Theoretical Economics 12 (2017), 863–908 1555-7561/20170863

Improving matching under hard distributional constraints

Daniel Fragiadakis
Department of Economics, Texas A&M University

Peter Troyan
Department of Economics, University of Virginia

Distributional constraints are important in many market design settings. Prominent examples include the minimum manning requirements at each Army branch in military cadet matching and diversity considerations in school choice, whereby school districts impose constraints on the demographic distribution of students at each school. Standard assignment mechanisms implemented in practice are unable to accommodate these constraints. This leads policymakers to resort to ad hoc solutions that eliminate blocks of seats ex ante (before agents submit their preferences) to ensure that all constraints are satisfied ex post (after the mechanism is run). We show that these current solutions ignore important information contained in the submitted preferences, resulting in avoidable inefficiency. We then introduce new *dynamic quotas mechanisms* that result in Pareto superior allocations while at the same time respecting all distributional constraints and satisfying important fairness and incentive properties. We expect the use of our mechanisms to improve the performance of matching markets with distributional constraints in the field.

Keywords. Minimum quotas, floors, ceilings, affirmative action, school choice, diversity, strategyproofness, deferred acceptance.

JEL classification. C78, D61, D63, I20.

1. Introduction

The theory of matching has been extensively developed and applied to solve a wide array of practical allocation problems in recent years. Important formal settings include

Daniel Fragiadakis: dfragiadakis@tamu.edu
Peter Troyan: troyan@virginia.edu

We are extremely grateful to Fuhito Kojima, Muriel Niederle, and Al Roth for numerous conversations regarding this project. We would like to thank Brian Baisa, Scott Baker, Andrey Fradkin, Guillaume Haeringer, Matt Jackson, Scott Kominers, Vikram Manjunath, Paul Milgrom, Krishna Rao, Itay Saporta-Eksten, Ilya Segal, Rodrigo Velez, Pablo Villanueva, Alex Westkamp, Alex Wolitzky, and Edison Yu for helpful comments and suggestions, as well as seminar participants at Virginia, Iowa, Cal State East Bay, Texas A&M, Western Ontario, and Illinois. We also thank a co-editor and an anonymous referee for helpful comments that improved both the results and the exposition of the paper. Generous support from the B. F. Haley and E. S. Shaw Fellowship (Troyan) and the Leonard W. Ely and Shirley R. Ely Fellowship (Fragiadakis), both through the Stanford Institute for Economic Policy Research, is gratefully acknowledged. This paper was originally circulated under the title "Market design under distributional constraints: Diversity in school choice and other applications."

DOI: 10.3982/TE2195

matching doctors to hospitals, elementary and high school students to school seats, and military cadets to Army branches; less formally, similar problems arise in many other areas, such as assigning groups of students to projects in a class or groups of employees to tasks in a firm. As the scope of applications has expanded, designers have increasingly encountered various types of practical constraints that were not present in the early analysis. In this paper, we study an important class of these constraints and show that an intuitive and common way in which they are handled results in avoidable inefficiencies. In addition, we propose new mechanisms, and show that they Pareto dominate current ad hoc approaches without compromising key fairness or incentive properties.

The constraints we consider are called *distributional constraints*, because they relate to the distribution of the final assignments of the agents across objects, projects, or institutions. As an example, consider medical residency matching, in which hospitals in rural areas commonly suffer doctor shortages relative to their urban counterparts.[1] Many governments are concerned about access to health care in rural communities, and try to implement policies to balance the distribution of doctors between urban and rural areas. One policy instituted by the Japanese government is to reduce the capacities of the urban hospitals to ensure that more doctors are assigned to rural hospitals. To produce the final assignment, they then run the very popular deferred acceptance mechanism (which is widely used in many medical residency markets, including in the United States) under these lower ("artificial") capacities, a mechanism we call *artificial caps deferred acceptance* (ACDA).[2]

As a second example of distributional constraints, consider the United States Military Academy (USMA), which every year must assign newly graduated cadets to positions in Army branches (for example, aviation or infantry). An October 1, 2007 memorandum from the Army Deputy Chief of Staff to the USMA titled "Branch Allocation Methodology" describes the following three phase procedure: first, minimum and maximum quotas for the number of cadets who must be assigned to each branch—or, "floors" and "ceilings," respectively—are calculated based on the Army's current staffing needs; second, artificial caps are calculated such that any final assignment that obeys these caps necessarily satisfies all of the *true* ceilings and floors as well; and third, the matching algorithm is run under the artificial (rather than the true) capacities.[3] The memo states that the assignment is done in three phases because "there is no ex-ante

[1]For example, Talbott (2007) notes that the United States as a whole has 280 doctors per 100,000 people, but the 18-county Mississippi delta area has only 103 doctors per 100,000 people. Similar doctor shortages in rural areas are present in many countries (e.g., Nambiar and Bavas 2010 document such shortages in Australia).

[2]More precisely, the Japanese government imposes *regional maximum quotas* that guarantees that no more than a certain number of doctors are assigned to each *region* of the country. In practice, however, this is implemented by simply reducing the individual capacities of the hospitals by some fixed percentage that guarantees that, when aggregated, the regional quotas will not be violated. See Kamada and Kojima (2015) for more details on this market.

[3]Sönmez and Switzer (2013) proposes new algorithms for the third (final) phase of the branching procedure, but does not consider the problem of how the capacities themselves are determined (second phase).

closed form algorithm that optimizes program participation *subject to manning requirements*" (emphasis added). Providing such algorithms and studying their properties are precisely the goals of this paper.[4]

A final well known example of distributional constraints arises in school choice mechanisms, which many cities have recently adopted to give parents more choice over schools (examples include New York, Boston, Chicago, and Denver). An additional consideration for many school districts when implementing school choice plans is achieving (demographic) diversity, which can be thought of as distributional constraints on the demographic distributions of the students at the schools. Usually, this is done using socioeconomic status (SES) or some proxy for it. For example, Chicago classifies students into four SES tiers and requires that all selective high schools enroll enough students from each tier.[5] To take another concrete example, consider Cambridge, MA, which divides students into high and low SES. They then require that the percentage of students at each school from each class must lie within a certain range; in other words, there is both a floor and a ceiling for the number of students of each type at each school.[6] Louisville, KY and White Plains, NY have very similar controlled choice plans.[7]

The common theme that ties all of these applications together is the presence of *floors*, i.e., a minimum number of agents who must be assigned to each institution. (In many cases, there are actually multiple floors at each institution, such as floors for each socioeconomic class in school choice.) While the literature thus far has been extremely successful at developing mechanisms that work very well in the field for markets with ceiling constraints, most of these existing mechanisms are inadequate for the many institutions that also have floor constraints. For instance, the original school choice mechanisms found in the seminal paper of Abdulkadiroğlu and Sönmez (2003) would allow a school to have (for example) a total capacity of 100 seats in addition to ceiling constraints of at most 50 high SES students and at most 50 low SES students. Note, however, that an assignment of 50 high SES students would satisfy these constraints, yet would be completely segregated, and thus would not satisfy the true diversity objectives of many school districts.

[4]Military branching can also be thought of as a special case of a firm (the Army) that must assign employees to projects (the branches), with each project having a minimum staffing requirement. For example, some technology firms in Silicon Valley use centralized mechanisms to assign new interns to positions. In a similar vein, for many new medical residents, the first year after medical school is a *transitional year* in which they rotate through various departments of a hospital. While the hospitals try to accommodate preferences as much as possible, each department has a minimum staffing requirement. All of these problems can also be modeled as a matching problem with floor constraints, and our mechanisms can be applied.

[5]See the website of the Chicago Public Schools Office of Access and Enrollment at http://www.cpsoae.org/apps/news/show_news.jsp?REC_ID=184188.

[6]Each school is required to be within 10% of the districtwide average for each SES class, which translates into ranges of approximately 25–45% for low SES students and 55–75% for high SES students (these numbers may vary slightly from year to year). See "Case studies of school choice and open enrollment in four cities" (Cowen Institute for Public Education Initiatives 2011).

[7]Achieving socioeconomic diversity in schools is also an issue in many European countries. For example, Sweden implements a school choice procedure, but is considering "restricting the ability of some parents to choose their children's schools by introducing 'controlled choice schemes that supplement parental choice to ensure a more diverse distribution of students in schools'" (Orange 2015). See http://www.matching-in-practice.eu for more examples.

Suppose the school in the above example also imposed floors of 25 high and 25 low SES students (in addition to the ceilings of 50). This ensures a minimum level of diversity at the school. One convenient way the district can ensure that these floors are met is to (i) lower the ceilings at other schools and then (ii) run an "off-the-shelf" mechanism designed to handle only ceiling constraints. As mentioned above, we call this approach *artificial caps* (and when the off-the-shelf mechanism is the deferred acceptance (DA) mechanism of Gale and Shapley (1962), we obtain the artificial caps deferred acceptance (ACDA) mechanism). It works by the simple intuitive principle that restricting someone from one school results in their being assigned to another.

Artificial caps is in fact a commonly used approach, likely because it is so intuitive. As described above, this is precisely how the Japan Residency Matching Program (JRMP) guarantees enough doctors are assigned to rural hospitals, and how the USMA ensures enough cadets are assigned to each Army branch.[8] They first eliminate sufficiently many positions ex ante (i.e., before preferences are submitted) to guarantee that all of desired floors are satisfied ex post (i.e., when the final matching is reached).

In this paper, we first show that imposing artificial caps results in avoidable inefficiency. The reason is that, to satisfy all of the constraints, ACDA must eliminate positions aggressively. To understand why, note that in a given matching problem, only one set of preferences, P, is submitted. It may be the case that the ceilings required for DA to satisfy all floors under P are not as low as those that were imposed by ACDA (which must be low enough to ensure that the floors are met ex post for *any* possible preferences that *could* have been submitted). This is problematic, since eliminating a seat at a school makes every student weakly worse off. Accordingly, we introduce the idea of a dynamic quotas (DQ) mechanism, where *dynamic quotas deferred acceptance* (DQDA) runs as follows: Start with the original ceilings and run DA. If the matching is feasible (i.e., meets all floors), end the algorithm. Otherwise, lower the ceiling at one school by 1. This causes a *rejection chain*, where that school rejects a student, who then applies to her next most preferred school, which then (may) reject a student, and so forth, until a student applies to a school with an open seat. When this rejection chain ends, if the matching is feasible, end the algorithm; if not, lower the ceiling of another school by 1, and so forth. We show that if we choose the order in which ceilings are lowered carefully, DQDA (i) always produces a matching that satisfies all distributional constraints and (ii) Pareto dominates artificial caps DA.

By construction, the final ceilings implemented depend on the submitted preferences. This is precisely how we obtain the efficiency gain over ACDA: by using information contained in the submitted preferences to determine the final ceilings, we are

[8]The case of the Japan Residency Matching Program is slightly subtle because, while they do not have *explicit* floors, it seems clear that the end goal of the artificial capacities they impose is not to limit the number of doctors in urban areas *per se*; instead, artificial caps are likely used as a simple, ad hoc way to increase the number of doctors in rural areas. Kamada and Kojima (2015) provide improvements to the current JRMP mechanism if the regional maximum quotas are taken at face value as the true objective, but their mechanisms cannot handle floors explicitly. If the true objectives are actually as-to-now implicit floor constraints, then we argue it would be an improvement to model these constraints explicitly and use the mechanisms provided in this paper. Doing so satisfies the actual distributional goals (the true floors and ceilings), and, as we show, makes all doctors better off, compared to the current approach of imposing regional caps and hoping that the resulting distribution of doctors turns out to be satisfactory.

able to eliminate fewer seats at each school than artificial caps. A concern that arises from this modification is that individuals now have the ability to change the ceilings by submitting different preferences. This poses a potential incentive problem: perhaps an individual can do better by misreporting her preferences than by stating them truthfully, i.e., the mechanism may not be *strategyproof*. Non-strategyproof mechanisms are often unattractive because they allow some agents to "game the system" and profit at the expense of less sophisticated players (this is especially true in school choice, where school districts are often worried that some parents may not fully understand the mechanism and will be harmed by not strategizing appropriately).[9] Thus, before implementing a dynamic quotas mechanism, it is crucial to fully understand its incentive properties. While it may seem that allowing the final ceilings to depend on the preferences introduces obvious avenues for manipulation, one of our main results is to show that this is not the case. Indeed, we show that if the algorithm is constructed carefully, DQDA is in fact strategyproof.

A final key concern is how to determine who is assigned where when demand exceeds supply. In school choice, many school districts create priority lists for each school, giving students higher priority for neighborhood schools, sibling attendance, languages spoken, or various other factors. The districts then wish to respect priorities in the following sense: if student i is assigned to school A but prefers the assignment of a student j (e.g., school B) who is of the same socioeconomic status, then j must have a higher priority at school B than i. Thus, i's envy toward j is not justified. Eliminating such justified envy concerns is often an important fairness consideration in other settings as well. For example, in the military, cadets are ranked according to an Order of Merit list, which combines academic, physical fitness, and leadership scores (among others). To give good incentives for the cadets to achieve high scores, the Army desires that higher ranking cadets be assigned their more preferred branches, i.e., they also want to eliminate justified envy. We show that DQDA does indeed satisfy this property.

We would also like to note that the paper makes not only a practical contribution, but a theoretical one as well. While the incredibly influential paper of Gale and Shapley (1962) first introduced the now very popular deferred acceptance (DA) mechanism, it did not study the mechanism's incentive properties. It has since come to be widely accepted that good incentive properties are integral to success, not only in matching, but in a broad array of economic contexts. With regard to DA, Dubins and Freedman (1981) and Roth (1982) were the first to prove that it is strategyproof for the proposing side in simple one-to-one matching models. As the potential applications of DA have rapidly grown, an active topic of research in the literature has been trying to understand the most general settings that are compatible with strategyproofness (Martínez et al. 2000, Abdulkadiroğlu 2005, Hatfield and Milgrom 2005, Hatfield and Kominers 2011, 2012, Hatfield and Kojima 2010). The presence of the floors makes our model quite different from these papers in a technical sense. In particular, they often rely on the existence of

[9]Non-strategyproof mechanisms also make it much more difficult for participants to reach equilibrium in practice, and so it is unclear whether real-world outcomes will conform to theoretical predictions of such mechanisms. Additionally, the need to strategize creates additional costs to participants from participating in the mechanism that are not present if it is in their best interest to simply report their true preferences.

student-optimal stable matchings, a condition that fails in our setting. In addition, all of the previous papers assume that the choice functions of the receiving side are fixed throughout the algorithm. We are the first to show that dynamically adapting choice functions are compatible with strategyproofness.[10] In the Appendix, we study general sufficient conditions on the evolution of the choice functions that guarantee good properties of DA. Interestingly, our conditions are related to the key substitutability conditions in the model of Hatfield and Milgrom (2005) (see also Kelso and Crawford 1982 and Roth and Sotomayor 1990). Both guarantee *monotonicity* of the corresponding cumulative offer process, which is crucial in proving our main results. Besides being useful in practice, these results deepen our understanding of the enormously successful deferred acceptance mechanism.

Related literature

Early papers that discussed distributional constraints in matching focused on the rural hospital problem and obtained mostly negative results. Papers such as Gale and Sotomayor (1985a, 1985b), Roth (1984, 1986), Martínez et al. (2000), and Hatfield and Milgrom (2005) prove various versions of the "rural hospital theorem," which says that if a doctor or a position at a hospital is unmatched at some stable matching, then they are unmatched at any stable matching.[11] This suggests that the rural hospital problem is difficult to solve without imposing any additional structure on the market, which is what led the Japan Residency Matching Program (JRMP) to impose regional caps on the number of doctors in urban areas, an issue studied in detail by Kamada and Kojima (2015).

Since it is an important goal of many school districts, diversity constraints have been much discussed in the school choice literature. Much of the work thus far has dealt with upper quotas/ceilings. In their seminal paper on school choice from a mechanism design perspective, Abdulkadiroğlu and Sönmez (2003) show how type-specific ceilings can be easily incorporated into standard matching mechanisms. Ceiling constraints do not fully capture diversity constraints, however, since they can still result in completely segregated schools. In addition, in a model with two types of students (majority and minority), Kojima (2012) points out that simple ceiling constraints can actually make all minority students (the supposed beneficiaries) worse off. Hafalir et al. (2013) correct this by proposing deferred acceptance with minority reserves, a mechanism further generalized by Kominers and Sönmez (2016), who introduce slot-specific priorities.

[10]Our use of the word "dynamic" should not be confused with the strand of literature on "dynamic matching markets" that introduces a time dimension and allows agents to enter/exit the market and matches to change each period (as in Kennes et al. 2014a, 2014b, Kadam and Kotowski 2014, and Akbarpour et al. 2016). In our model, agents play a static game where they submit their preferences once, at the start of the game, and take no further action. Once lists are submitted, the quotas/choice functions of one side of the market may change as the algorithm progresses, hence operating in a possibly dynamic manner, but ultimately, a single final match is produced.

[11]Afacan (2013) studies whether hospitals can manipulate their preferences to change the number of positions filled. Sönmez (1997) studies the complementary question of whether hospitals can manipulate their capacities to obtain a more preferred assignment of doctors.

Doğan (2016) notes that in the mechanism of Hafalir et al. (2013), stronger affirmative action constraints may actually harm some minority students without helping others, and proposes a modification to rectify this. Abdulkadiroğlu (2005), Erdil and Kumano (2013), Aygün and Bó (2016), and Echenique and Yenmez (2015) study various generalizations of school priorities over sets of students and how they can capture certain types of diversity goals.[12]

The main difference between our model and the aforementioned works is that we treat the ceilings and floors as *hard* constraints, i.e., constraints that must be satisfied at any feasible matching. Most prior papers interpret floors as *soft* constraints; that is, the constraints are more like "guidelines," and they may end up being violated at the final assignment. Working with hard constraints complicates the problem considerably, and leads to the incompatibility of several important properties (non-wastefulness, elimination of justified envy and strategyproofness) that could be achieved simultaneously in prior models.

Recently, there is a growing literature that has begun to deal with hard floor constraints.[13] Ehlers et al. (2014) start by looking at a school choice model with hard floors and ceilings similar to that studied here. After noting the above impossibility results, they advocate for a "soft" interpretation of the constraints where the floors and ceilings can be violated, and provide new mechanisms in this context.[14] While such a soft bounds approach may be acceptable in certain settings, there are many situations in which it is inadequate, such as medical markets suffering from the rural hospital problem, school districts with court-mandated desegregation guidelines, or the military, where minimum manning requirements must be filled. Our paper provides a strategyproof mechanism that ensures that all floors are filled, and so can be used in these settings. Fragiadakis et al. (2015) also study a model with hard floor constraints. The model here is more general, as the model in Fragiadakis et al. (2015) is restricted to the case of aggregate floor constraints; for example, they would not allow a school to have separate floors for different types of students. We must construct different solutions in this paper so as to handle multiple types (diversity constraints), and the arguments become more complex. In addition, we show that our mechanisms Pareto dominate mechanisms that are used in practice. In a more recent paper, Kojima et al. (2014) analyze certain classes of distributional constraints using the tools of discrete convex analysis. While they are able to encompass many types of constraints (including the simpler mechanisms of Fragiadakis et al. 2015), they note that they are unable to accommodate the constraints we consider, because of the complexities introduced by type-specific floors and ceilings. Allowing for more general constraint structures that accommodate such goals introduces substantial complications in constructing appropriate mechanisms and in proving important incentive and efficiency results. Given the importance

[12] See also Westkamp (2013), who proposes similar mechanisms in the context of German university admissions, and Braun et al. (2014), who conduct an experimental analysis of these mechanisms.

