school on their list and apply. It may be the case that when a student applies to a school, she is rejected immediately, and must continue down their list. Formally, if $i^k \neq i^{k+1}$ we say step $\Xi(k)$ is **effective**. If a step is ineffective ($i^k = i^{k+1}$), then the same student who applied is also the one rejected, and nothing would change if i^k simply skipped her application to a^k. 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 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 $\mu^R(j)$). Consider step 1. In the former case, a student j is rejected from their initial match $c = \mu^R(j)$.[31] The rest of their preference list $\bar{P}_j$ coincides with their true preferences P_j so they go down their true list P_j until they reach a school where they have 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 $k = 1$ 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 $k = 2$ matching under Γ and so on for each additional matching, until the same student i^K applies to the first school c^K that has an empty seat, at which point both Γ and Ξ' end at the same final matching.[32] □

[31]Since i is assumed to have a claim at c under μ^R (and μ^R is nonwasteful), we have $i \succ_c j$, for some $j \in \mu^R(c)$.

[32]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

B Comparison to other weakenings of stability

In this appendix, first we show formally that our definition of essential stability is formally 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^* 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 Morrill (2016) and Tang and Zhang (2016) also are satisfied by the EADA mechanism and so it is less obvious that they are formally distinct. However, as we show here, they do not necessarily lead to the same prediction.

We first consider Morrill (2016). Rather than defining stability on a matching itself, Morrill (2016) defines stability on a set of matchings; i.e., an individual matching μ is not "stable" independently, but is only stable (or fair, in his terminology) in relation to other matchings. More formally, say that 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 $j \in \nu(a)$. Given a set of matchings M, a matching μ is **possible** for M if μ is not blocked by any $\nu \in M$. Denote the set of possible matchings for a set M by $\pi(M)$. Then, a set of matchings F is **fair** if[33]

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.

[33] Morrill (2016) first introduces the concept of a *legal* set of assignments. He then introduces the definition of fairness given here. This definition is similar, but slightly different, from the definition of a legal set of assignments, but he shows that the fair set of assignments is actually equivalent to the legal set of assignments. Thus, the result below applies to both the fair set and the legal set of assignments.

1. For all $\mu \in F$, μ is not blocked by any $\nu \in F \cup \pi(F)$

2. For all $\mu \notin F$, μ is blocked by some $\nu \in \pi(F)$.

Example 1 can be used to show that essential stability is different from fairness as defined in Morrill (2016) (and, by extension, from weak stability as defined in Tang and Zhang (2016)). More precisely, we exhibit a matching μ that must be included in any fair set of matchings F, 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.

Proposition 1. *Let F be a fair set of matchings, and let $\mu = BACDE$. Then, $\mu \in F$, but μ is not essentially stable.*

Proof. 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 this does not return to i_3, the claim (i_3, B) is non-vacuous and so μ is not essentially stable.

Next, we show that if F is a fair set of matchings, then $\mu \in F$. Let $\pi(F)$ be the set of possible matchings for F. First, note that the DA outcome is $\mu^{DA} = ABCDE$, and $\mu^{DA} \in F$ for any F (because it is not blocked by anything). Next, observe that each student i has the highest priority at his DA school. So, i can use the DA matching to block any other matching ν that gives him a school he disprefers to his DA school. This implies that for all $\nu \in \pi(F)$, ν Pareto dominates μ^{DA}.[34]

Now, assume that $\mu = BACDE \notin F$. By part (2) of the definition of fairness, there exists some $\nu \in \pi(F)$ that blocks it. The only potential student who can block μ is i_3, who can block with B. Let ν be some $\nu \in \pi(F)$ at which $\nu(i_3) = B$. Since ν must Pareto dominate μ^{DA}, there is only one possibility: $\nu = ACBDE$.[35] Thus, $\nu = ACBDE \in \pi(F)$.

[34]If ν does not Pareto dominate μ^{DA}, then there is some i such that $\mu^{DA}(i) P_i \nu(i)$. Then, ν is not possible for F, because i would block ν using μ^{DA}, which is always included in any F.

[35]Since ν must Pareto dominate μ^{DA}, i_1 must get A (because i_3 is assigned B). Then,

Since $\nu \in \pi(F)$, there is no $\rho \in F$ 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 $\rho \notin F$; in particular, $\rho = ABDCE \notin F$.

Since $\rho \notin F$, there must be some $\sigma \in \pi(F)$ 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 their DA assignment,[36] which contradicts that every $\sigma \in \pi(F)$ Pareto dominates μ^{DA}. □

The above proposition shows that our definition is not equivalent to that of Morrill (2016). As far as the definition of weak stability from Tang and Zhang (2016), a result of Morrill (2016) shows that the set of fair matchings F is equivalent to the vNM stable set. Tang and Zhang (2016) show that every matching that is in the vNM stable set is weakly stable in their sense. Thus, the same matching μ from the above proposition is weakly stable in the sense of Tang and Zhang (2016), but is not essentially stable.

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 $\nu(i_5) = E$.

[36]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 their DA assignment.