[13] In the context of object allocation, Budish et al. (2013) study what types of constraints admit expected assignments that can be implemented as lotteries over deterministic assignments.

[14] They also provide an algorithm for the hard constraints case but, unlike ours, their mechanism is not strategyproof.

of these types of constraints in real-world markets, we view this as one of the key contributions of the current paper.

Finally, the problem of distributional constraints has also garnered interest from computer scientists. For example, Biró et al. (2010) study college admissions in Hungary, in which colleges are allowed to declare minimum quotas for their programs, and Hamada et al. (2011) study hospital–resident matching with lower bounds. Both papers focus mainly on computational hardness results: the former shows that the problem of determining the existence of a stable matching is NP-complete, while the latter shows that the same is true of finding a matching that minimizes the number of blocking pairs. These papers provide another perspective which says that introducing floors into matching markets complicates the problem substantially.

2. Model

There is a set of agents $I = \{i_1, \dots, i_n\}$ and a set of objects to which they can be assigned, $S = \{s_1, \dots, s_m\}$, each of which has total capacity Q_s. The finite set of types for agents is $\Theta = \{\theta_1, \dots, \theta_r\}$, and each agent is of exactly one type. The function $\tau : I \to \Theta$ gives the type of each agent, and I_θ is the set of agents of type θ. Types are fixed and are publicly observable (i.e., types cannot be misreported). In addition to a capacity Q_s, each $s \in S$ has a type-specific *floor* $L_{s,\theta}$ (or lower quota) and a type-specific *ceiling* $U_{s,\theta}$ (or upper quota) for the number of agents of each type θ who can be assigned to it. We assume $0 \le L_{s,\theta} \le U_{s,\theta} \le Q_s$ for all (s, θ). Let $Q = (Q_s)_{s \in S}$ be the vector of all capacities and let $L = (L_{s,\theta})_{s \in S, \theta \in \Theta}$ and $U = (U_{s,\theta})_{s \in S, \theta \in \Theta}$ be the matrices of all type-specific floors and ceilings, respectively.

One application of the model is that I is a set of students and S is a set of schools. Then Θ can be interpreted as a set of socioeconomic classes corresponding to the diversity constraints of the school district.[15] Other potential applications include the military assigning cadets to branches, residency programs assigning doctors to hospitals (as in Japan), firms assigning workers to tasks, or business schools assigning students to projects.[16] For concreteness, from here on we mostly stick to the language of students

[15] School districts sometimes state diversity constraints in terms of percentages (rather than absolute numbers) because they are simpler to communicate, but these percentages are usually converted into absolute numbers of seats when actually running the algorithm. This is done in Cambridge, for example (see http://www3.cpsd.us/video/controlled_choice_video for a video describing the implementation of the Cambridge algorithm, targeted toward parents). From a technical perspective, the use of percentages introduces complementarities, which leads to many impossibility results (see Echenique and Yenmez 2015), and percentages likely do not truly capture a school district's goals, since they also want the total population of a school to lie in some range (for example, a school with one high SES student and one low SES student would be "unsegregated," but having so few students in a school would not be cost effective from the school district's perspective). For these reasons, most formal models use absolute numbers (e.g., Ehlers et al. 2014 and Hafalir et al. 2013). When there is only a single type (e.g., in military cadet matching), this distinction is immaterial.

[16] In fact, the true prevalence of these types of distributional constraints in the real world is likely to be underestimated. This is because artificial caps may be set internally in such a way that they are unobserved to outsiders. That is, before running an algorithm, capacities may be set so as to implicitly satisfy some floors (as in the JRMP in Japan), but to an outside analyst, it would appear as if the artificial caps were

and schools, since this is a framework with which most readers are likely familiar and so will ease the exposition.

Each student i has a strict *preference relation* P_i over S, while each school s has a strict *priority relation* $\succ_s$ over I. Profiles of such relations, one for each agent, are denoted $P_I = (P_i)_{i\in I}$ and $\succ_S = (\succ_s)_{s\in S}$. Let $\mathcal{P}$ denote the set of all individual preference relations, and let $\mathcal{P}^n$ denote the set of all preference profiles P_I. The student preferences are their own private information. The priorities are fixed and known to all students. In school choice applications, priorities are often set by law, and depend on such things as distance from a school, whether a student has a sibling attending the school, or whether a student speaks a certain language. In the context of the military, priorities are determined by combinations of academic, physical fitness, and leadership scores, among other things.

A *matching* is a correspondence $\mu : I \cup S \to I \cup S$ that describes which students are assigned to which schools. Formally, μ must satisfy (i) $\mu(i) \in S$ for all $i \in I$, (ii) $\mu(s) \subseteq I$ for all $s \in S$, and (iii) $\mu(i) = s$ if and only if $i \in \mu(s)$. Let $\mathcal{M}$ denote the set of matchings. For any $\mu \in \mathcal{M}$, we let $\mu_\theta(s)$ be the set of type θ students assigned to school s under matching μ. Matching μ is *feasible* if $L_{s,\theta} \le |\mu_\theta(s)| \le U_{s,\theta}$ and $|\mu(s)| \le Q_s$ for all (s, θ). In words, a feasible matching is one that satisfies all of the type-specific floors and ceilings as well as the overall capacities. Let $\mathcal{M}_f \subseteq \mathcal{M}$ denote the set of feasible matchings. We assume throughout the paper that $\mathcal{M}_f \neq \varnothing$; this is the (obviously necessary) requirement that the distributional constraints be consistent with the number of students of each type actually present in the market.

A *mechanism* $\psi : \mathcal{P}^n \to \mathcal{M}$ is a function that maps preference profiles to matchings. If the students submit $P_I \in \mathcal{P}^n$, then $\psi(P_I) \in \mathcal{M}$ is the resulting matching. We write $\psi_i(P_I)$ for the school to which student i is assigned, and write $\psi_s(P_I)$ for the set of students assigned to school s. We say that ψ is feasible if $\psi(P_I) \in \mathcal{M}_f$ for all $P_I \in \mathcal{P}^n$.

Given two matchings $\mu, \nu \in \mathcal{M}_f$, μ *weakly Pareto dominates* ν if $\mu(i)R_i\nu(i)$ for all $i \in I$; if, in addition, $\mu(i)P_i\nu(i)$ for some $i \in I$, then we say μ *Pareto dominates* ν.[17] If $\mu \in \mathcal{M}_f$ is not Pareto dominated by any other $\nu \in \mathcal{M}_f$, then we say that μ is *Pareto efficient*.[18]

We say student i of type θ *claims an empty seat at school* s if (i) $sP_i\mu(i)$, (ii) $|\mu(s)| < Q_s$ and $|\mu_\theta(s)| < U_{s,\theta}$, and (iii) $|\mu_\theta(\mu(i))| > L_{\mu(i),\theta}$. In words, a student claims an empty seat if there is a school s she prefers that has an open seat for her type *and* leaving her current school $\mu(i)$ would not violate feasibility by causing it to drop below one of its floors. If no student claims an empty seat under matching μ, then μ is *non-wasteful*.

the "true" ceilings and there were no floors. The JRMP and military cadet matching are two cases where artificial caps are quite explicitly used, but since the idea is an intuitive one that seems to naturally occur to policymakers, it is not hard to imagine other cases in which they are imposed discreetly. Thus, we believe that our mechanisms may be applicable to many markets that have up to now dealt with floors in such a suboptimal manner simply for lack of a better option.

[17]We use R_i to denote the weak preference relation corresponding to P_i, i.e., $s\, R_i\, s'$ if and only if $s\, P_i\, s'$ or $s = s'$.

[18]Only the welfare of the students is considered, which is consistent with the school choice mechanism design literature in which school seats are viewed as objects to be consumed by students (see Abdulkadiroğlu and Sönmez 2003).

In words, non-wastefulness means that whenever a student prefers a school s to her current assignment, it is impossible to move her to s without violating feasibility.

A second property is a fairness requirement called elimination of justified envy.[19] Student $i \in \mu(s)$ *justifiably envies student* $i' \in \mu(s')$ if (i) $s' \, P_i \, s$, (ii) $i \succ_{s'} i'$, and (iii) there exists an alternative matching $\nu \in \mathcal{M}_f$ such that $\nu(i) = s'$, $\nu(i') \neq s'$, and $\nu(j) = \mu(j)$ for all $j \neq i, i'$. If no student justifiably envies any other, then the matching *eliminates justified envy*. In words, student i justifiably envies i' if she prefers the school of student i', has higher priority than i' at this school, and i and i' can be reassigned without violating any distributional constraints (and without altering the allocation of any other student).[20]

The above properties have counterparts for mechanisms. Mechanism ψ is non-wasteful if $\psi(P_I)$ is a non-wasteful matching for all $P_I \in \mathcal{P}^n$, and ψ eliminates justified envy if $\psi(P_I)$ is a matching that eliminates justified envy for all $P_I \in \mathcal{P}^n$. We say that the *mechanism* ψ *Pareto dominate mechanism* φ if

$$\text{for all } P_I, \quad \psi_i(P_I) \, R_i \, \varphi_i(P_I) \quad \text{for all } i \in I,$$

$$\text{for some } P_I, \quad \psi_i(P_I) \, P_i \, \varphi_i(P_i) \quad \text{for some } i \in I.$$

As for matchings, we say that ψ *weakly Pareto dominates* φ whenever the first part holds.

Since the student preferences are their own private information, the last important property we must discuss is the incentives for students to report these preferences truthfully to a mechanism. Mechanism ψ is *strategyproof* if $\psi_i(P_I) \, R_i \, \psi_i(P'_i, P_{-i})$ for all $i \in I$, $P_I \in \mathcal{P}^n$, and $P'_i \in \mathcal{P}$. In words, a mechanism is strategyproof if no student can ever gain by misreporting her preferences, no matter what the other students report.

Strategyproofness is a strong form of incentive compatibility, and is viewed as an important property for many reasons. First, strategyproof mechanisms advance the so-called Wilson doctrine (Wilson 1987), which argues that to be successful, market designs should not be sensitive to specific assumptions on agent beliefs (see also Bergemann and Morris 2005). Strategyproof mechanisms satisfy the Wilson doctrine in its strongest sense, since truthful reporting is optimal for any beliefs agents may have. Second, from a practical perspective, policymakers in general (and school districts in particular) are often interested in strategyproof mechanisms because they are strategically simple for agents to play. Agents can be informed that all they must do is submit their true preferences, and unsophisticated players who are unable to strategize effectively will not be

[19]For example, this was an important criterion to administrators of the Boston school district when they were redesigning their school assignment mechanism (Abdulkadiroğlu et al. 2005). In the military, elimination of justified envy is an important normative criterion that ensures that higher performing cadets receive their more preferred assignments.

[20]In two-sided matching models without distributional constraints, non-wastefulness and elimination of justified envy are often combined into one definition called *stability*, which is usually then given a positive interpretation. We must separate the two definitions due to impossibility results caused by the introduction of the floors (discussed below). In addition, in many school choice settings, these properties are more usefully interpreted in a normative manner (see also Kamada and Kojima 2015, who use normative justifications for alternative stability concepts in hospital residency matching in Japan, where the standard (positive) notion of stability fails). Balinski and Sönmez (1999), Ehlers et al. (2014), and Fragiadakis et al. (2015) use similar distinctions between non-wastefulness and elimination of justified envy as we do here.

disadvantaged.[21] For these reasons, many cities have opted for school choice mechanisms that are strategyproof (among them, New York City, Boston, and New Orleans). Strong incentive constraints have been an important design consideration in other market design settings as well, such as hospital–resident matching (Roth 1991, Roth and Peranson 1999) and auction design.[22]

3. Deferred acceptance and artificial caps

When there are no floor constraints, a widely used solution to the assignment problem is to use some variation of the deferred acceptance (DA) algorithm of Gale and Shapley (1962). The following description is a simple generalization of Gale and Shapley's algorithm with ceiling constraints and type-specific reserves.

Deferred acceptance

Step 1. Each student applies to the first school on her preference list. Each school s considers all students who have applied to it and *tentatively* accepts students as follows:

(i) Type-specific seats: For each type θ, school s accepts the $L_{s,\theta}$ highest ranked type θ students according to $\succ_s$.

(ii) Open seats: For any students remaining in the applicant pool, school s admits students one by one from the top of its priority order, unless either some type-specific ceiling $U_{s,\theta}$ would be violated or $Q_s - \sum_{\theta\in\Theta} L_{s,\theta}$ open seats have already been filled. All students not accepted are rejected.

Step k. Each student who was rejected in step $k-1$ applies to her most preferred school that has not yet rejected her. Each school s considers its new applicants in step k *jointly* with the students tentatively admitted from step $k-1$, and again tentatively accepts students in its applicant pool in the same manner as above. All students not accepted are rejected.

In this version of DA, each school reserves $L_{s,\theta}$ seats exclusively for students of type θ; the remaining $Q_s - \sum_{\theta\in\Theta} L_{s,\theta}$ seats are open seats, that can go to students of any type, subject to the ceiling constraints $U_{s,\theta}$. When the floors are set to 0 at all schools, the above algorithm reduces to that defined by Abdulkadiroğlu and Sönmez (2003), and when there is only a single type, it reduces further to the algorithm of Gale and Shapley (1962). In these simpler environments, DA is a very successful mechanism because it is non-wasteful, eliminates justified envy, and is strategyproof.

[21] See Pathak and Sönmez (2008, 2013).

[22] Strategyproofness is of course not costless, as is shown in a recent strand of the school choice literature started by Abdulkadiroğlu et al. (2011) that finds that non-strategyproof mechanisms may sometimes outperform strategyproof ones on welfare grounds, at least in equilibrium (see also Featherstone and Niederle 2014, Troyan 2012, and Akyol 2013). However, the equilibria of these mechanisms can be complex, and it may be difficult for many agents to calculate best responses, which is what has often lead to the adoption of strategyproof alternatives.

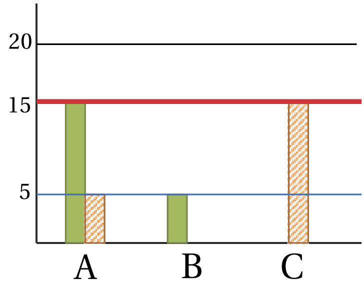

Figure 1. Outcome of DA with ceilings of 15 for each type (represented by the thick horizontal line). The solid bars represent the number of type h students at each school, while the striped bars represent the number of type ℓ students at each school.

However, while in the above mechanism the schools *reserve* $L_{s,\theta}$ students for each type θ, these seats may not actually be filled at the end of the algorithm. Thus, under a model of hard floor constraints, DA may produce a matching that is not feasible. This is shown in the following example.

Example 1. Let there be three schools A, B, and C, each with a capacity of 20 seats. There are 40 students, divided into two types: students $h_1, \dots, h_{20}$ are of high socioeconomic status (type h), and students $\ell_1, \dots, \ell_{20}$ are of low socioeconomic status (type ℓ). The distributional constraints are to have between 5–15 students of each type at each school. Let the preferences of all of the h students be P_{h_i}: A, B, C and let the preferences of all of the ℓ students be P_{ℓ_i}: A, C, B. For simplicity, let the priorities of all schools rank $h_1 \succ_s h_2 \cdots \succ_s h_{20} \succ_s \ell_1 \cdots \succ_s \ell_{20}$ (this can easily be generalized). DA using the true ceilings of 15 for each type at each school produces the output shown in Figure 1.

The problem is that schools B and C are not meeting their floors. A simple and oft used solution to this problem is to run DA under some lower ceilings, a mechanism we call artificial caps deferred acceptance (ACDA). However, to guarantee that all of the floors will be filled for *any* possible submitted preference profile, the artificial caps must be quite strict. For example, consider running DA on the same preferences as above, but set the ceilings for each type at each school to be 8 (left panel of Figure 2). The figure shows that under the above preferences, even artificial caps of 8 still leave some floors unfilled.

It is indeed possible to set the artificial caps in such a way that all floors will be filled, no matter what preferences are submitted: for example, setting ceilings of 7 do just this (right panel of Figure 2). The benefit of setting ceilings of 7 is that we can guarantee that the final outcome will be feasible for any preference profile that is submitted. The cost is that, for some preference profiles, such low ceilings may be unnecessarily stringent and, hence, very wasteful. For example, consider an alternative preference profile in which 10 type h students and 10 type ℓ students have A as their first choice, 5 type h students and 5 type ℓ students have B as their first choice, and 5 type h students and 5 type ℓ students have C as their first choice. An assignment that gave every student her first choice would be feasible, but would violate the artificial caps of 7. Thus, if such

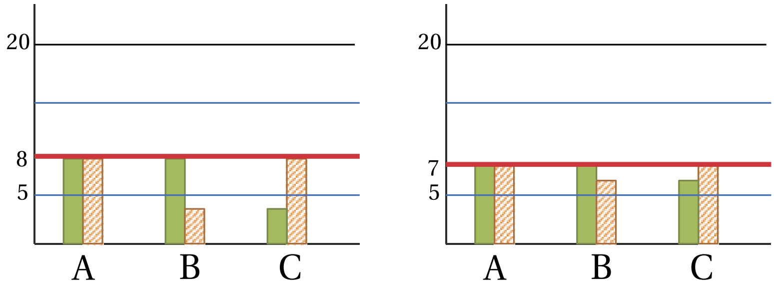

Figure 2. Left panel: Artificial caps of 8 at each school leave schools B and C below their floors. Right panel: Artificial caps of 7 ensure that all lower and upper quotas are satisfied. Again, the solid bars represent the type h students and the striped bars represent the type ℓ students.

rigid artificial caps were used, the resulting assignment will be inefficient. It is these inefficiencies we hope to recover by designing a new mechanism. ◊

To define artificial caps DA formally, let $\bar{U} = (\bar{U}_{s,\theta})_{s\in S,\theta\in\Theta}$ be some alternative type-specific ceilings and let $\bar{Q} = (\bar{Q}_s)_{s\in S}$ be some alternative total capacities that may be different from the primitive U and Q. We call any such $(\bar{U}, \bar{Q})$ *artificial caps*. We then formally define the *artificial caps deferred acceptance algorithm* (ACDA) as the deferred acceptance algorithm using some $(\bar{U}, \bar{Q})$, not necessarily equal to the primitive (U, Q). We denote the artificial caps deferred acceptance mechanism under $(\bar{U}, \bar{Q})$ by $\mathrm{DA}^{(\bar{U},\bar{Q})}(\cdot)$.

Under ACDA, the chosen $(\bar{U}, \bar{Q})$ will surely be satisfied (by definition), but if $(\bar{U}, \bar{Q})$ are not picked carefully, the final outcome of ACDA may not be feasible, either because some floors are not filled or because some students are unmatched. If, for any submitted preference profile P_I, running ACDA under $(\bar{U}, \bar{Q})$ produces a final matching such that all students are assigned and all type-specific floors, type-specific ceilings, and overall capacities are satisfied, then we say that $(\bar{U}, \bar{Q})$ *ensures a feasible match*. The following theorem shows that such feasibility-ensuring choices of $(\bar{U}, \bar{Q})$ always exist.

Theorem 1. *The set of $(\bar{U}, \bar{Q})$ that ensures a feasible match is nonempty.*

The proof chooses some feasible $\mu \in \mathcal{M}_f$ and sets $\bar{U}_{s,\theta} = |\mu_\theta(s)|$ and $\bar{Q}_s = |\mu(s)|$ for all (s, θ). This corresponds to predetermining exactly the number of students of each type θ who will be assigned to each school before students even submit their preferences.[23] While the proof only provides one example, there are in general many choices of $(\bar{U}, \bar{Q})$ that ensure a feasible match (the exact details depend on the specifics of the market in question). Henceforth, we assume that ACDA is run under some $(\bar{U}, \bar{Q})$ that ensures a feasible match.

Properties of ACDA

Without floors, it is well known that DA satisfies non-wastefulness, elimination of justified envy, and strategyproofness (Gale and Shapley 1962, Abdulkadiroğlu and Sönmez

[23]This of course implicitly depends on the assumption that the set of feasible matchings itself is nonempty, an assumption that was discussed above.

2003). In the presence of floors, however, an impossibility result obtains: matchings that eliminate justified envy may not even exist (Ehlers et al. 2014). This is intuitive, since floors are often used to give certain groups access to schools they would not be able to obtain based on priority alone. This observation leads to a natural alternative fairness criterion: a matching/mechanism *eliminates justified envy among same types* if no student justifiably envies another student of her same type. This is a reasonable criterion, because any remaining priority violations are caused by the distributional constraints, which the market organizer finds inherently valuable.

Theorem 2. *ACDA eliminates justified envy among same types and is strategyproof.*

The strategyproofness and (same type) envy-freeness of ACDA are immediately inherited from the fact that DA is strategyproof and eliminates justified envy among same types. ACDA is likely a popular mechanism because it satisfies these two properties, while at the same time filling all floors and, crucially, being very easy to implement. However, ACDA does have one significant drawback: the potential inefficiencies that arise from setting rigid capacities that are too low (recall Example 1 or see Example 2 below). In the next section, we introduce our dynamic quotas DA mechanism to recover these inefficiencies.

4. Dynamic quotas deferred acceptance

The major problem with ACDA is that its capacities are set once and for all from the beginning of the mechanism, and have no ability to respond to information that is contained in the submitted preferences. As we saw in Example 1 above, this can be important, because while for some preference profiles, very low ceilings may be necessary to ensure all floors are satisfied, for others, it may not be necessary to lower the ceilings at all. This idea is what motivates our new dynamic quotas mechanism, which uses information contained in the submitted preferences to endogenously determine the final ceilings and capacities in the course of running the algorithm and in so doing, prevents too many seats from being eliminated.

One nice feature of using rigid artificial caps is that ACDA immediately (and trivially) inherits the key strategyproofness property of DA. However, now that the submitted preferences can alter the final capacities, the main issue we need to be concerned with is strategyproofness: if a student's submission can alter the final capacities in such a way as to confer a strategic advantage from not reporting truthfully, one of the key properties that is important to the success of DA (and ACDA) will be lost. We will see that by designing the algorithm carefully, it is possible to retain strategyproofness.

To define our new algorithm, we first introduce the concept of a reduction sequence.

Definition 1. A *reduction sequence* is a sequence $\eta = \{(U^1, Q^1), (U^2, Q^2), \dots, (U^K, Q^K)\}$ of ceiling–capacity pairs that satisfies the following statements:

(i) For all k, $(U^{k+1}, Q^{k+1}) \leq (U^k, Q^k) \leq (U, Q)$.

(ii) The set (U^K, Q^K) ensures a feasible matching.

DEFINITION 2. A reduction sequence is *minimal* if the following statements hold for all k:

(i) For one (s, θ), $U^{k+1}_{s,\theta} = U^k_{s,\theta} - 1$ and $Q^{k+1}_s = Q^k_s - 1$.

(ii) For all $(s', \theta') \neq (s, \theta)$, $U^{k+1}_{s',\theta'} = U^k_{s',\theta'}$.

(iii) For all $s'' \neq s$, $Q^{k+1}_{s''} = Q^k_{s''}$.

In words, a reduction sequence is simply a monotonically decreasing sequence of ceiling–capacity pairs. Minimality means that in moving from stage k to $k+1$, we choose a school-type pair (s, θ) and lower the type-θ ceiling at s and the capacity of s by exactly one seat; the ceilings and capacities of the remaining schools are unchanged. To avoid trivialities, we assume that for any feasibility-ensuring artificial caps (U^K, Q^K), such a nondegenerate reduction sequence exists.[24]

Dynamic quotas deferred acceptance

Let $\eta = \{(U^1, Q^1), (U^2, Q^2), \dots, (U^K, Q^K)\}$ be a reduction sequence.

Stage 1. Compute the outcome of standard DA (as defined in Section 2) under (U^1, Q^1). If the resulting matching is feasible, end the algorithm and output this matching. If not, proceed to stage 2.

Stage k for $k \geq 2$.

k.0. Lower the ceilings and capacities to (U^k, Q^k). Divide each school s into $L_{s,\theta}$ type-specific seats for each type θ and $Q^k_s - \sum_\theta L_{s,\theta}$ open seats that can be assigned to any type.

k.1. Beginning with an applicant pool equal to the set of students held at the end of stage $k-1$, school s tentatively fills the type-specific seats for each type θ with the $L_{s,\theta}$ highest ranked type θ students according to $\succ_s$. If there are students remaining in the applicant pool, school s tentatively admits students one by one from the top of its priority order to the open seats, unless either some stage k type-specific ceiling $U^k_{s,\theta}$ would be violated or $Q^k_s - \sum_\theta L_{s,\theta}$ open seats have already been filled. All students not accepted are rejected.

[24]We say a reduction sequence is *degenerate* if for all $k < K$ and all P_I, $\mathrm{DA}^{(U^k,Q^k)}(P_I)$ is feasible only if $\mathrm{DA}^{(U^k,Q^k)}(P_I) = \mathrm{DA}^{(U^K,Q^K)}(P_I)$. For an example of a market with only degenerate reduction sequences, consider the special case where the sum of the floor constraints is exactly equal to the total number of students. When this is true, the ceiling constraints are effectively irrelevant because the floors alone exactly pin down the distribution of students across schools. Because there is no flexibility in the final distribution, the problem is trivial. Our model is only interesting when the primitives allow for some flexibility, which is indeed the case in many real-world markets. Examples 1 and 2 both exhibit markets that admit nondegenerate reduction sequences.

$k.j$. Each student who was rejected in stage $k.(j-1)$ applies to her most preferred school that has not yet rejected her. Each school s considers its new applicants jointly with the students held at the end of stage $k.(j-1)$, and tentatively accepts students in the same manner as described in $k.1$. All students not accepted are rejected.

Stage k continues until the substage $k.j$ at which no students are rejected. If the tentative matching at this point is feasible, end the algorithm and output this matching. If not, proceed to stage $k+1$.

The basic idea behind DQDA is to start with high ceilings and capacities, and check whether given the submitted preferences, the output of DA satisfies the floors as well.[25] If so, the algorithm ends with the high ceilings. If not, only then do we lower the ceilings. This initiates a rejection chain. The rejected students then apply to their next most preferred school, which rejects its lowest priority students, and so forth, continuing until no further students are rejected. We continue gradually lowering the ceilings until all floors are filled. The key is that the dynamic adjustment process of DQDA only lowers ceilings after taking the submitted preferences of the students into account and stops as soon as all floors are filled, which results in fewer seats being eliminated unnecessarily.

Example 2. The following example provides an illustration of the DQDA algorithm. Let $S=\{s_1,s_2,s_3,s_4\}$, $\Theta=\{\ell,h\}$, and $I=\{\ell_1,h_1,h_2\}$. Consider the school quotas/priorities and student preferences given in the following table. All floors are zero except for the type h floor at s_4.

Schools

	$L_{s,\ell}$	$L_{s,h}$	$U_{s,\ell}$	$U_{s,h}$	Q_s	$\succ_s$
s_1	0	0	1	1	1	$h_1 \succ_{s_1} h_2 \succ_{s_1} \ell_1$
s_2	0	0	1	1	1	$\ell_1 \succ_{s_2} h_1 \succ_{s_2} h_2$
s_3	0	0	1	1	1	$\ell_1 \succ_{s_3} h_2 \succ_{s_3} h_1$
s_4	0	1	1	2	2	$h_1 \succ_{s_4} h_2 \succ_{s_4} \ell_1$

Students

P_{ℓ_1}: s_2, s_3, s_1, s_4
P_{h_1}: s_2, s_1, s_4, s_3
P_{h_2}: s_3, s_4, s_1, s_2

The last object needed is a reduction sequence. Consider

$$\eta=\left\{\left(\begin{bmatrix}1&1\\1&1\\1&1\\1&2\end{bmatrix},\begin{bmatrix}1\\1\\1\\2\end{bmatrix}\right),\left(\begin{bmatrix}1&0\\1&1\\1&1\\1&2\end{bmatrix},\begin{bmatrix}0\\1\\1\\2\end{bmatrix}\right),\left(\begin{bmatrix}1&0\\1&0\\1&1\\1&2\end{bmatrix},\begin{bmatrix}0\\0\\1\\2\end{bmatrix}\right)\right\}.$$

In words, this reduction sequence starts with $(U^1,Q^1)=(U,Q)$. At the beginning of stage 2, we lower the capacity and type h ceiling at s_1 by 1, while at the beginning of

[25] A potential concern of the reader may be that the DQDA algorithm does not produce a feasible assignment and/or may leave some students unmatched. Given our construction of the algorithm and the reduction sequence, this is not an issue. In particular, we choose the reduction sequence such that the final capacities (U^K,Q^K) ensure a feasible match (and hence all students are assigned). We show that under minimality, DQDA Pareto dominates DA under (U^K,Q^K), and so these properties hold under DQDA as well.

stage 3, we lower the capacity and type h ceiling at s_2 by 1. Note that the final entry, (U^3, Q^3), ensures a feasible matching, while (U^1, Q^1) and (U^2, Q^2) do not.

The output of stage $k = 1$ is

$$\mu^1 = \begin{pmatrix} s_1 & s_2 & s_3 & s_4 \\ h_1 & \ell_1 & h_2 & \varnothing \end{pmatrix}.$$

The type h floor at s_4 is not satisfied, and so we must move to stage 2. At the beginning of stage 2, $U_{s_1,h}$ and Q_{s_1} are lowered by 1, and so student h_1 is rejected from s_1, beginning a rejection chain. When this rejection chain ends, the output at the end of stage 2 is

$$\mu^2 = \begin{pmatrix} s_1 & s_2 & s_3 & s_4 \\ \varnothing & \ell_1 & h_2 & h_1 \end{pmatrix}.$$

Matching μ^2 satisfies all of the primitive floors and ceilings, and thus there is no need to move to stage 3. The final output is μ^2, which Pareto dominates the matching that would have been implemented by ACDA under artificial caps of $(\bar{U}, \bar{Q}) = (U^3, Q^3)$:

$$\mu^{\text{ACDA}} = \begin{pmatrix} s_1 & s_2 & s_3 & s_4 \\ \varnothing & \varnothing & \ell_1 & \{h_1, h_2\} \end{pmatrix}.$$

Student h_1 is indifferent between the two outcomes, but both ℓ_1 and h_2 strictly prefer DQDA. ◊

There is an alternative way to define a dynamic quotas algorithm that is very natural. At the end of each stage, rather than leaving everyone at their assigned schools and lowering the ceilings from (U^{k-1}, Q^{k-1}) to (U^k, Q^k), we could instead remove all students from their assigned schools and run the entire deferred acceptance algorithm from the beginning under the lower ceilings (U^k, Q^k). We call this version of the algorithm *sequential deferred acceptance* (SDA).

Sequential deferred acceptance

Recall that $\text{DA}^{(U',Q')}(\cdot)$ denotes the DA mechanism using ceilings and capacities (U', Q'). Given a reduction sequence $\eta = \{(U^1, Q^1), (U^2, Q^2), \dots, (U^K, Q^K)\}$, the sequential deferred acceptance algorithm is defined as follows.

Stage 1. Compute $\text{DA}^{(U^1,Q^1)}(P_I)$, the outcome of DA under (U^1, Q^1). If $\text{DA}^{(U^1,Q^1)}(P_I)$ is a feasible matching, end the algorithm and output this matching. If not, proceed to stage 2.

In general, proceed as follows.

Stage k. Lower the ceilings and capacities to (U^k, Q^k) and compute $\text{DA}^{(U^k,Q^k)}(P_I)$, the outcome of DA under (U^k, Q^k). If $\text{DA}^{(U^k,Q^k)}(P_I)$ is a feasible matching, end the algorithm and output this matching. If not, proceed to stage $k+1$.

While DQDA and SDA are similar, in general they are not equivalent mechanisms. However, there is an important class of reduction sequences where they in fact are equivalent: minimal reduction sequences.

THEOREM 3. *Let η be a minimal reduction sequence. For any preferences P_I, the matching produced by DQDA under η is equivalent to that produced by SDA under η.*

The important implication of Theorem 3 is that the final matching output by DQDA is ex post equivalent to the matching that would have been produced by standard DA under some (fixed) ceilings and capacities (U^k, Q^k). This will be helpful in proving our main result below. It is imperative to note, however, that ex ante, it is not known what the final (U^k, Q^k) will be, as this will be determined by the submitted preferences.

The main result

With these preliminary results in hand, we can now state and prove Theorem 4, the main result.

THEOREM 4. *(i) Every ACDA mechanism is Pareto dominated by some DQDA mechanism under a minimal reduction sequence η.*

Further, for any minimal reduction sequence η, the following statements hold:

(ii) DQDA under η eliminates justified envy among same types.

(iii) DQDA under η is strategyproof.

Before discussing the intuition behind the proofs of these results, we first comment on their interpretation. Note that DQDA takes the reduction sequence η as an input, and different reduction sequences lead to different DQDA mechanisms. DQDA under any minimal $\eta = \{(U^1, Q^1), \dots, (U^K, Q^K)\}$ Pareto dominates ACDA under (U^K, Q^K), and while DQDA under η may not Pareto dominate ACDA under some other (U', Q'), the important point to note is that, given such a (U', Q'), there is an *alternative* reduction sequence η' ($\neq \eta$) that *will* Pareto dominate ACDA under (U', Q').[26] Thus, the upshot

[26]Given two different reduction sequences η and η', DQDA under η and DQDA under η' in general are Pareto incomparable, as different students will have different preferences over the order in which quotas are reduced. However, all students will prefer DQDA to the relevant artificial caps. As an analogy, consider the serial dictatorship, a popular mechanism for allocating discrete objects when there are no priorities. The serial dictatorship works by first fixing some ordering of the students, and then allowing the students to pick their favorite schools according to this ordering. The formal definition of the serial dictatorship takes the fixed student ordering as an input, similar to how DQDA takes η as an input, and different orderings lead to different serial dictatorship mechanisms. Obviously, there will be no Pareto dominance relation between serial dictatorships with different student orderings (each student will prefer the ordering that allows her to choose first), just as there will be no Pareto dominance relation between DQDA mechanisms with different choices for η. However, the importance of this result is that it is *always* possible to improve on ACDA without compromising fairness or incentives. The key feature needed to retain strategyproofness is that η be fixed ex ante, i.e., the submitted student preferences cannot affect the order in which ceilings

of the theorem is that any choice of artificial caps (including those that are used in practice by institutions like the JRMP or the military) is inefficient in a strong sense: it is always possible to achieve a Pareto improvement using a dynamic quotas mechanism. The remaining parts of Theorem 4 show that, importantly, these Pareto improvements can be achieved *without harming the fairness or incentive properties of the algorithm*. Thus, there seems to be little reason for any market using an artificial caps mechanism in practice not to consider switching to a dynamic quotas mechanism instead.

Intuition for Theorem 4 Next, we discuss the intuition for Theorem 4 (the full proof is given in Appendix B). First, we note that Theorem 4 is actually a special case of a more general result that we build in Appendix B.

Remark 1. While early hospital–resident and school choice matching models assume simple linear preference/priority relations over individual agents on the other side of the market (as we do in the main text here), as the theory has developed, this has been continually expanded to accommodate other types of priority structures that do not take this simple form. At the most general level, one can define abstract *choice functions* $\mathrm{Ch}_s(\cdot)$ for the schools, where $\mathrm{Ch}_s(I') \subseteq I'$ returns the students school s admits from any potential subset of applicants I' (see, for example, Hatfield and Milgrom 2005 or Abdulkadiroğlu 2005). It is possible to generalize our model in a similar way and to define a generalized DQDA algorithm where η is a sequence of choice functions. To ensure the good properties from Theorem 4 hold in general, we must place additional structure on the choice functions. The most important condition is *monotonicity*, which says that, fixing the set of applicants at a school, the set of rejected students expands as we move to later stages. This purpose of this condition is analogous to the substitutability condition of Hatfield and Milgrom (2005): both guarantee that as the algorithm progresses, a school never wants to admit a student it has previously rejected. Our condition is different, however, because of the dynamic nature of the choice functions in our generalized algorithm: as the choice functions evolve from stage k to stage $k+1$, we must ensure that they evolve in such a way that no school ever wants to admit a student it previously rejected in an earlier stage. Minimal reduction sequences are a special case of monotonic choice functions. So as not to obscure the main insight of the paper, we refer interested readers to Appendix A for details.

Part (i) For part (i), note that DQDA ends under some ceilings and capacities $(U^k, Q^k) \geq (U^K, Q^K)$. At first glance, it may seem obvious that the final matching produced by DA under higher ceilings and capacities should be preferred by all students to the outcome under lower ceilings and capacities. However, the full argument is more subtle, as lower ceilings for one type of student may actually be beneficial to other types. For example, consider a model with two types, $\Theta = \{\ell, h\}$, and a school s with priority

are reduced. Within these constraints, policymakers may actively choose the ceiling that is to be reduced at each stage so as to achieve some desirable policy goals; alternatively, at each stage k we can randomly choose some pair (s, θ), subject to feasibility constraints (just as the random serial dictatorship randomly chooses a student ordering and implements the serial dictatorship using this ordering).

relation $\succ_s$ and upper quotas and capacities in stages k and $k+1$ as given by

	$U_{s,\ell}$	$U_{s,h}$	Q_s
k	2	2	2
$k+1$	2	1	2

$\succ_s$: $h_1, h_2, \ell_1, \ldots$.

Note that the preceding table is *not* a minimal reduction sequence, because $U_{s,h}$ is reduced, while Q_s is not. If the set of students who apply to school s under the stage k quotas is $I' = \{h_1, h_2, \ell_1\}$, the students admitted would be $\{h_1, h_2\}$. However, under the stage $k+1$ quotas, the admitted students would be $\{h_1, \ell_1\}$. So if ℓ_1 prefers school s (if s is the first choice for all students, for example), then ℓ_1 will actually be strictly *worse off* in stage k, under the "higher" capacities. This insight can be embedded into a full market example to show that DQDA does not Pareto dominate DQDA for any arbitrary reduction sequence.

The reason this occurs in the above example is because in both stages, school s has a total capacity of 2, but in stage k, both of these seats can go to a type h student, while in stage $k+1$, only one of these seats can go to a type h student. The higher ceiling for type h students in stage k makes student ℓ_1 worse off, because the type h students then prevent her from getting a seat at s. Formally, to ensure the Pareto dominance result holds, what is needed is that the set of students chosen from any given set of applicants must be weakly larger (in the set inclusion sense) in earlier stages compared to later stages. This is exactly the monotonicity assumption that was discussed in Remark 1. This is not true in the above example; however, the assumption that η is a minimal reduction sequence guarantees that this holds.

Part (ii) Part (ii) is the simplest part of the theorem: it follows from the fact that standard DA with type-specific ceilings eliminates justified envy among same types together with the fact that, ex post, the final matching produced by DQDA is equivalent to standard DA for some (U^k, Q^k) (of course, it is not known ex ante what the final stage k will be, which is precisely why we need to run the dynamic quotas algorithm in the first place).[27]

Part (iii) Parts (i) and (ii) can be shown by analyzing DA within each stage k separately. Analyzing the incentive properties of DQDA is significantly more complicated, because

[27] Recall that there is an impossibility result that says that matchings that eliminate justified envy (across types) may not even exist (see Section 3), and so the most we can hope to achieve is elimination of justified envy among same types. If, in addition, we require non-wastefulness, another impossibility result obtains: matchings that simultaneously eliminate justified envy among same types and are non-wasteful may also fail to exist (see Ehlers et al. 2014 and Fragiadakis et al. 2015). Thus, one of these properties must be weakened. Markets that use ACDA are opting to keep elimination of justified envy among same types and weaken non-wastefulness. However, as Theorem 4 shows, ACDA weakens non-wastefulness more than necessary, as DQDA Pareto dominates ACDA while still satisfying elimination of justified envy among same types and strategyproofness.

they cannot be understood by looking at each stage independently. The submitted preferences themselves have the potential to alter the final capacities, which may (potentially) give the agents the ability to manipulate their preferences in such a way as to change when the algorithm ends and make themselves better off.

To understand the intuition behind why DQDA is strategyproof, it is easiest to restrict to the case of a single type, so that each school has only one aggregate floor constraint and one aggregate capacity constraint. Then fix the preferences of all students other than i at P_{-i}. Let k be the final stage of DQDA if i submits her true preferences P_i. The potential manipulations P_i' that i can make can be classified into three types: contractions (P_i' causes DQDA to end at some $k' < k$), extensions (P_i' causes DQDA to end at some stage $k' > k$), and equivalences (P_i' causes DQDA to end in stage $k' = k$). The last category is the easiest to rule out as being profitable: if both P_i and P_i' cause DQDA to end in stage k, then, since DQDA is equivalent to standard DA in stage k, strategyproofness of DA implies that any such manipulation is not profitable. For extensions, if P_i' causes DQDA to continue beyond stage k, then there are less seats available for all students. This makes every student worse off (by the same argument used for part (i) above), and so extensions are never profitable.

The most difficult types of manipulations to rule out are contractions. Contractions seem to have the most potential for profitability, because just as less seats tend to make students worse off, more seats may have the potential to make students better off. The question, though, is whether the algorithm allows student i to manipulate her preferences to end the algorithm under higher ceilings, and to do so in such a way that makes her better off. The first part is true (it is possible for a student to manipulate to cause the algorithm to end under higher ceilings), but the second is not (due to the way we have constructed the algorithm, any false report she can submit to do so makes her weakly worse off).

To understand why, consider a student i with preferences P_i: $s_1, s_2, s_3, \dots$. Begin by running DA on all students other than i,[28] and assume after this is done that there is only one floor seat left to be filled, at school s_2. Now, DQDA ends the next time any student applies to s_2. One option available to student i is to lie and list school s_2 as her true first choice, thereby ending the algorithm immediately (a contraction) with her receiving s_2. If i instead submits her true preferences and first applies to s_1, she may initiate a chain of rejections that ends with some other student applying to s_2, in which case i receives s_1, her true first choice. The key observation is that even if i is eventually rejected from s_1 (e.g., if a seat at s_1 is eliminated at some point during the running of the algorithm), she then simply applies to s_2 and the algorithm ends. The seat at s_2 is always available until someone applies to it, at which point the algorithm ends and all assignments are made permanent. Thus, *there is no harm* in i reporting her top choices truthfully, because seats at lower ranked schools are always available to her. Note that when there is a single type, monotonicity of the school choice functions (as discussed above) holds trivially, since lowering the capacity at a school automatically eliminates

[28]The order in which students are allowed to apply is irrelevant, a result first shown by McVitie and Wilson (1971).

the seat for all students, since everyone is the same type. More generally, with multiple types, we may run into problems similar to those in the discussion of part (i) above. In the general model in Appendix A, monotonicity of the choice functions is a key property used in the proof of strategyproofness.[29]

Note that while Theorem 4 shows that DQDA Pareto dominates ACDA, it does not say that DQDA always selects a Pareto efficient matching. This is unsurprising because, even without floors, the DA outcome may not be Pareto efficient due to the fairness (no justified envy) constraints. DA is still widely used, however, because fairness is considered an important property in many markets. Given this, it is natural to ask whether DA and/or our DA-type mechanisms are constrained efficient. We say that a feasible matching μ that eliminates justified envy among same types is *constrained efficient* if there is no other feasible matching μ' that eliminates justified envy among same types and Pareto dominates μ.[30] We say a mechanism ψ is constrained efficient if it always selects a constrained efficient matching.

When there are no floor constraints, DA is strategyproof, eliminates justified envy, and is constrained efficient (Gale and Shapley 1962, Abdulkadiroğlu and Sönmez 2003). As we have seen, the introduction of floor constraints creates complications that lead to the incompatibility of many of the good properties of DA in the standard model, and this continues here: neither ACDA nor DQDA is a constrained efficient mechanism. This follows from the following impossibility result, shown by Ehlers et al. (2014).

Proposition 1 (Ehlers et al. 2014). *There is no feasible mechanism that is strategyproof, eliminates justified envy among same types, and is constrained efficient.*[31]

Since DQDA is strategyproof and eliminates justified envy among same types, it is not constrained efficient. The impossibility result shows that this is a general problem caused by the presence of floors, and so trade-offs are required between these three desirable properties. Markets that currently make use of ACDA are choosing to prioritize strategyproofness and elimination of justified envy, while weakening efficiency. What we show is that ACDA weakens efficiency more than necessary, and we are the first to provide a new mechanism that is more efficient without compromising important fairness or incentive properties.

[29] An implicit, yet important, feature of the definition of DQDA is that the reduction sequence η be determined exogenously to the submitted preferences, which is key in proving strategyproofness. In the Supplemental Material (available in a supplementary file on the journal website, http://econtheory.org/supp/2195/supplement.pdf), we define an alternative endogenous reduction DQDA (EDQDA) algorithm that allows the reduction sequence to be determined endogenously. EDQDA is not strategyproof.

[30] We focus on eliminating justified envy among same types only because even with this weaker condition, there is an impossibility result (Proposition 1).

[31] Ehlers et al. (2014) actually show a stronger result, namely that there is no mechanism that is strategyproof, eliminates justified envy among same types, and is *constrained non-wasteful*. Their definition of constrained non-wastefulness is implied by constrained efficiency, and so there also is no mechanism that is strategyproof, eliminates justified envy among same types, and is constrained efficient.

5. General mechanisms with dynamic quotas

For most of this paper, we have focused specifically on deferred acceptance. However, the main ideas behind artificial caps and dynamic quotas can easily be applied to any mechanism that has inputs for ceiling constraints but not for floors. This includes most standard mechanisms used in practice, such as the immediate acceptance mechanism, the serial dictatorship, and top trading cycles.

Formally, we define a class of mechanisms called *upper quota mechanisms*. Upper quota mechanisms are indexed by vectors of ceilings and capacities (U', Q'), which we refer to jointly as upper quotas. Let $\mathcal{M}(U', Q') = \{\mu \in \mathcal{M} : |\mu_\theta(s)| \leq U'_{s,\theta}$ and $|\mu(s)| \leq Q'_s$ for all $(s, \theta)\}$ be the set of all matchings that respect (U', Q'). An upper quota mechanism is a function $\psi^{(U',Q')} : \mathcal{P}^n \to \mathcal{M}$ such that $\psi^{(U',Q')}(P_I) \in \mathcal{M}(U', Q')$ for all $P_I \in \mathcal{P}^n$. A collection of mechanisms, one for each (U', Q'), is denoted $\Psi := \{\psi^{(U',Q')}\}_{(U',Q')}$. We refer to Ψ as a *class* of upper quota mechanisms. As an example, Ψ could be the class of DA mechanisms as defined in Section 2: $\Psi = \{\mathrm{DA}^{(U',Q')}\}_{(U',Q')}$. Note that upper quota mechanisms always satisfy the given ceilings and capacities (U', Q'), but again need not be feasible, for reasons similar to those discussed for DA in Section 2. If $(\bar{U}, \bar{Q})$ are such that $\psi^{(\bar{U},\bar{Q})}(P_I)$ is feasible for all P_I, then we call $\psi^{(\bar{U},\bar{Q})}$ an artificial caps mechanism. For example, when Ψ is the class of DA mechanisms, then $\psi^{(\bar{U},\bar{Q})}$ is ACDA under artificial caps $(\bar{U}, \bar{Q})$.

Definition 3. Let (U', Q') and (U'', Q'') be such that $(U', Q') \leq (U'', Q'')$ and $\sum_{\theta\in\Theta}(U''_{s,\theta} - U'_{s,\theta}) \leq Q''_s - Q'_s$ for all $s \in S$. The class of mechanisms Ψ is *resource monotonic* if, for all such (U', Q') and (U'', Q''), $\psi^{(U'',Q'')}$ weakly Pareto dominates $\psi^{(U',Q')}$.

Resource monotonicity means that raising the ceilings and capacity of a school makes all students weakly better off, provided that the type-specific ceilings are not raised more than the capacity.[32]

We now generalize dynamic quotas to any class of mechanisms Ψ.

Dynamic quotas Ψ (DQ^Ψ)

Fix a sequence of ceiling–capacity vectors $\eta = \{(U^1, Q^1), (U^2, Q^2), \ldots, (U^K, Q^K)\}$ such that (i) $(U^{k+1}, Q^{k+1}) \leq (U^k, Q^k) \leq (U, Q)$ for all k and (ii) $\psi^{(U^K,Q^K)}(P_I)$ is feasible for all P_I. The algorithm then proceeds in a series of stages.

Stage 1. Calculate $\psi^{(U^1,Q^1)}(P_I)$. If $\psi^{(U^1,Q^1)}(P_I)$ is a feasible matching, end the algorithm and output this matching. If not, proceed to stage 2.

[32] Resource monotonicity has been used as an important axiom in many allocation settings (see, for example, Ehlers and Klaus 2004, Kesten 2006, and Thomson 2010). The previous works do not have type-specific ceilings, and prior notions of resource monotonicity say that raising just the capacity of a school makes all agents better off (in a Pareto sense). This is implied by Definition 3, but for our purposes, we want to allow the type-specific ceilings to be raised as well. However, we must do this in a "controlled" manner: raising the type-specific ceilings by more than the number of capacity seats may make students of other types whose ceilings were not raised worse off (see the discussion after Theorem 4).

In general, proceed as follows.

Stage k. Calculate $\psi^{(U^k,Q^k)}(P_I)$. If $\psi^{(U^k,Q^k)}(P_I)$ is a feasible matching, end the algorithm and output this matching. If not, proceed to stage $k+1$.

We define dynamic quotas Ψ as the function $\mathrm{DQ}^{\Psi}: \mathcal{P}^n \to \mathcal{M}$ that produces, for each input, the matching at the end of the above algorithm.

Theorem 5. *Assume that Ψ is resource monotonic and η is a minimal reduction sequence. Then the dynamic quotas mechanism* DQ^{Ψ} *weakly Pareto dominates the artificial caps mechanism $\psi^{(U^K,Q^K)}$.*

DQDA is a special case of DQ^{Ψ} when Ψ is the set of deferred acceptance mechanisms and η is a minimal reduction sequence (by Theorem 3). Dynamic quotas can also be applied to many other choices of Ψ. Besides DA, common upper quota mechanisms used in practice include the serial dictatorship (SD), immediate acceptance (IA), and top trading cycles (TTC).[33] Versions of immediate acceptance are used in the Cambridge, MA and Minneapolis, MN school districts, and the algorithm has previously been used in Seattle, WA and Boston, MA. Versions of TTC are currently used in school districts in San Francisco, CA and New Orleans, LA.

Since these are popular mechanisms, they are all natural candidates for use in the presence of floor constraints as well. Again, by imposing sufficiently strict artificial caps, it is possible to ensure a feasible matching. However, doing so results in the same inefficiencies as with deferred acceptance, because artificial caps still eliminates seats ex ante, ignoring important information contained in the student preferences. We thus argue that any market that uses artificial caps SD/IA/TTC would be better served by switching to dynamic quotas SD/IA/TTC. Because SD and IA are indeed resource monotonic, dynamic quotas SD/IA will Pareto dominate artificial caps SD/IA. TTC, alternatively, is not resource monotonic (Kesten 2006), and so, the Pareto dominance of dynamic quotas over artificial caps will not hold for TTC. However, it is still be true that the final matching implemented under dynamic quotas have higher ceilings than artificial caps, which intuitively makes the students better off on average. We thus suggest that policymakers may want to consider dynamic quotas over artificial caps, even if resource monotonicity does not hold formally.

6. Conclusion

This paper shows that a common approach used in many matching markets with distributional constraints may result in avoidable inefficiencies. We introduce a new class of mechanisms based on a concept of dynamically adapting quotas. By using agents' submitted preferences to allocate positions more flexibly given the reported demand, we

[33]The serial dictatorship is a simple mechanism where students are ordered and then one at a time, choose their favorite school that has space remaining. Immediate acceptance (also known as the Boston mechanism) and TTC (first defined by Shapley and Scarf 1974) are mechanisms that are common in the field and are well known in the literature, and so we refer the reader to other papers for their full definitions and analysis. A good introduction can be found in Abdulkadiroğlu and Sönmez (2003).

are able to recover these inefficiencies. A key feature of our dynamic quotas approach is its straightforward and intuitive nature, which is essential for practical implementation in market design settings where the mechanism used must be easily communicable to participants. We also provide a rigorous theoretical analysis with respect to crucial properties such as incentives and efficiency. We show that it is possible to improve upon the existing approaches in a Pareto sense without compromising fairness or incentives, suggesting that market participants would benefit from implementing our mechanisms. Methodologically, we identify conditions under which the popular deferred acceptance remains strategyproof even using dynamic quotas, and, in showing this result, introduce techniques that may be helpful in designing strategyproof mechanisms for other settings as well.

Our theoretical results have applications to a potentially wide array of market design contexts, including hospital–resident matching, school choice, and military cadet branching, as well as a variety of less formal settings such as a firm assigning employees to projects, each of which must have at least a certain number of workers assigned to it. We take no position on the merits of imposing distributional constraints in any of these contexts: the Army does so out of necessity, medical residency programs may do so out of a concern over access to health care in rural areas, and many school districts believe there are social benefits to diverse educational environments. While debating the merits of such constraints is undoubtedly necessary, equally important is the determination of the precise mechanism by which they will actually be implemented in practice. This paper contributes to the latter research by extending the market design approaches advocated by Roth (2002) to settings with hard constraints. We then provide practical mechanisms within the framework of such constraints that have the potential to be successful in real-world applications. We do argue, however, that markets that impose artificial caps as a way to satisfy some "implicit" floor constraints (such as the Japan Residency Matching Program or other markets in which the true distributional constraints are not publicly stated) should consider modeling their goals more explicitly and switching to a dynamic quotas mechanism similar to those provided in this paper. As we have shown, doing so improves welfare without compromising fairness, incentives, or the true underlying distributional goals.

Appendix A: A general model of matching under distributional constraints

In this section, we define a model that allows for more general choice functions and feasibility constraints. We then introduce some conditions on choice functions and preliminary theorems that will be useful in the proof of strategyproofness (found in the next section), though they also may be of independent interest. To simplify the flow of the argument, the proofs of some lemmas are to be found in Appendix C.

A.1 *Primitives*

The primitives of the general model once again consist of a set of agents $I = \{i_1, \dots, i_n\}$ and objects $S = \{s_1, \dots, s_m\}$, which, for consistency, we continue to refer to as students and schools, respectively. Each student has a strict preference relation P_i on S. How schools determine which students to admit is now described differently. For any set X, let 2^X denote the power set of X. For each school s, we define a choice function $\mathrm{Ch}_s : 2^I \to 2^I$, where, for every $I' \in 2^I$, $\mathrm{Ch}_s(I') \subseteq I'$ denotes the set of students admitted to school s when its choice set is I'. This setup is similar to the model of Hatfield and Milgrom (2005).[34] In the baseline model, $\mathrm{Ch}_s(I')$ would be the highest priority subset of students in I' subject to the floors and ceilings (see Section 2). Corresponding to each choice function is a *rejection function* $\mathrm{Rej}_s(I') = I' \setminus \mathrm{Ch}_s(I')$. Let $\mathrm{Ch} := \{\mathrm{Ch}_{s_1}, \dots, \mathrm{Ch}_{s_m}\}$ denote a vector of choice functions, one for each school.

The following two conditions on choice functions were identified by Hatfield and Milgrom (2005) as key for strategyproofness of DA in a model without floor constraints.

Definition 4. Choice function Ch_s is *substitutable* if $I' \subseteq I''$ implies $\mathrm{Rej}_s(I') \subseteq \mathrm{Rej}_s(I'')$.

Definition 5. Choice function Ch_s satisfies the *law of aggregate demand* if $I' \subseteq I''$ implies $|\mathrm{Ch}_s(I')| \le |\mathrm{Ch}_s(I'')|$.

We assume that all choice functions defined from here forward satisfy both substitutability and the law of aggregate demand.

Let $\mathcal{I}(s) = \{I' \subseteq I : I' = \mathrm{Ch}_s(I'') \text{ for some } I'' \in 2^I\}$.[35] In words, $\mathcal{I}(s)$ is a set consisting of all possible assignments for school s, obtained by considering every potential set of applicants I'' that s may have the opportunity to choose from.

The Ch_s functions above describe the primitive choice functions of the individual schools. Beyond this, we allow the school district to impose additional *distributional constraints* at the district level. In the main text, these are the type-specific floors and ceilings, and the district can declare any matching that does not satisfy them as infeasible. In this more general setting, we assume for each school s that the school district defines a priori a subset $\mathcal{I}_f(s) \subseteq \mathcal{I}(s)$ of *feasible assignments*.[36] The set of *feasible matchings* $\mathcal{M}_f$ is then defined as

$$\mathcal{M}_f = \{\mu \in \mathcal{M} : \mu(s) \in \mathcal{I}_f(s) \text{ for all } s \in S\}.$$

Note that the definition of $\mathcal{I}_f(s)$ and IRS (footnote 34) imply that if $\mu \in \mathcal{M}_f$, then $\mathrm{Ch}_s(\mu(s)) = \mu(s)$ for all s. As in the main text, we continue to assume that the set of feasible matchings is nonempty.

[34] Aygün and Sönmez (2013) point out a technical ambiguity in the model of Hatfield and Milgrom (2005), noting that to ensure the choice functions are derived from a well defined underlying priority relation $\succ_s$ over *sets* of students, one must assume a condition called *irrelevance of rejected students* (IRS), which, in our setting, says that if $\mathrm{Ch}_s(I'') \subseteq I' \subseteq I''$, then $\mathrm{Ch}_s(I') = \mathrm{Ch}_s(I'')$. We assume below that all of our choice functions satisfy substitutability and the law of aggregate demand, which Aygün and Sönmez (2013) show implies IRS, and so we are justified in working directly with the choice functions, rather than the underlying priority relation. See also Fleiner (2003), who discusses similar points.

[35] Note that by IRS (footnote 34), $I' \in \mathcal{I}(s)$ implies that $I' = \mathrm{Ch}_s(I')$.

[36] In the main text, $\mathcal{I}_f(s)$ consists of all assignments that satisfy a school's type-specific floor and ceiling constraints and capacities. In the standard school choice model (e.g., Abdulkadiroğlu and Sönmez 2003), $\mathcal{I}_f(s) = \mathcal{I}(s)$.

In this context, a reduction sequence is now a sequence of vectors of choice functions $\eta = \{\text{Ch}^1, \ldots, \text{Ch}^K\}$, where each $\text{Ch}^k = (\text{Ch}_s^k)_{s \in S}$ is a vector of choice functions (one for each school) such that (i) $\text{Ch}_s(\text{Ch}_s^k(I')) = \text{Ch}_s^k(I')$ for all $s \in S$, $k = 1, \ldots, K$ and all $I' \in 2^I$, and (ii) deferred acceptance under choice function vector Ch^K always results in a feasible match for any preference profile P_I.[37] Assumption (i) guarantees that the reduction sequence never violates each school's primitive choice function (e.g., by accepting too many students such that the school's capacity would be violated) and (ii) is the same assumption made for artificial caps DA that ensures, under restrictive enough choice functions, we always obtain a feasible match. As discussed above, we continue to assume that all $\text{Ch}_s^k(\cdot)$ are substitutable and satisfy the law of aggregate demand.

Definition 6. Reduction sequence $\eta = \{\text{Ch}^1, \ldots, \text{Ch}^K\}$ is *monotonic* if $\text{Rej}_s^{k'}(I') \subseteq \text{Rej}_s^{k''}(I')$ for all $s \in S$, $I' \subseteq I$, and $k'' \geq k'$.

Remark 2. If η is monotonic and Ch_s^k is substitutable for all s and k, we say η is *monotonically substitutable*. Note that if η is monotonically substitutable, then $I' \subseteq I''$ and $k' \leq k'' \implies \text{Rej}_s^{k'}(I') \subseteq \text{Rej}_s^{k''}(I'')$.[38]

Definition 7. Reduction sequence η is *minimal* if the following statements hold for all k: (i) for exactly one s, $0 \leq |\text{Ch}_s^k(I')| - |\text{Ch}_s^{k+1}(I')| \leq 1$ for all $I' \subseteq I$ and (ii) for all remaining $s' \neq s$, $\text{Ch}_{s'}^{k+1}(\cdot) = \text{Ch}_{s'}^k(\cdot)$.

Minimality guarantees that when moving from stage k to $k+1$, at most one student will be rejected. Lemma 2 below shows that any (quota) reduction sequence (as defined in Section 4) that is minimal in the sense of Definition 2 induces a choice function reduction sequence that is monotonic and minimal (in the sense of Definitions 6 and 7), and each $\text{Ch}_s^k(\cdot)$ satisfies substitutability and the law of aggregate demand for all k and all s.

A.2 *Generalized DQDA*

We now define a generalized version of the DQDA algorithm that takes as an input a reduction sequence of choice functions. The description of the algorithm makes use of the *cumulative offer process* of Hatfield and Milgrom (2005) (see also Hatfield and Kojima 2009). As the cumulative offer process progresses, schools continually accumulate applications from students, and at each point, hold on to their most preferred set students among all of those who have cumulatively applied to it. Students who are not currently held by any school make new applications to their most preferred school that

[37]As in the main text, it is always possible to find such feasibility-ensuring choice functions: simply take any feasible match μ and set $\text{Ch}_s^K(I') = \mu(s) \cap I'$ for all I'. This is of course very restrictive, and there are in general many less restrictive ways to achieve this. In addition, a (quota) reduction sequence as defined in Definition 1 induces a (choice function) reduction sequence in the natural way.

[38]$\text{Rej}_s^{k'}(I') \subseteq \text{Rej}_s^{k'}(I'') \subseteq \text{Rej}_s^{k''}(I'')$, where the first inclusion is by substitutability and the second is by monotonicity.

has not yet rejected them. As a matter of notation, we use $\mathcal{A}_s^k(t)$ to denote the *cumulative* set of students who have applied to school s, up to and including stage k, step t of the algorithm (i.e., $\mathcal{A}_s^k(t)$ includes all students who ever made an application to s in the algorithm, up to and including stage k, step t).

Generalized dynamic quotas deferred acceptance (GDQDA) Fix a reduction sequence $\eta = \{\mathrm{Ch}^1, \ldots, \mathrm{Ch}^K\}$.

Stage 1. Step 0. Set $\mathcal{A}_s^1(0) = \varnothing$ for all $s \in S$.

Step 1. Choose some student i^1 who applies to her favorite school, s^1.[39] Let $\mathcal{A}_{s^1}^1(1) = \{i^1\}$ and $\mathcal{A}_s^1(1) = \mathcal{A}_s^1(0)$ for all other $s \in S$. Each school $s \in S$ tentatively accepts the students in $\mathrm{Ch}_s^1(\mathcal{A}_s^1(1))$ and rejects the rest.

Step t. Choose a student i^t who is not tentatively accepted by any school, and let her apply to her most preferred school s^t that has not yet rejected her. Let $\mathcal{A}_{s^t}^1(t) = \mathcal{A}_{s^t}^1(t-1) \cup \{i^t\}$ and $\mathcal{A}_s^1(t) = \mathcal{A}_s^1(t-1)$ for all $s \neq s^t$. Each school $s \in S$ tentatively accepts the students in $\mathrm{Ch}_s^1(\mathcal{A}_s^1(t))$ and rejects all other students.

Stage 1 terminates when every student is either tentatively accepted by some school $s \in S$ or has applied to all schools and been rejected.[40] This happens in a finite number of steps T^1. Let the resulting matching be ν^1, where $\nu^1(s) = \mathrm{Ch}_s^1(\mathcal{A}_s^1(T^1))$ for all $s \in S$. If $\nu^1 \in \mathcal{M}_f$, end the algorithm and output matching ν^1. If not, proceed to stage 2.

In general, proceed as follows.

Stage k. Step 0. Set $\mathcal{A}_s^k(0) = \mathcal{A}_s^{k-1}(T^{k-1})$ for all $s \in S$, and let each school tentatively accept $\mathrm{Ch}_s^k(\mathcal{A}_s^k(0))$ and reject all remaining students.

Step 1. Choose a student i^1 who is not tentatively accepted by any school, and let her apply to her most preferred school s^1 that has not yet rejected her (in this stage or any previous stages). Let $\mathcal{A}_{s^1}^k(1) = \mathcal{A}_{s^1}^k(0) \cup \{i^1\}$ and $\mathcal{A}_s^k(1) = \mathcal{A}_s^k(0)$ for all other $s \in S$. Each school $s \in S$ tentatively accepts the students in $\mathrm{Ch}_s^k(\mathcal{A}_s^k(1))$ and rejects the rest.

Step t. Choose a student i^t who is not tentatively accepted by any school, and let her apply to her most preferred school s^t that has not yet rejected her (in this stage or any previous stages). Let $\mathcal{A}_{s^t}^k(t) = \mathcal{A}_{s^t}^k(t-1) \cup \{i^t\}$ and $\mathcal{A}_s^k(t) = \mathcal{A}_s^k(t-1)$ for all other $s \in S$. Each school $s \in S$ tentatively accepts the students in $\mathrm{Ch}_s^k(\mathcal{A}_s^k(t))$ and rejects the rest.

Stage k terminates when every student is tentatively accepted by some school $s \in S$ or has applied to all schools and been rejected. This happens

[39] It is well known that the order in which students are allowed to apply is irrelevant (see, e.g., McVitie and Wilson 1971 or Hatfield and Kojima 2010).

[40] Note that in this definition, we allow the possibility that a student applies to and is rejected from every school, in which case she is unmatched at the end of the algorithm. Under suitable choice functions that ensure feasible matchings, this is not an issue.

in a finite number of steps T^k. Let the resulting matching be defined by $\nu^k(s) = \mathrm{Ch}_s^k(\mathcal{A}_s^k(T^k))$ for all $s \in S$. If $\nu^k \in \mathcal{M}_f$, end the algorithm and output matching ν^k. If not, proceed to stage $k+1$.

Note that, as it stands, using the cumulative offer process, there is no assumption of consistency imposed on the algorithm, since some student could be assigned to more than one school or no school at all. However, we show below that under our conditions on η, this is not an issue.

Recall that in the main text, we introduced a second way to run a dynamic version of DA that we call sequential deferred acceptance. It is also possible to define a generalized sequential deferred acceptance (GSDA) algorithm in the context of the general model.

Generalized sequential deferred acceptance Fix a reduction sequence $\eta = \{\mathrm{Ch}^1, \dots, \mathrm{Ch}^K\}$, and let $\mathrm{DA}^{\mathrm{Ch}'} : \mathcal{P}^n \to \mathcal{M}$ denote the deferred acceptance algorithm under choice function vector Ch'.

Stage 1. Starting with the empty matching, compute $\mathrm{DA}^{\mathrm{Ch}^1}(P_I)$. If $\mathrm{DA}^{\mathrm{Ch}^1}(P_I) \in \mathcal{M}_f$, end the algorithm and output this matching. If not, proceed to stage 2.

In general, proceed as follows.

Stage k. Starting with the empty matching, compute $\mathrm{DA}^{\mathrm{Ch}^k}(P_I)$. If $\mathrm{DA}^{\mathrm{Ch}^k}(P_I) \in \mathcal{M}_f$, end the algorithm and output this matching. If not, proceed to stage $k+1$.

Theorem 6. *Let ν^k be the matching at the end of stage k of the GDQDA algorithm, and let μ^k be the matching at the end of stage k of the GSDA algorithm. Assume that η is monotonic and minimal, and that $\mathrm{Ch}_s^k(\cdot)$ satisfies substitutability and the law of aggregate demand for all k and s. Then $\nu^k = \mu^k$ for all k, and μ^k (and hence ν^k) assigns each student to at most one school.*

Proof. As in the definition of the algorithm above, let $\mathcal{A}_s^k(t)$ denote the cumulative offer set of school s at step t of stage k under GDQDA. For each stage k, let T^k denote the final step of stage k.

Now consider GSDA. Within a stage k, we are simply running DA under choice functions Ch^k. We can define an analogous *within-stage* cumulative offer process. Let $\mathcal{B}_s^k(t)$ denote the cumulative set of applicants to school s through step t of the within-stage cumulative offer process. Because GSDA starts each stage k from the empty matching, $\mathcal{B}_s^k(0) = \varnothing$ for all s and k. Let $\hat{T}^k$ denote the last step of stage k. Hatfield and Milgrom (2005) show that the matching produced by this cumulative offer process is equivalent to deferred acceptance, i.e., $\mu^k(s) = \mathrm{Ch}_s^k(\mathcal{B}_s^k(\hat{T}^k))$ for all $s \in S$ and all k.

The following lemma, which is proved in Appendix C, is key to the argument.

Lemma 1. *For all k and all s, $\mathcal{A}_s^k(T^k) = \mathcal{B}_s^k(\hat{T}^k)$.*

Note that, in general, $T^k \neq \hat{T}^k$, but the result says that the cumulative applicant sets at the end of each stage k are still the same. With this lemma, the result follows easily: $\mu^k(s) = \mathrm{Ch}_s^k(\mathcal{B}(\hat{T}^k)) = \mathrm{Ch}_s^k(\mathcal{A}_s^k(T^k)) = \nu^k(s)$ for all k and s, where the first and third equalities are by definition, and the second is by Lemma 1.

That μ^k assigns each student to at most one school follows from substitutability: assume not, i.e., assume that $i \in \mu^k(s)$ and $i \in \mu^k(s')$. Without loss of generality, assume $s \succ_i s'$. Then, by definition of $\mu^k(s')$, we have $i \in \mathcal{B}^k_{s'}(\hat{T}^k)$, which means that i applied to s and was rejected at some earlier step $t < \hat{T}^k$: $i \in \mathrm{Rej}^k_s(\mathcal{B}^k_s(t))$. But $\mathcal{B}^k_s(t) \subseteq \mathcal{B}^k_s(\hat{T}^k)$, and so substitutability implies $i \in \mathrm{Rej}^k_s(\mathcal{B}^k_s(\hat{T}^k))$, which contradicts that $i \in \mu^k(s)$. □

The following result is an immediate corollary of Theorem 6.

Corollary 1. *If η is monotonic and minimal, and $\mathrm{Ch}^k_s(\cdot)$ satisfies substitutability and the law of aggregate demand for all k and all s, then the final matching produced by GDQDA under η is equivalent to the final matching produced by GSDA under η.*

The next theorem is a more general statement of the result that dynamic quotas DA Pareto dominates artificial caps. Part (i) of Theorem 4 follows from this theorem.

Theorem 7. *Let $\eta = \{\mathrm{Ch}^1, \ldots, \mathrm{Ch}^K\}$ be monotonic and minimal, and assume that $\mathrm{Ch}^k_s(\cdot)$ satisfies substitutability and the law of aggregate demand for all k and all s. Then all students weakly prefer the outcome of GDQDA under η to DA under Ch^K.*

Proof. By Corollary 1, the final matching produced by GDQDA is equivalent to the matching produced by DA under choice functions $\mathrm{Ch}^{k'}$ for some $k' \leq K$. The outcome of DA in stage k is equivalent to the outcome of the cumulative offer process under Ch^k (Hatfield and Milgrom 2005, Hatfield and Kojima 2010). By monotonicity, $\mathrm{Ch}^K_s(I') \subseteq \mathrm{Ch}^{k'}_s(I')$ for all $I' \subseteq I$. We then apply Lemma 1 of Kamada and Kojima (2015) to obtain the result. □

Corollary 2. *Assume that $\eta = \{\mathrm{Ch}^1, \ldots, \mathrm{Ch}^K\}$ is monotonic and minimal, and that $\mathrm{Ch}^k_s(\cdot)$ satisfies substitutability and the law of aggregate demand for all k and all s. Let ν^k be the tentative matching at the end of stage k of the GDQDA algorithm (and hence, by Theorem 6, ν^k is also the tentative matching at the end of stage k of GSDA). No student is unmatched under ν^k for any $k = 1, \ldots, K$.*

Proof. By the same argument used to show Theorem 7, ν^k Pareto dominates ν^K. The matching under Ch^K, denoted ν^K, leaves no student unmatched (since Ch^K ensures a feasible match), so $\nu^K(i) \neq \varnothing$ for all $i \in I$. Assume that $\nu^k(i) = \varnothing$ for some $i \in I$. Because all students prefer every school to being unmatched, we have $\nu^K(i)\ P_i\ \nu^k(i)$. However, this contradicts that ν^k Pareto dominates ν^K. □

Appendix B: Proofs from the main text

Proofs of any lemmas not given here can be found in Appendix C.

Proof of Theorem 1

Consider any feasible matching $\mu \in \mathcal{M}_f$, and define $\bar{Q}_s = |\mu(s)|$ and $\bar{U}_{s,\theta} = |\mu_\theta(s)|$ for all (s, θ). Note that $\sum_{s' \in S} \bar{Q}_{s'} = n$, $\sum_{s' \in S} \bar{U}_{s',\theta} = |I_\theta|$, and $\sum_{\theta' \in \Theta} \bar{U}_{s,\theta'} = \bar{Q}_s$ for all s and all θ. We show that $(\bar{U}, \bar{Q})$ ensures a feasible match. Let P_I be an arbitrary profile of preferences, and let ν denote the final assignment produced by DA under preferences P_I. It is obvious that $|\nu_\theta(s)| \le \bar{U}_{s,\theta} \le U_{s,\theta}$ and $|\nu(s)| \le \bar{Q}_s \le Q_s$ by definition. What remains to be shown is that every student is assigned a school (i.e., no student is unmatched) and $|\nu_\theta(s)| \ge L_{s,\theta}$.

First, assume that some student i is unmatched. An equivalent way to run the deferred acceptance algorithm is to use the cumulative offer process of Hatfield and Milgrom (2005) (see the description of GDQDA in Appendix A). As shown in the proof of Lemma 2 below, the school choice functions for each school induced by $(\bar{U}, \bar{Q})$, denoted $\overline{\mathrm{Ch}}_s(\cdot)$, are substitutable and satisfy the law of aggregate demand. By definition, $\nu(s) = \overline{\mathrm{Ch}}_s(\mathcal{A}_s(T))$, where $\mathcal{A}_s(T)$ denotes the cumulative set of offers received by school s at the end of the cumulative offer process.

Let $\tau(i) = \theta$, and for all θ', let $\nu_{\theta'}^{\text{open}}(s)$ be the number of type θ' students who are assigned to a school s through its open seats. Since we assume that i is unmatched, we have $i \in \mathcal{A}_s(T)$ for all s (i.e., i has applied to every school) and $i \notin \nu(s)$ for all s. Since i is not chosen by any school, one of the following (or both) must hold at every school s: (i) $|\nu_\theta(s)| = \bar{U}_{s,\theta}$ or (ii) the type θ seats at s are all filled with $L_{s,\theta}$ other type θ students and the open seats of s are filled with $\bar{Q}_s - \sum_{\theta' \in \Theta} L_{s,\theta'}$ students of any type (i.e., $\sum_{\theta'} \nu_{\theta'}^{\text{open}}(s) = \bar{Q}_s - \sum_{\theta' \in \Theta} L_{s,\theta'}$). Note first that it cannot be true that $|\nu_\theta(s)| = \bar{U}_{s,\theta}$ for all s. If this were to hold, then $\sum_{s \in S} |\nu_\theta(s)| = \sum_{s \in S} \bar{U}_{s,\theta} = |I_\theta|$, but since i is not matched, at most $|I_\theta| - 1$ type θ students are assigned under ν. Therefore, there must be at least one school for which (i) does not hold and, therefore, (ii) does hold.

Thus, let $\hat{s}$ be such a school, where $\sum_{\theta'} |\nu_{\theta'}^{\text{open}}(\hat{s})| = \bar{Q}_{\hat{s}} - \sum_{\theta' \in \Theta} L_{\hat{s},\theta'}$ holds but $|\nu_\theta(\hat{s})| < \bar{U}_{\hat{s},\theta}$. By construction, at school $\hat{s}$, for every type θ', $|\nu_{\theta'}^{\text{open}}(\hat{s})| \le \bar{U}_{\hat{s},\theta'} - L_{\hat{s},\theta'}$. Because $\sum_{\theta' \in \Theta} \bar{U}_{s,\theta'} = \bar{Q}_s$, the only way for the $\bar{Q}_{\hat{s}} - \sum_{\theta' \in \Theta} L_{\hat{s},\theta'}$ open seats at $\hat{s}$ to be filled is if $|\nu_{\theta'}^{\text{open}}(\hat{s})| = \bar{U}_{\hat{s},\theta'} - L_{\hat{s},\theta'}$ for all θ'.[41] If $\bar{U}_{\hat{s},\theta} > L_{\hat{s},\theta}$, then $|\nu_\theta(\hat{s})| = \bar{U}_{\hat{s},\theta}$ (because all type θ floor seats must be assigned before any open seats can go to type θ students). If $\bar{U}_{\hat{s},\theta} = L_{\hat{s},\theta}$, then once again, $|\nu_\theta(\hat{s})| = \bar{U}_{\hat{s},\theta}$ (because otherwise, one of the type θ floor seats would be empty and i would have been accepted to s). In either case, $|\nu_\theta(\hat{s})| = \bar{U}_{\hat{s},\theta}$, which is a contradiction.

Now that we know all students are matched to a school under ν, we can show that all floors are satisfied. Assume not, i.e., assume there is some pair (s, θ) such that $|\nu_\theta(s)| < L_{s,\theta}$. Then $\sum_{s' \in S} |\nu_\theta(s')| < \sum_{s' \in S} \bar{U}_{s',\theta} = |I_\theta|$.[42] But this implies that under ν, there exists some student i of type θ who is not assigned to any school, which is a contradiction to what we just showed in the preceding paragraph.

Proof of Theorem 2

Lemma 2 below shows that for any stage k of DQDA, the induced within-stage choice functions of the schools satisfy substitutability and the law of aggregate demand. Since

[41]If $|\nu_{\theta'}^{\text{open}}(\hat{s})| < \bar{U}_{\hat{s},\theta'} - L_{\hat{s},\theta'}$ for some θ', then we can sum over θ' to get $\sum_{\theta'} |\nu_{\theta'}^{\text{open}}(\hat{s})| < \sum_{\theta'} (\bar{U}_{\hat{s},\theta'} - L_{\hat{s},\theta'}) = \bar{Q}_s - \sum_{\theta'} L_{\hat{s},\theta'}$, which contradicts that the open seats are filled.

[42]The first inequality follows from the fact that $\sum_{s' \in S \setminus \{s\}} |\nu_\theta(s')| \le \sum_{s' \in S \setminus \{s\}} \bar{U}_{s',\theta}$ (because ν respects $(\bar{U}, \bar{Q})$ by construction) and that $|\nu_\theta(s)| < L_{s,\theta} \le \bar{U}_{s,\theta}$.

ACDA is equivalent to DA using the stage K choice functions, strategyproofness follows from Hatfield and Milgrom (2005), who show that substitutability and the law of aggregate demand are sufficient for strategyproofness.

To show that ACDA eliminates justified envy among same types, note that ACDA is equivalent to DA under (U^K, Q^K). Denote the matching produced by DA under (U^K, Q^K) as μ. Assume that some student i envies another student j of her same type: $\mu(j)\ P_i\ \mu(i)$ and $\tau(i) = \tau(j) = \theta$. Let step t be the step of the DA algorithm at which i is rejected from $\mu(j)$. In step t, i is rejected because the type θ specific seats are filled with $L_{\mu(j),\theta}$ students of type θ ranked higher than i according to $\succ_{\mu(j)}$, and the open seats are filled with either (i) $U^K_{\mu(j),\theta} - L_{\mu(j),\theta}$ students of type θ ranked higher than i according to $\succ_{\mu(j)}$ or (ii) $Q^K_{\mu(j)} - \sum_{\theta\in\Theta} L_{\mu(j),\theta}$ students of any type ranked higher than i according to $\succ_{\mu(j)}$. As the algorithm progresses, a student accepted in step t can be rejected from the type θ specific seats only if a higher ranked student of type θ applies, and the same is true of the students at the open seats. Thus, at the end of the algorithm, all students assigned to $\mu(j)$ through either the type θ specific seats or the open seats must be ranked higher than i. Since $\tau(j) = \theta$ as well, this implies that $j \succ_{\mu(j)} i$, i.e., i does not justifiably envy j.[43]

Proof of Theorem 3

We first note the following lemma, which is proved in Appendix C. With slight abuse of notation, let $\eta = \{\mathrm{Ch}^1, \dots, \mathrm{Ch}^K\}$ be the sequence of choice functions induced by a reduction sequence $\{(U^1, Q^1), \dots, (U^K, Q^K)\}$ as in Definition 1.

Lemma 2. *The reduction sequence $\eta = \{\mathrm{Ch}^1, \dots, \mathrm{Ch}^K\}$ is monotonic, and $\mathrm{Ch}^k_s(\cdot)$ satisfies substitutability and the law of aggregate demand for all k and s. If, in addition, $\{(U^1, Q^1), \dots, (U^K, Q^K)\}$ is minimal in the sense of Definition 2, then $\eta = \{\mathrm{Ch}^1, \dots, \mathrm{Ch}^K\}$ is minimal in the sense of Definition 7.*

Given Lemma 2, Theorem 3 then is a special case of Corollary 1.

Proof of Theorem 4

Consider an ACDA mechanism with corresponding caps denoted (U^K, Q^K), and consider any nondegenerate reduction sequence $\{(U^1, Q^1), \dots, (U^K, Q^K)\}$ that is minimal (in the sense of Definition 2).[44] Let $\eta = \{\mathrm{Ch}^1, \dots, \mathrm{Ch}^K\}$ be the reduction sequence of induced choice functions. Note that by Lemma 2, η is monotonic, is minimal (in the sense of Definition 7), and $\mathrm{Ch}^k_s(\cdot)$ satisfies substitutability and the law of aggregate demand for all k and s.

[43] Of course, there may be students of other types $\theta' \neq \theta$ whom i envies and has higher priority over, since these students could be assigned through the type θ' specific seats.

[44] Recall that our model considers only markets where such nondegenerate reduction sequences exist. When they do not, the problem is trivial and uninteresting (see footnote 24).

Part (i). That all students weakly prefer the outcome of DQDA under η to ACDA under (U^K, Q^K) for all P_I follows from Lemma 2 and Theorem 7. It is simple to construct a preference profile that exhibits a strict improvement for some student under DQDA.

Part (ii). By Corollary 1, the final matching produced by DQDA is equivalent to SDA, which in turn is equivalent to DA under Ch^k for some $k \leq K$. That DA under Ch^k eliminates justified envy among same types can be shown using the same argument as in the proof of Theorem 2.

Part (iii). Fix the reports of the other students at P_{-i}, and let i's true preferences be P_i. We show that there is no report P_i' that gives i a better assignment than reporting the truth.

In the proof, we move back and forth between the GDQDA and GSDA algorithms, which, by Corollary 1, are equivalent. We begin by working with GSDA, and make use of the fact that the DA algorithm within each stage k is equivalent to the cumulative offer process of Hatfield and Milgrom (2005). It is well known (McVitie and Wilson 1971; see also Dubins and Freedman 1981 or Hatfield and Kojima 2010) that the order in which students are allowed to apply does not matter, as any such order will lead to the same final matching. Consider an ordering in which student i applies last. In stage k of the GSDA algorithm, let $\tilde{\mathcal{B}}_s^k$ denote the cumulative set of applicants that school s receives in the cumulative offer process under Ch^k on all students other than i.

Now, let i enter the market. This causes a rejection chain, which simply records the action of the algorithm:[45]

Step	Action	Cumulative offer sets
0	None	$\mathcal{B}_{\hat{s}}^k(0) = \tilde{\mathcal{B}}_{\hat{s}}^k$ for all $\hat{s}$
1	Student i applies to school s	$\mathcal{B}_s^k(1) = \mathcal{B}_s^k(0) \cup \{i\}$ $\mathcal{B}_{\hat{s}}^k(1) = \mathcal{B}_{\hat{s}}^k(0)$ for all $\hat{s} \neq s$
2	s rejects i'	$\mathcal{B}_{\hat{s}}^k(2) = \mathcal{B}_{\hat{s}}^k(1)$ for all $\hat{s}$
3	i' applies to s'	$\mathcal{B}_{s'}^k(3) = \mathcal{B}_{s'}^k(2) \cup \{i'\}$ $\mathcal{B}_{\hat{s}}^k(3) = \mathcal{B}_{\hat{s}}^k(2)$ for all $\hat{s} \neq s'$
$\vdots$	$\vdots$	$\vdots$

The rejection chain in stage k ends at the first step $\hat{T}^k$ for which some student i' applies to some school s' and s' does not reject a student. At this point, the applicant pools for each school are $\mathcal{B}_s^k(\hat{T}^k)$ and, as described above, the stage k output μ^k is defined by $\mu^k(s) = \mathrm{Ch}_s^k(\mathcal{B}_s^k(\hat{T}^k))$. Note that by substitutability, if $i \in \mathrm{Rej}_s^k(\mathcal{B}_s^k(t))$ for some step t, then $i \in \mathrm{Rej}_s^k(\mathcal{B}_s^k(t'))$ for all $t' \geq t$. In particular, $i \in \mathrm{Rej}_s^k(\mathcal{B}_s^k(\hat{T}^k))$, which implies that under μ^k, each student is assigned to exactly one school (by Corollary 2, no student is unmatched). Also, note that the law of aggregate demand guarantees at each rejection step, at most one student is rejected, which implies that in each application step, there is a unique student i' who applies to the next school on his preference list s'.

[45]In the above description of the cumulative offer process, both the application and the rejection phases occurred in the same step. Here, we have broken the steps up into rejection steps and application steps for clarity. This changes the numbering of the steps, but does not affect the algorithm or the results in any other way.

For any school s and applicant pool $\mathcal{B} \subseteq I$, define $\delta_s(\mathcal{B}) = \sum_{\theta\in\Theta} \max\{L_{s,\theta} - |\mathcal{B} \cap I_\theta|, 0\}$. In words, $\delta_s(\mathcal{B})$ is the number of floor seats unfilled at s when its applicant pool is $\mathcal{B}$.[46]

Let $\Delta^k = \sum_{s\in S} \delta_s(\tilde{\mathcal{B}}_s^k)$. In words, Δ^k is the total number of floor seats unfilled at all schools in stage k after all students but i have applied.

Lemma 3. *The following statements hold for all k:*

(i) If $\Delta^k = 0$, then $\mathrm{DA}^{\mathrm{Ch}^k}(P_i', P_{-i}) \in \mathcal{M}_f$ for all $P_i' \in \mathcal{P}$.

(ii) If $\Delta^k > 1$, then $\mathrm{DA}^{\mathrm{Ch}^k}(P_i', P_{-i}) \notin \mathcal{M}_f$ for all $P_i' \in \mathcal{P}$.

(iii) We have $\Delta^k \geq \Delta^{k+1} \geq \Delta^k - 1$.

If $\Delta^1 = 0$, then all floor seats have been filled before i enters the market in stage 1, and the GDQDA algorithm ends in stage 1 for any report P_i' of agent i. Therefore, from the perspective of agent i, the mechanism is equivalent $\mathrm{DA}^{\mathrm{Ch}^1}$ which is known to be strategyproof, and so she has no profitable manipulation. So assume that $\Delta^1 \geq 1$. Lemma 3, part (iii) then implies that there is some critical stage k^* for which $\Delta^{k^*} = 1$ and $\Delta^{k'} > 1$ for all $k' < k^*$. By Lemma 3, part (ii), the algorithm does not end in stage $k' < k^*$ for any report of student i. By the construction of GSDA, we can ignore the first $k^* - 1$ stages of the algorithm, and begin at stage k^* using choice functions Ch^{k^*}.

From here forward we switch from GSDA and work with the GDQDA algorithm definition. Starting at k^*, let $\{\tilde{\mathcal{A}}_s^{k^*}\}_{s\in S}$ be the cumulative offer sets of the schools after running DA under Ch^{k*} on all students other than i.[47] Then let i enter the market with some reported preferences P_i'. As above, her entering again causes a rejection chain, an example of which is

Stage	Step	Action	Cumulative offer sets
k^*	0	None	$\mathcal{A}_{\hat{s}}^{k^*}(0) = \tilde{\mathcal{A}}_{\hat{s}}^{k^*}$ for all $\hat{s}$
	1	Student i applies to school s	$\mathcal{A}_s^{k^*}(1) = \mathcal{A}_s^{k^*}(0) \cup \{i\}$ $\mathcal{A}_{\hat{s}}^{k^*}(1) = \mathcal{A}_{\hat{s}}^{k^*}(0)$ for all $\hat{s} \neq s$
	2	s rejects i'	$\mathcal{A}_{\hat{s}}^{k^*}(2) = \mathcal{A}_{\hat{s}}^{k^*}(1)$ for all $\hat{s}$
	3	i' applies to s'	$\mathcal{A}_{s'}^{k^*}(3) = \mathcal{A}_{s'}^{k^*}(2) \cup \{i'\}$ $\mathcal{A}_{\hat{s}}^{k^*}(3) = \mathcal{A}_{\hat{s}}^{k^*}(2)$ for all $\hat{s} \neq s'$
$k^* + 1$	1	Choice functions become Ch^{k^*+1}	$\mathcal{A}_{\hat{s}}^{k^*+1}(0) = \mathcal{A}_{\hat{s}}^{k^*}(3)$ for all $\hat{s}$
	2	School s'' rejects student i''	$\mathcal{A}_{\hat{s}}^{k^*+1}(1) = \mathcal{A}_{\hat{s}}^{k^*+1}(0)$ for all $\hat{s}$
	$\vdots$	$\vdots$	$\vdots$

[46]Recall that at every stage k, each school s always reserves $L_{s,\theta}$ seats for students of type θ, and so if $|\mathcal{B} \cap I_\theta| \leq L_{s,\theta}$, then all type θ students will be chosen, while if $|\mathcal{B} \cap I_\theta| > L_{s,\theta}$, at least $L_{s,\theta}$ students of type θ will be chosen.

[47]These are equivalent to $\{\tilde{\mathcal{B}}_s^{k^*}\}_{s\in S}$, the cumulative offer sets of standard DA under choice functions Ch^{k^*} (by Lemma 1) on all students other than i. We switch the notation from $\mathcal{B}$ to $\mathcal{A}$ to emphasize the switch from GSDA to GDQDA and to remain consistent with the notation used above in their respective definitions.

As before, the rejection chain simply records the action of the algorithm. However, now note that because we are using the GDQDA algorithm, we include quota reductions as part of the rejection chain (stage $k^* + 1$, step 1 above).

There are several features to note about rejection chains. First, at stage k, step t of the rejection chain, the set of applicants being tentatively held by school s is $\mathrm{Ch}_s^k(\mathcal{A}_s^k(t))$. Second, monotonicity and substitutability guarantee that if at some stage and step (k, t) we have $i \in \mathrm{Rej}_s^k(\mathcal{A}_s^k(t))$, then $i \in \mathrm{Rej}_s^{k'}(\mathcal{A}_s^{k'}(t'))$ for all (k', t') such that either (i) $k' > k$ or (ii) $k' = k$ and $t' \geq t$. This, together with the law of aggregate demand and minimality, guarantees that at each step, there is *at most* one student who is not currently being held by any school, and it is this student who makes the next application according to her preferences.[48] Third, within a stage, the school rejecting a student at some step is the same as the last school to receive an application (since the cumulative offer sets of the other schools have not changed). However, across stages, this may not be true (because the school that (potentially) must reject a student at the beginning of a stage is determined by the reduction sequence). For example, at stage k^*, step 3, s' receives an application. When the choice functions become Ch^{k^*+1}, school s'' rejects a student, but it may be that $s'' \neq s'$.

At stage k^*, $\Delta^{k^*} = 1$ (by definition of k^*), which means that $\delta_y(\tilde{\mathcal{A}}_y^{k^*}) = 1$ for some school y and $\delta_s(\tilde{\mathcal{A}}_s^{k^*}) = 0$ for all $s \neq y$. By definition of $\delta_s(\cdot)$, it must be that $|\tilde{\mathcal{A}}_y^{k^*} \cap I_\phi| = L_{y,\phi} - 1$ for some type $\phi \in \Theta$, and $|\tilde{\mathcal{A}}_s^{k^*} \cap I_\theta| \geq L_{s,\theta}$ for all $(s, \theta) \neq (y, \phi)$. In words, this means that every school has enough students in its choice set to fill all type-specific floors except for school y, which is one student short of filling its type ϕ floor.

We next note the following important fact about rejection chains:

$$\text{The GDQDA algorithm ends the next time a type } \phi \text{ student applies to } y. \tag{1}$$

This is an "if and only if" statement: the algorithm ends at the next stage step (k, t) at which a type ϕ student applies to y, and cannot end earlier.[49]

The next part of the proof is inspired by the scenario lemma of Dubins and Freedman (1981). Define a *scenario* $\mathbb{S}_i$ as a sequence of applications for agent i, i.e., a partial rank ordering over S for student i. So a scenario could be $\mathbb{S}_i = \{s, u, v\}$, which means that i first applies to s, then to u, then to v. The list *need not* include all schools. Since the preferences of the other students are fixed at P_{-i}, each scenario induces a corresponding rejection chain. We use $\mathcal{R}(\mathbb{S}_i)$ to denote the rejection chain corresponding to scenario $\mathbb{S}_i$. The rejection begins with i applying to the first school in $\mathbb{S}_i$, and then records all subsequent applications, rejections, and quota reductions. The rejection chain for any scenario $\mathbb{S}_i$ ends in one of two ways:

[48] The law of aggregate demand guarantees this within a stage, while minimality guarantees it across stages.

[49] The algorithm obviously cannot end *before* (1) occurs. To see that it ends immediately once (1) occurs, note that by definition of the school choice functions, once a school s fills a type-specific floor, it never drops below it, because all schools s always reserve $L_{s,\theta}$ seats for each type θ. Thus, once (1) occurs, the current stage ends (because no further student is rejected from s), and all floors are filled at all schools, so the algorithm ends.

(i) Some student j of type ϕ applies to school y:

Stage	Step	Action	Cumulative offer sets
k	t	j applies to y	$\mathcal{A}_y^k(t) = \mathcal{A}_y^k(t-1) \cup \{j\}$ $\mathcal{A}_{\hat{s}}^k(t) = \mathcal{A}_{\hat{s}}^k(t-1)$ for all $\hat{s} \neq s$

(ii) Agent i is rejected by the last school in $\mathbb{S}_i$:

Stage	Step	Action	Cumulative offer sets
k	t	i is rejected by v	$\mathcal{A}_s^k(t) = \mathcal{A}_s^k(t-1)$ for all s

The following lemma is key to the remainder of the proof.

Lemma 4. *Consider two scenarios $\mathbb{S}_i$ and $\hat{\mathbb{S}}_i$ such that every school in $\hat{\mathbb{S}}_i$ is also named in $\mathbb{S}_i$ (order is immaterial), and assume that in $\mathcal{R}(\mathbb{S}_i)$, student i applies to every school in $\mathbb{S}_i$. Then every step in $\mathcal{R}(\hat{\mathbb{S}}_i)$ also occurs at some point in $\mathcal{R}(\mathbb{S}_i)$.*

With this lemma in hand, we can continue the proof. Suppose, without loss of generality, that i's true preferences are P_i: $s_1, s_2, \ldots, s_m$. Say that if i submits her true preferences, she receives school s_h, and suppose that there is some scenario $\hat{\mathbb{S}}_i = \{u, \ldots, v\}$, where i gets some school v such that $v\, P_i\, s_h$. Then rejection chain $\mathcal{R}(\hat{\mathbb{S}}_i)$ must end with some student j of type ϕ applying to y, while i is assigned to v.

Case (i): $s\, P_i\, v$ for all $s \in \hat{\mathbb{S}}_i \setminus \{v\}$. Compare $\hat{\mathbb{S}}_i$ to a scenario $\mathbb{S}_i = \{s_1, \ldots, s_{h-1}\}$. By assumption, $\hat{\mathbb{S}}_i \subseteq \mathbb{S}_i$, and in $\mathcal{R}(\mathbb{S}_i)$, i makes every application in $\mathbb{S}_i$.[50] In particular, the last step of $\mathcal{R}(\mathbb{S}_i)$ is "s_{h-1} rejects i."

By Lemma 4, every application in $\mathcal{R}(\hat{\mathbb{S}}_i)$ is also made in $\mathcal{R}(\mathbb{S}_i)$. In particular, j must also apply to y in $\mathcal{R}(\mathbb{S}_i)$, which contradicts the fact that $\mathcal{R}(\mathbb{S}_i)$ ends with i being rejected by s_{h-1}.

Case (ii): $v\, P_i\, s$ for at least one $s \in \hat{\mathbb{S}}_i$. Delete all schools $s \in \hat{\mathbb{S}}_i$ such that $v\, P_i\, s$ to create a smaller scenario $\hat{\hat{\mathbb{S}}}_i \subset \hat{\mathbb{S}}_i$. By case (i), $\mathcal{R}(\hat{\hat{\mathbb{S}}}_i)$ must end with i rejected by v. Since $\hat{\hat{\mathbb{S}}}_i \subset \hat{\mathbb{S}}_i$, Lemma 4 implies that i must also be rejected by v in $\mathcal{R}(\hat{\mathbb{S}}_i)$, which is a contradiction.

Proof of Theorem 5

Since (U^K, Q^K) ensures a feasible match under Ψ, we know that DQ^Ψ ends no later than stage K for every P_I; that is, $\mathrm{DQ}^\Psi(P_I) = \psi^{(U^k, Q^k)}(P_I)$ for some $k \le K$. Since η is a minimal reduction sequence, $\sum_{\theta \in \Theta}(U_{s,\theta}^k - U_{s,\theta}^K) \le Q_s^k - Q_s^K$ for all $s \in S$. By resource monotonicity, $\mathrm{DQ}_i^\Psi(P_I) = \psi_i^{(U^k, Q^k)}(P_I) R_i \psi_i^{(U^K, Q^K)}(P_I)$ for all $i \in I$. Since this holds for all P_I, the result follows.

Appendix C: Omitted proofs of lemmas

Proof of Lemma 1

We first start with two sub-lemmas.

[50] This follows because i receives s_h when he submits his true preferences.

Lemma 5. *If $k' \geq k$, then $\mathcal{B}_s^k(\hat{T}^k) \subseteq \mathcal{B}_s^{k'}(\hat{T}^{k'})$ for all s.*

Proof. By monotonicity, $\mathrm{Ch}_s^{k'}(I') \subseteq \mathrm{Ch}_s^k(I')$ for all $s \in S$ and all $I' \subseteq I$. The statement then follows from Lemma 1, part (2) of Kamada and Kojima (2015).[51] □

The next sub-lemma makes use of the following definition of stability. Note that this is only a technical definition used to prove the results below and is not related to the justified envy definitions used in the main text. Let $\mathrm{Ch}' := \{\mathrm{Ch}'_{s_1}, \dots, \mathrm{Ch}'_{s_m}\}$ be a vector of choice functions (which again need not be equal to the primitive Ch).

Definition 8. A matching μ is *stable with respect to* Ch$'$ if the following statements hold:

(i) We have $\mu(s) = \mathrm{Ch}'_s(\mu(s))$ for all $s \in S$.

(ii) There exists no pair (i, s) such that $s \, P_i \, \mu(i)$ and $i \in \mathrm{Ch}'_s(\mu(s) \cup \{i\})$.

Part (i) says that a school does not unilaterally reject any student assigned to it. The corresponding "individual rationality" property holds for students automatically, because we assume that all students find all schools acceptable. Given this definition, we have the following lemma.

Lemma 6. *Under ν^k, each student is assigned to at most one school, and ν^k is stable with respect to Ch^k.*

Proof. For the first part, note that within a stage k, substitutability implies $\mathrm{Rej}_s^k(\mathcal{A}_s^k(t-1)) \subseteq \mathrm{Rej}_s^k(\mathcal{A}_s^k(t))$ for all $t = 1, \dots, T^k$, and across stages, monotonicity ensures that $\mathrm{Rej}_s^{k-1}(\mathcal{A}_s^{k-1}(T^{k-1})) \subseteq \mathrm{Rej}_s^k(\mathcal{A}_s^k(0))$ for all s. Thus, if i is rejected by a school s at some step t of some stage k, then i is rejected in all later steps and stages, implying that no student is assigned to more than one school. For stability, first note that by irrelevance of rejected students (see footnote 34), $\mathrm{Ch}_s^k(\mathcal{A}_s^k(T^k)) = \nu^k(s) = \mathrm{Ch}_s^k(\nu^k(s))$. For the second part, if $sP_i\nu^k(i)$, then i was at some point rejected from s. This implies that $i \in \mathcal{A}_s^k(T^k)$ but $i \notin \nu^k(s)$. This means that $\nu^k(s) = \mathrm{Ch}_s^k(\mathcal{A}_s^k(T^k)) \subseteq \nu^k(s) \cup \{i\} \subseteq \mathcal{A}_s^k(T^k)$, which, together with irrelevance of rejected students, implies that $\mathrm{Ch}_s^k(\nu^k(s) \cup \{i\}) = \nu^k(s)$, i.e., $i \notin \mathrm{Ch}_s^k(\nu^k(s) \cup \{i\})$. Therefore, ν^k is stable with respect to Ch^k. □

We now use induction to show that

$$\mathcal{A}_s^k(T^k) = \mathcal{B}_s^k(\hat{T}^k) \quad \text{for all } s \in S \tag{2}$$

holds for all k. Since stage 1 of both algorithms is just the cumulative offer process starting with the empty matching and using choice functions Ch^1, (2) holds for $k = 1$. Assume the inductive hypothesis that (2) holds for $1, \dots, k-1$. We show that this implies it holds for k as well.

[51] See also the "capacity lemma" of Konishi and Ünver (2006) for a related result.

First, note that since stage k of GSDA is simply the DA algorithm under Ch^k, we know that μ^k is the student-optimal stable match with respect to Ch^k (Hatfield and Milgrom 2005). Further, by Lemma 6, ν^k is *some* match that is stable with respect to Ch^k. This implies that $\mu^k(i)\, R_i\, \nu^k(i)$ for all $i \in I$. This implies that $\mathcal{B}_s^k(\hat{T}^k) \subseteq \mathcal{A}_s^k(T^k)$ for all $s \in S$.[52]

Last, we must show that $\mathcal{A}_s^k(T^k) \subseteq \mathcal{B}_s^k(\hat{T}^k)$ for all $s \in S$. Assume to the contrary, i.e., that there exists some s such that $\mathcal{A}_s^k(T^k) \nsubseteq \mathcal{B}_s^k(\hat{T}^k)$, and let t' be the first step of stage k of GDQDA such that $\mathcal{A}_s^k(t) \subseteq \mathcal{B}_s^k(\hat{T}^k)$ for all $t < t'$ and all $s \in S$, but $\mathcal{A}_s^k(t') \nsubseteq \mathcal{B}_s^k(\hat{T}^k)$.[53] Let i be the student who applies to s at step t'. This means that i is rejected from $\mu^k(i)$ (the school she is matched to under GSDA) in some step of stage k of GDQDA; let the earliest of these steps be t'', so that $i \in \mathrm{Rej}^k_{\mu^k(i)}(\mathcal{A}^k_{\mu^k(i)}(t''))$.[54] Further, note that $t'' < t'$, which, by the definition of t', implies that $\mathcal{A}^k_{\mu^k(i)}(t'') \subseteq \mathcal{B}^k_{\mu^k(i)}(\hat{T}^k)$. Substitutability of the choice functions within stage k then implies that $\mathrm{Rej}^k_{\mu^k(i)}(\mathcal{A}^k_{\mu^k(i)}(t'')) \subseteq \mathrm{Rej}^k_{\mu^k(i)}(\mathcal{B}^k_{\mu^k(i)}(\hat{T}^k))$, which means $i \in \mathrm{Rej}^k_{\mu^k(i)}(\mathcal{B}^k_{\mu^k(i)}(\hat{T}^k))$, which contradicts the fact that i is assigned to school $\mu^k(i)$ under GSDA in stage k.

Proof of Lemma 2

Substitutability. Consider a school s, stage k, and set of students I' such that $i \in \mathrm{Rej}_s^k(\mathcal{A}')$, and another set of students $\mathcal{A}''$ such that $\mathcal{A}' \subseteq \mathcal{A}''$. Let $\tau(i) = \theta$. When the set of applicants is $\mathcal{A}'$, student i is rejected because the type θ specific seats are filled with $L_{s,\theta}$ higher ranked type θ students, and the open seats are filled with either (i) $U^k_{s,\theta} - L_{s,\theta}$ higher ranked type θ students or (ii) $Q^k_s - \sum_{\theta \in \Theta} L_{s,\theta}$ higher ranked students of any type. In either case, since all students in $\mathcal{A}'$ are also in $\mathcal{A}''$, when school s is choosing from $\mathcal{A}''$, the type θ specific seats are once again filled with $L_{s,\theta}$ higher ranked type θ students, and either condition (i) or (ii) also still holds. So $i \in \mathrm{Rej}_s^k(\mathcal{A}'')$ as well.

Monotonicity. If s' is not the school whose quotas are reduced in moving from stage k to $k+1$, then $\mathrm{Rej}^k_{s'}(\mathcal{A}) = \mathrm{Rej}^{k+1}_{s'}(\mathcal{A})$ trivially. So let s be the school whose capacity is reduced in moving from stage k to $k+1$, $Q_s^{k+1} = Q_s^k - 1$ and $U^{k+1}_{s,\theta} = U^k_{s,\theta} - 1$, while $U^{k+1}_{s,\theta'} = U^k_{s,\theta'}$ for all other $\theta' \neq \theta$.

We want to show that $\mathrm{Rej}_s^k(\mathcal{A}) \subseteq \mathrm{Rej}_s^{k+1}(\mathcal{A})$ for all k and all $\mathcal{A} \subseteq I$. To do so, we show the contrapositive:

$$i \in \mathrm{Ch}_s^{k+1}(\mathcal{A}) \quad \Longrightarrow \quad i \in \mathrm{Ch}_s^k(\mathcal{A}).$$

Assume not, and let i be the highest ranked student according to $\succ_s$ such that $i \in \mathrm{Ch}_s^{k+1}(\mathcal{A})$, but $i \notin \mathrm{Ch}_s^k(\mathcal{A})$ (equivalently, $i \in \mathrm{Rej}_s^k(\mathcal{A})$). Let $\tau(i) = \theta'$, which may or may

[52]If this were not the case, then there exists some s such that $\mathcal{B}_s^k(\hat{T}^k) \nsubseteq \mathcal{A}_s^k(T^k)$. This means that in GSDA in stage k, some student i is rejected from $\nu^k(i)$ (her match under GDQDA). But this contradicts the fact that $\mu^k(i)\, R_i\, \nu^k(i)$.

[53]Such a t' exists because $\mathcal{A}_s^k(0) = \mathcal{A}_s^{k-1}(T^{k-1}) = \mathcal{B}_s^{k-1}(\hat{T}^{k-1}) \subseteq \mathcal{B}_s^k(\hat{T}^k)$ for all $s \in S$, where the first equality is by definition, the second is by the inductive hypothesis, and the set inclusion is by Lemma 5.

[54]Note that i must be rejected at some step t'' of stage k (and not in an earlier stage). To see this, assume that i was rejected from $\mu^k(i)$ in some earlier stage k' of GDQDA. This implies that $\mu^k(i)\, P_i\, \nu^{k'}(i)$. Since $k' < k$, Lemma 1, part (1) of Kamada and Kojima (2015) implies that $\mu^{k'}(i)\, R_i\, \mu^k(i)$. By combining these two inequalities, we conclude that $\mu^{k'}(i)\, P_i\, \nu^{k'}(i)$, which contradicts the inductive hypothesis.

not be equal to θ. If i is admitted through the type θ' specific seats in stage $k+1$, then she is one of the $L_{s,\theta'}$ highest ranked type θ' students in $\mathcal{A}$, and so she will be admitted in stage k as well. So i must be admitted through an open seat in stage $k+1$. Therefore, when i's application is considered in stage $k+1$, the following statements both hold: (a) at most $U_{s,\theta'}^{k+1} - L_{s,\theta'} - 1$ higher ranked type θ' students have been accepted to the open seats and (b) at most $Q_s^{k+1} - \sum_{\theta\in\Theta} L_{s,\theta} - 1$ students in total have been accepted to the open seats. Define $J = \{j \in \mathrm{Rej}_s^{k+1}(\mathcal{A}) : j \succ_s i\}$. Note that (b) implies that for all $j \in J$, the type-specific ceiling $U_{s,\tau(j)}^{k+1}$ is reached before j's application is considered in stage $k+1$, which, in particular, means that $\tau(j) \neq \theta'$ for all $j \in J$.

Since $Q_s^k = Q_s^{k+1} + 1$, for i to be rejected under stage k quotas, there must be two $j_1, j_2 \in J$ such that $j_1, j_2 \in \mathrm{Ch}_s^k(\mathcal{A})$; without loss of generality, let $j_1 \succ_s j_2 \succ_s i$. Since $j_1 \in \mathrm{Ch}_s^k(\mathcal{A})$, it must be that $\tau(j_1) = \theta$.[55] Then, after j_1 is admitted, the $U_{s,\theta}^k$ ceiling is now binding (since $U_{s,\theta}^k = U_{s,\theta}^{k+1} + 1$ and all $j \succ_s j_1$ who are admitted under stage $k+1$ quotas are also admitted under stage k quotas by the assumption that i is the highest ranked such student for which this is not the case). So when j_2's application is considered, she will be rejected, which is a contradiction.[56]

Minimality. As for monotonicity, we only need consider the school s whose quotas are reduced in moving from k to $k+1$. Let s be the school such that $Q_s^{k+1} = Q_s^k - 1$ and $U_{s,\theta}^{k+1} = U_{s,\theta}^k - 1$ for some θ, while $U_{s,\theta'}^{k+1} = U_{s,\theta'}^k$ for all $\theta' \neq \theta$. Consider a set of applicants $\mathcal{A} \subseteq I$, with $\mathrm{Ch}_s^k(\mathcal{A})$ the set that is admitted in stage k. There are four cases.

Case (i): $|\mathrm{Ch}_s^k(\mathcal{A})| < Q_s^k$ and $|\mathrm{Ch}_s^k(\mathcal{A}) \cap I_\theta| < U_{s,\theta}^k$. In this case, $\mathrm{Ch}_s^{k+1}(\mathcal{A}) = \mathrm{Ch}_s^k(\mathcal{A})$, which implies $|\mathrm{Ch}_s^k(\mathcal{A})| - |\mathrm{Ch}_s^{k+1}(\mathcal{A})| = 0$.

Case (ii): $|\mathrm{Ch}_s^k(\mathcal{A})| = Q_s^k$ and $|\mathrm{Ch}_s^k(\mathcal{A}) \cap I_\theta| < U_{s,\theta}^k$. Let i' be the lowest ranked student admitted through the open seats in stage k. Then $\mathrm{Ch}_s^{k+1}(\mathcal{A}) = \mathrm{Ch}_s^k(\mathcal{A}) \setminus \{i'\}$, which implies $|\mathrm{Ch}_s^k(\mathcal{A})| - |\mathrm{Ch}_s^{k+1}(\mathcal{A})| = 1$.

Case (iii): $|\mathrm{Ch}_s^k(\mathcal{A})| < Q_s^k$ and $|\mathrm{Ch}_s^k(\mathcal{A}) \cap I_\theta| = U_{s,\theta}^k$. Let i' be the lowest ranked type θ student admitted through the open seats in stage k. Then $\mathrm{Ch}_s^{k+1}(\mathcal{A}) = \mathrm{Ch}_s^k(\mathcal{A}) \setminus \{i'\}$, which implies $|\mathrm{Ch}_s^k(\mathcal{A})| - |\mathrm{Ch}_s^{k+1}(\mathcal{A})| = 1$.

Case (iv): $|\mathrm{Ch}_s^k(\mathcal{A})| = Q_s^k$ and $|\mathrm{Ch}_s^k(\mathcal{A}) \cap I_\theta| = U_{s,\theta}^k$. Let i' be the lowest ranked type θ student admitted through the open seats in stage k. Then $\mathrm{Ch}_s^{k+1}(\mathcal{A}) = \mathrm{Ch}_s^k(\mathcal{A}) \setminus \{i'\}$, which implies $|\mathrm{Ch}_s^k(\mathcal{A})| - |\mathrm{Ch}_s^{k+1}(\mathcal{A})| = 1$.

Law of aggregate demand. Within a stage, school s admits students one by one until either some type-specific ceiling $U_{s,\theta}^k$ or overall capacity Q_s^k has been reached. More students in the applicant pool clearly weakly increases the number of students admitted, and the law of aggregate demand is satisfied.

[55] If $\tau(j_1) \neq \theta$, then $U_{s,\tau(j_1)}^k = U_{s,\tau(j_1)}^{k+1}$. Then, since all $j \succ_s j_1$ such that $j \in \mathrm{Ch}_s^{k+1}(\mathcal{A})$ also satisfy $j \in \mathrm{Ch}_s^k(\mathcal{A})$ (by the assumption that i is the highest ranked student for which this is not the case), the ceiling for type $\tau(j_1)$ students will already have been reached in stage k when j_1's application is considered, and j_1 will be rejected.

[56] If $\tau(j_2) = \theta$, this follows from the previous sentence. If $\tau(j_2) \neq \theta$, it follows from the same argument as in footnote 55.

Proof of Lemma 3

We use the following facts about δ_s:

(a) We have $\mathcal{B} \subseteq \mathcal{B}' \implies \delta_s(\mathcal{B}') \leq \delta_s(\mathcal{B})$.

(b) If $\delta_s(\mathcal{B} \cup \{i\}) < \delta_s(\mathcal{B}) \implies \mathrm{Rej}_s^k(\mathcal{B} \cup \{i\}) = \mathrm{Rej}_s^k(\mathcal{B})$ for all k.

Fact (a) follows immediately from the definition of δ. Fact (b) follows because $\delta_s(\mathcal{B} \cup \{i\}) < \delta_s(\mathcal{B})$ implies that student i is of some type θ such that $|\mathcal{B} \cap I_\theta| < L_{s,\theta}$. But this means that when the applicant pool at school s is $\mathcal{B} \cup \{i\}$, student i is accepted through one of the type θ seats. This does not affect the students accepted by the type θ' seats for $\theta' \neq \theta$ or the open seats, and thus $\mathrm{Ch}_s^k(\mathcal{B} \cup \{i\}) = \mathrm{Ch}_s^k(\mathcal{B})$ or, equivalently, $\mathrm{Rej}_s^k(\mathcal{B} \cup \{i\}) = \mathrm{Rej}_s^k(\mathcal{B})$.

Part (i). Since $\Delta^k = 0$, the set $\tilde{\mathcal{B}}_s^k$ contains at least $L_{s,\theta}$ students of type θ for all schools s. Let $\mathcal{B}_s^k(\hat{T}^k)$ be the cumulative set of applicants to school s at the end of stage k. Since $\tilde{\mathcal{B}}_s^k \subseteq \mathcal{B}_s^k(\hat{T}^k)$ for any submitted preferences of student i, $\mathrm{Ch}^k(\mathcal{B}_s^k(\hat{T}^k))$ is a feasible assignment for school s, and the algorithm ends in stage k.

Part (ii). We can have $\Delta^k > 1$ in two ways: either $\delta_s(\tilde{\mathcal{B}}_s^k) > 1$ for some school or $\delta_s(\tilde{\mathcal{B}}_s^k) \geq 1$ for multiple schools. First, consider $\delta_s(\tilde{\mathcal{B}}_s^k) > 1$ for some school s. This means that there are at least two floor seats at s that are not yet filled because not enough students have applied to s. When student i enters the market in stage k, he causes a rejection chain. As described in the proof of Theorem 4, the rejection chain is such that in each application step, only one student makes an application. So stage k ends the *first* time a school gets an application from a student i' and does not reject an additional student. Since any time a floor seat is filled, no further student is rejected, at the end of stage k, at most one of the unfilled floor seats at s can be filled, and so the assignment of school s will still not be feasible.

The case where $\delta_s(\tilde{\mathcal{B}}_s^k) \geq 1$ for multiple schools is argued similarly.

Part (iii). By Theorem 6, $\tilde{\mathcal{B}}_s^{k+1}$ can equivalently be computed by starting with $\tilde{\mathcal{B}}_s^k$ and then reducing the choice functions to Ch^{k+1}. Doing so causes a rejection chain that ends the first time a student i' applies to a school s' and s' does not reject a student. Since $\tilde{\mathcal{B}}_s^k \subseteq \tilde{\mathcal{B}}_s^{k+1}$, we have $\delta_s(\tilde{\mathcal{B}}_s^{k+1}) \leq \delta_s(\tilde{\mathcal{B}}_s^k)$, which implies that $\Delta^k \geq \Delta^{k+1}$. To see that $\Delta^{k+1} \geq \Delta^k - 1$, note that in the rejection chain, the first time a student applies to a school and fills a floor, no further student is rejected (by fact (b)) and the rejection chain ends. Thus, at the end of the rejection chain, at most one floor seat that was not filled under $\tilde{\mu}^k$ can be filled under $\tilde{\mu}^{k+1}$, and so $\Delta^{k+1} \geq \Delta^k - 1$.

Proof of Lemma 4

We use the notation (k, t) to denote the line corresponding to stage k, step t of a rejection chain (note that we calculate the algorithm by starting at k^*, so in the remainder of this proof $k \geq k^*$ holds). Let $\mathcal{A}_s^k(t)$ denote the cumulative offer set of school s at line (k, t) of $\mathcal{R}(\mathbb{S}_i)$, and let $\hat{\mathcal{A}}_s^k(t)$ denote the corresponding set at line (k, t) of $\mathcal{R}(\hat{\mathbb{S}}_i)$. Similarly, let t^k denote the final step of stage k under scenario $\mathbb{S}_i$, and let $\hat{t}^k$ denote the final step of stage k under scenario $\hat{\mathbb{S}}_i$. Last, let k_{end} denote the last stage of $\mathcal{R}(\mathbb{S}_i)$ and let $\hat{k}_{\text{end}}$ denote the last stage of $\mathcal{R}(\hat{\mathbb{S}}_i)$.

We prove the result by induction on the line index (k, t). Line $(k^*, 1)$ of $\mathcal{R}(\hat{\mathbb{S}}_i)$ is "i applies to s," and this step occurs somewhere in $\mathcal{R}(\mathbb{S}_i)$ by assumption. So make the inductive assumption that all lines up to $(k, t-1)$ of $\mathcal{R}(\hat{\mathbb{S}}_i)$ also occur in $\mathcal{R}(\mathbb{S}_i)$. Then consider the next line in $\mathcal{R}(\hat{\mathbb{S}}_i)$. There are three cases:

Case (i): The next line (k, t) is an application line. Line (k, t) then reads "i' applies to s'." There are two cases. If $i' = i$, then this application also occurs in $\mathcal{R}(\mathbb{S}_i)$ by assumption. If $i' \neq i$, then let u be the school immediately before s' on the preference list of i'. Because (k, t) is an application line, $(k, t-1)$ must be a rejection line in which student i' is rejected by u. Since, by the inductive hypothesis, line $(k, t-1)$ occurs somewhere in $\mathcal{R}(\mathbb{S}_i)$, student i' must be rejected from u at some point in $\mathcal{R}(\mathbb{S}_i)$ and will then, according to his preferences, apply to s' in the following line.

Case (ii): The next line (k, t) is a rejection line. Line (k, t) then reads "s' rejects i'." Thus, student i' must have already applied to s', either before i entered the market or somewhere in rejection chain $\mathcal{R}(\hat{\mathbb{S}}_i)$. The choice function at s' when i' is rejected is $\mathrm{Ch}^k_{s'}$ and the set of cumulative applicants is $\hat{\mathcal{A}}^k_{s'}(t)$, which implies that $i' \in \mathrm{Rej}^k_{s'}(\hat{\mathcal{A}}^k_{s'}(t))$. By the inductive hypothesis, all students in $\hat{\mathcal{A}}^k_{s'}(t)$ also apply to s' under scenario $\mathbb{S}_i$; in other words, $\hat{\mathcal{A}}^k_{s'}(t) \subseteq \mathcal{A}^{k_{\text{end}}}_{s'}(T^{k_{\text{end}}})$. Further, the inductive hypothesis plus monotonic substitutability imply that $\mathrm{Rej}^k_{s'}(\hat{\mathcal{A}}^k_{s'}(t)) \subseteq \mathrm{Rej}^{k_{\text{end}}}_{s'}(\hat{\mathcal{A}}^k_{s'}(t)) \subseteq \mathrm{Rej}^{k_{\text{end}}}_{s'}(\mathcal{A}^{k_{\text{end}}}_{s'}(T^{k_{\text{end}}}))$.[57] So, i' must be rejected from s' at some point under scenario $\mathbb{S}_i$, i.e., line (k, t) must occur in $\mathcal{R}(\mathbb{S}_i)$.

Case (iii): The next line $(k+1, 1)$ is a choice function reduction line. In this case, line $(k+1, 1)$ reads "The choice functions become Ch^{k+1}." Assume to the contrary that this reduction does not occur under $\mathbb{S}_i$. Thus, $\mathcal{R}(\mathbb{S}_i)$ ends in stage k under Ch^k (by the inductive hypothesis, $\mathcal{R}(\mathbb{S}_i)$ reaches at least stage k). By Theorem 6, an alternative way to compute the outcome at the end of stage k under either scenario is to start with the empty matching and run the cumulative offer process under $\hat{\mathbb{S}}_i$ and $\mathbb{S}_i$. Recall that $\tilde{\mathcal{A}}^k_s$ is the cumulative set of applicants to school s before i enters the market (as before, since the preferences of all agents $-i$ do not change, this is the same under either case). Note that $\delta_s(\tilde{\mathcal{A}}^k_s) = 0$ for all $s \neq y$, and $\delta_y(\tilde{\mathcal{A}}^k_y) = 1$, where y has one type ϕ floor seat left to be filled (if the latter did not hold, all school choice sets would be feasible even before i enters, and the mechanism would not continue to stage $k+1$ under scenario $\hat{\mathbb{S}}_i$). Just as above, we can write a rejection chain for each scenario corresponding to running the cumulative offer process within stage k. Let $\mathcal{R}^k(\mathbb{S}_i)$ and $\mathcal{R}^k(\hat{\mathbb{S}}_i)$ denote these two rejection chains.

Since $\mathcal{R}(\mathbb{S}_i)$ ends in stage k, $\mathcal{R}^k(\mathbb{S}_i)$ must end in one of two ways:

(i) School y gets an application from some student j of type ϕ:

Stage	Step	Action	Offer sets
k	t	j applies to y	$\mathcal{A}^k_y(t) = \mathcal{A}^k_y(t-1) \cup \{j\}$ $\mathcal{A}^k_{\hat{s}}(t) = \mathcal{A}^k_{\hat{s}}(t-1)$ for all $\hat{s} \neq s$

[57]The inductive hypothesis implies that $k_{\text{end}} \geq k$, and so the first set inclusion follows by monotonicity; the second follows by substitutability and $\hat{\mathcal{A}}^k_{s'}(t) \subseteq \mathcal{A}^{k_{\text{end}}}_{s'}(T^{k_{\text{end}}})$.

(ii) Agent i is rejected by the last school in $\mathbb{S}_i$:

Stage	Step	Action	Offer sets
k	t	i is rejected by s	$\mathcal{A}_s^k(t) = \mathcal{A}_s^k(t-1)$ $\mathcal{A}_{\hat{s}}^k(t) = \mathcal{A}_{\hat{s}}^k(t-1)$ for all $\hat{s} \neq s$

Alternatively, $\mathcal{R}(\hat{\mathbb{S}}_i)$ does not end in stage k. This means that $\mathcal{R}^k(\hat{\mathbb{S}}_i)$ must end with

Stage	Step	Action	Offer sets
k	t	Student i' applies to school v	$\hat{\mathcal{A}}_v^k(t) = \hat{\mathcal{A}}_v^k(t-1) \cup \{i'\}$ $\hat{\mathcal{A}}_{\hat{s}}^k(t) = \hat{\mathcal{A}}_{\hat{s}}^k(t-1)$ for all $\hat{s} \neq s$

where either $v \neq y$ or i' is not of type ϕ.

Now, by the scenario lemma of Dubins and Freedman (1981), $\mathcal{R}^k(\mathbb{S}_i)$ must also end with a student applying to school v and not being rejected or filling the final floor seat. But, this contradicts (1) and (2) above.

References

Abdulkadiroğlu, Atila (2005), "College admissions with affirmative action." *International Journal of Game Theory*, 33, 535–549. [867, 869, 881]

Abdulkadiroğlu, Atila, Yeon-Koo Che, and Yusuke Yasuda (2011), "Resolving conflicting preferences in school choice: The "Boston" mechanism reconsidered." *American Economic Review*, 101, 399–410. [873]

Abdulkadiroğlu, Atila, Parag A. Pathak, Alvin E. Roth, and Tayfun Sönmez (2005), "The Boston Public School match." *American Economic Review: Papers and Proceedings*, 95, 368–371. [872]

Abdulkadiroğlu, Atila and Tayfun Sönmez (2003), "School choice: A mechanism design approach." *American Economic Review*, 93, 729–747. [865, 868, 871, 873, 875, 876, 884, 886, 888]

Afacan, Mustafa Oğuz (2013), "Filling position incentives in matching markets." Working paper, Sabanci University. [868]

Akbarpour, Mohammad, Shengwu Li, and Shayan Oveis Gharan (2016), "Thickness and information in dynamic matching markets." Unpublished paper. [868]

Akyol, Ethem (2013), "Welfare comparison of school choice mechanisms under incomplete information." Unpublished paper, Pennsylvania State University. [873]

Aygün, Orhan and Inácio Bó (2016), "College admission with multidimensional privileges: The Brazilian affirmative action case." Unpublished paper, WZB Berlin Social Science Center. [869]

Aygün, Orhan and Tayfun Sönmez (2013), "Matching with contracts: Comment." *American Economic Review*, 103, 2050–2051. [888]

Balinski, Michel and Tayfun Sönmez (1999), "A tale of two mechanisms: Student placement." *Journal of Economic Theory*, 84, 73–94. [872]

Bergemann, Dirk and Stephen Morris (2005), "Robust mechanism design." *Econometrica*, 73, 1771–1813. [872]

Biró, Peter, Tamás Fleiner, Robert W. Irving, and David F. Manlove (2010), "The college admissions problem with lower and common quotas." *Theoretical Computer Science*, 411, 3136–3153. [870]

Braun, Sebastian, Nadja Dwenger, Dorothea Kübler, and Alexander Westkamp (2014), "Implementing quotas in university admission: An experimental analysis." *Games and Economic Behavior*, 85, 232–251. [869]

Budish, Eric, Yeon-Koo Che, Fuhito Kojima, and Paul Milgrom (2013), "Designing random allocation mechanisms: Theory and applications." *American Economic Review*, 103, 585–623. [869]

Cowen Institute for Public Education Initiatives (2011), "Case studies of school choice and open enrollment in four cities." Technical report, Tulane University. [865]

Doğan, Battal (2016), "Responsive affirmative action in school choice." *Journal of Economic Theory*, 165, 69–105. [869]

Dubins, Lester E. and David A. Freedman (1981), "Machiavelli and the Gale–Shapley algorithm." *American Mathematical Monthly*, 88, 485–494. [867, 895, 897, 904]

Echenique, Federico and M. Bumin Yenmez (2015), "How to control controlled school choice." *American Economic Review*, 105, 2679–2694. [869, 870]

Ehlers, Lars, Isa E. Hafalir, M. Bumin Yenmez, and Muhammed A. Yildirim (2014), "School choice with controlled choice constraints: Hard bounds versus soft bounds." *Journal of Economic Theory*, 153, 648–683. [869, 870, 872, 876, 882, 884]

Ehlers, Lars and Bettina Klaus (2004), "Resource-monotonicity for house allocation problems." *International Journal of Game Theory*, 32, 545–560. [885]

Erdil, Aytek and Taro Kumano (2013), "Prioritizing diversity in school choice." Unpublished paper, University of Cambridge. [869]

Featherstone, Clayton and Muriel Niederle (2014), "Improving on strategyproof school choice mechanisms: An experimental investigation." Unpublished paper, Stanford University. [873]

Fleiner, Tamás (2003), "A fixed-point approach to stable matchings and some applications." *Mathematics of Operations Research*, 28, 103–126. [888]

Fragiadakis, Daniel, Atsushi Iwasaki, Peter Troyan, Suguru Ueda, and Makoto Yokoo (2015), "Strategyproof matching with minimum quotas." *ACM Transactions on Economics and Computation*, 4, 6.1–6.40. [869, 872, 882]

Gale, David and Lloyd S. Shapley (1962), "College admissions and the stability of marriage." *American Mathematical Monthly*, 69, 9–15. [866, 867, 873, 875, 884]

Gale, David and Marilda A. Oliveira Sotomayor (1985a), "Ms. Machiavelli and the stable matching problem." *American Mathematical Monthly*, 92, 261–268. [868]

Gale, David and Marilda A. Oliveira Sotomayor (1985b), "Some remarks on the stable matching problem." *Discrete Applied Mathematics*, 11, 223–232. [868]

Hafalir, Isa E., M. Bumin Yenmez, and Muhammed A. Yildirim (2013), "Effective affirmative action in school choice." *Theoretical Economics*, 8, 325–363. [868, 869, 870]

Hamada, Koki, Kazuo Iwama, and Shuichi Miyazaki (2011), "The hospitals/residents problem with quota lower bounds." In *Proceedings of the 19th Annual European Symposium on Algorithms (ESA 2011)*, 180–191, Springer, Berlin. [870]

Hatfield, John W. and Fuhito Kojima (2009), "Group incentive compatibility for matching with contracts." *Games and Economic Behavior*, 67, 745–749. [889]

Hatfield, John W. and Fuhito Kojima (2010), "Substitutes and stability for matching with contracts." *Journal of Economic Theory*, 145, 1704–1723. [867, 890, 892, 895]

Hatfield, John W. and Scott Duke Kominers (2011), "Contract design and stability in matching markets." Unpublished paper, Harvard University. [867]

Hatfield, John W. and Scott Duke Kominers (2012), "Matching in networks with bilateral contracts." *American Economic Journal: Microeconomics*, 4, 176–208. [867]

Hatfield, John W. and Paul R. Milgrom (2005), "Matching with contracts." *American Economic Review*, 95, 913–935. [867, 868, 881, 888, 889, 891, 892, 893, 894, 895, 900]

Kadam, Sangram V. and Maciej H. Kotowski (2014), "Multi-period matching." Unpublished paper, Harvard University. [868]

Kamada, Yuichiro and Fuhito Kojima (2015), "Efficient matching under distributional concerns: Theory and application." *American Economic Review*, 105, 67–99. [864, 866, 868, 872, 892, 899, 900]

Kelso, Alexander S. and Vincent P. Crawford (1982), "Job matching, coalition formation, and gross substitutes." *Econometrica*, 50, 1483–1504. [868]

Kennes, John, Daniel Monte, and Norovsambuu Tumennasan (2014a), "The day care assignment: A dynamic matching problem." *American Economic Journal: Microeconomics*, 6, 362–406. [868]

Kennes, John, Daniel Monte, and Norovsambuu Tumennasan (2014b), "Dynamic matching markets and the deferred acceptance mechanism." Unpublished paper, Dalhousie University. [868]

Kesten, Onur (2006), "On two competing mechanisms for priority-based allocation problems." *Journal of Economic Theory*, 127, 155–171. [885, 886]

Kojima, Fuhito (2012), "School choice: Impossibilities for affirmative action." *Games and Economic Behavior*, 75, 685–693. [868]

Kojima, Fuhito, Akihisa Tamura, and Makoto Yokoo (2014), "Designing matching mechanisms under constraints: An approach from discrete convex analysis." Unpublished paper, Munich Personal RePEc Archive. [869]

Kominers, Scott Duke and Tayfun Sönmez (2016), "Matching with slot-specific priorities: Theory." *Theoretical Economics*, 11, 683–710. [868]

Konishi, Hideo and M. Utku Ünver (2006), "Games of capacity manipulation in hospital-intern markets." *Social Choice and Welfare*, 27, 3–24. [899]

Martínez, Ruth, Jordi Massó, Alejandro Neme, and Jorge Oviedo (2000), "Single agents and the set of many-to-one stable matchings." *Journal of Economic Theory*, 91, 91–105. [867, 868]

McVitie, David G. and Leslie B. Wilson (1971), "The stable marriage problem." *Communications of the ACM*, 14, 486–490. [883, 890, 895]

Nambiar, Meera and Josh Bavas (2010), "Rudd plan won't fix rural doctor shortage: RDA." ABC. Available at http://www.abc.net.au/news/2010-03-16/rudd-plan-wont-fix-rural-doctor-shortage-rda/365524. [864]

Orange, Richard (2015), "Sweden urged to rethink parents' choice over schools after education decline." The Guardian. Available at http://www.theguardian.com/world/2015/may/04/sweden-school-choice-education-decline-oecd. [865]

Pathak, Parag A. and Tayfun Sönmez (2008), "Leveling the playing field: Sincere and sophisticated players in the Boston mechanism." *American Economic Review*, 98, 1636–1652. [873]

Pathak, Parag A. and Tayfun Sönmez (2013), "School admissions reform in Chicago and England: Comparing mechanisms by their vulnerability to manipulation." *American Economic Review*, 103, 80–106. [873]

Roth, Alvin E. (1982), "The economics of matching: Stability and incentives." *Mathematics of Operations Research*, 7, 617–628. [867]

Roth, Alvin E. (1984), "The evolution of the labor market for medical interns and residents: A case study in game theory." *Journal of Political Economy*, 92, 991–1016. [868]

Roth, Alvin E. (1986), "On the allocation of residents to rural hospitals: A general property of two-sided matching markets." *Econometrica*, 54, 425–427. [868]

Roth, Alvin E. (1991), "A natural experiment in the organization of entry-level labor markets: Regional markets for new physicians and surgeons in the United Kingdom." *American Economic Review*, 81, 415–440. [873]

Roth, Alvin E. (2002), "The economist as engineer: Game theory, experimentation, and computation as tools for design economics." *Econometrica*, 70, 1341–1378. [887]

Roth, Alvin E. and Elliott Peranson (1999), "The redesign of the matching market for American physicians: Some engineering aspects of economic design." *American Economic Review*, 89, 748–780. [873]

Roth, Alvin E. and Marilda A. Oliveira Sotomayor (1990), *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*, volume 18 of Econometric Society Monographs. Cambridge University Press, Cambridge. [868]

Shapley, Lloyd S. and Herbert E. Scarf (1974), "On cores and indivisibility." *Journal of Mathematical Economics*, 1, 23–37. [886]

Sönmez, Tayfun (1997), "Manipulation via capacities in two-sided matching markets." *Journal of Economic Theory*, 77, 197–204. [868]

Sönmez, Tayfun and Tobias B. Switzer (2013), "Matching with (branch-of-choice) contracts at United States Military Academy." *Econometrica*, 81, 451–488. [864]

Talbott, Chris (2007), "Shortage of doctors affects rural U.S." The Washington Post. Available at http://www.washingtonpost.com/wp-dyn/content/article/2007/07/21/AR2007072100432.html. [864]

Thomson, William (2010), "Fair allocation rules." In *Handbook of Social Choice and Welfare*, Vol. 2 (Kenneth J. Arrow, Amartya K. Sen, and Kotaro Suzumura, eds.), 393–506, Elsevier, Amsterdam. [885]

Troyan, Peter (2012), "Comparing school choice mechanisms by interim and ex-ante welfare." *Games and Economic Behavior*, 75, 936–947. [873]

Westkamp, Alexander (2013), "An analysis of the German university admissions system." *Economic Theory*, 53, 561–589. [869]

Wilson, Robert (1987), "Game-theoretic analyses of trading processes." In *Advances in Economic Theory: Fifth World Congress* (Truman F. Bewley, ed.), Econometric Society Monographs, 33–70, Cambridge University Press, Cambridge. [872]

Co-editor Faruk Gul handled this manuscript.

Manuscript received 29 May, 2015; final version accepted 8 April, 2016; available online 20 April, 2016.