First-Choice Maximal and First-Choice Stable School Choice Mechanisms

UMUT DUR, North Carolina State University, USA
TIMO MENNLE, University of Zurich, Switzerland
SVEN SEUKEN, University of Zurich, Switzerland

We investigate the class of school choice mechanisms that are *first-choice maximal (FCM)* (i.e., they match a maximal number of students to their reported first choices) and *first-choice stable (FCS)* (i.e., no students form blocking pairs with their reported first choices). FCM is a ubiquitous desideratum in school choice, and we show that FCS is the only rank-based relaxation of stability that is compatible with FCM. The class of FCM and FCS mechanisms includes variants of the well-known Boston mechanism as well as certain Asymmetric Chinese Parallel mechanisms. Regarding incentives, we show that while no mechanism in this class is strategyproof, the Pareto efficient ones are least susceptible to manipulation. Regarding student welfare, we show that the Nash equilibrium outcomes of these mechanisms correspond precisely to the set of stable matchings. By contrast, when some students are sincere, we show that more students may be matched to their true first choices in equilibrium than under any stable matching. Finally, we show how our results can be used to obtain a new characterization of the Boston mechanism (i.e., the most widely used FCM and FCS mechanism). On a technical level, this paper provides new insights about an influential class of school choice mechanisms. For practical market design, our results yield a potential rationale for the popularity of FCM and FCS mechanisms in practice.

Full paper including appendix available at SSRN: https://ssrn.com/abstract=3180824

1 INTRODUCTION

School choice programs give students an opportunity to express their preferences over which public schools they would like to attend. Ideally, one would like to match all students to their respective true first choices. However, this ideal may not be achievable because schools have limited capacities and some schools may be more popular than others. Therefore, administrators need to design school choice mechanisms that reconcile students' conflicting interests with capacity constraints. Generating high student welfare is one of the key objectives in this task [4].

One way to measure student welfare is to consider the *number of students who are matched to their first choices.* This measure is particularly tangible because maximizing it is an obvious compromise between capacity constraints and the desire to ideally match all students to their top

We would like to thank (in alphabetical order) Antonio Miralles, Axel Ockenfels, Alexander Westkamp, as well as participants of the ZEW Market Design Research Seminar, the Design and Behavior Research Seminar (Uni Cologne), the 12th MiP Workshop, MATCH-UP 2017, and the EEA/ESEM 2017 for helpful comments. Part of this research was supported by the SNSF (Swiss National Science Foundation) under grant #156836.

Authors' addresses: Umut Dur, North Carolina State University, Department of Economics, 2801 Founders Drive, Raleigh, NC, 27695, USA, udur@ncsu.edu; Timo Mennle, University of Zurich, Switzerland, mennle@ifi.uzh.ch; Sven Seuken, University of Zurich, Department of Informatics, Binzmuhlestrasse 14, Zurich, CH-8050, Switzerland, seuken@ifi.uzh.ch.

ACM EC'18, June 18--22, 2018, Ithaca, MA, USA. ACM ISBN 978-1-4503-4529-3/18/06…$15.00
https://doi.org/10.1145/3219166.3219201

choices. It comes as no surprise that the share of students who are matched to their first choices receives attention in the media with headlines such as

> *"One in six secondary pupils in England doesn't get first school choice"*[1]

and

> *"45% of New York City 8th-graders got into top high school choice [...]."*[2]

Furthermore, administrators report the share of students who are matched to their first choices as part of their public communication. For example, Denver Public Schools prominently feature this measure on the website that informs parents about school choice.[3]

In line with the popularity of this measure, many school choice mechanisms in practice attempt to maximize it, the most prominent example being the *Boston mechanism (BM)* [4]. Arguably, BM owes much of its popularity to the intuitive way in which it attempts to maximize first choices. The common focus on first choices motivates our definition of *first-choice maximality (FCM)*, which requires that a mechanism matches a maximal number of students to their reported first choices.

A second important desideratum in school choice is stability. A matching is called *stable* if it is *non-wasteful* (i.e., no student would rather be matched to a school with unfilled seats), *individually rational* (i.e., no matched student would rather be unmatched), and if it *eliminates justified envy* (i.e., if a student prefers a different school to her match, then any student matched to that other school must have higher priority[4]) [7]. Equivalently, stability can be defined as the absence of blocking pairs.[5] A student in a blocking pair may feel as though she has been treated unfairly by the mechanism. On top of that, she may even have a basis for pursuing legal actions against the school district.[6] Administrators would naturally try to avoid both, perceived unfairness and the risk of legal actions. Stability is therefore a common criterion in school choice.

Our first insight in this paper is that FCM and stability are incompatible. This raises the question whether there exists a relaxed notion of stability that can serve as a useful second best but is not in conflict with FCM. We answer this question in the affirmative: A matching is called *first-choice stable (FCS)* if no student forms a blocking pair with her reported first choice (but students may form blocking pairs with other choices). We show that among all rank-based relaxations of stability, FCS is the only one that is compatible with FCM.[7]

This insight gives rise to a natural class of school choice mechanisms: We start with the ubiquitous desideratum to maximize the number of students who are matched to their first choices (i.e., FCM). Given this, it would be unreasonable to match these first choices in a way that violates priorities; thus, we also require FCS. On the other hand, all more demanding notions of stability are in conflict with FCM. This motivates our analysis of the class of mechanisms that satisfy both FCM and FCS. For the sake of brevity, we refer to these mechanisms as *first-choice (FC) mechanisms*.

---

[1] Sally Weale (June 14, 2016). *The Guardian*. Retrieved October 10, 2016: www.theguardian.com

[2] Sophia Rosenbaum and Erica Pearson (March 11, 2014). *New York Daily News*. Retrieved October 10, 2016: nydailynews.com

[3] Denver Public Schools Website. Retrieved October 10, 2016: schoolchoice.dpsk12.org/schoolchoice/

[4] Priorities are common in school choice. For example, a school may grant priority to students whose siblings attend the same school, who live nearby, or based on grades [3].

[5] A student-school-pair (i, s) is *blocking* if i would prefer s to her current match and s has unfilled seats, s is i's outside option, or some student who is matched to s has lower priority at s than i.

[6] A Wisconsin student's lawsuit succeeded on the grounds of *justified envy* [4]. It appears natural that students with justified envy are more likely to sue for violations of priorities at first choices than at other ranks. The admission to medical degree programs in Germany is organized via a centralized mechanism [33]. Rejected applicants sometimes obtain a seat by suing universities for leaving seats unfilled. These lawsuits succeed on the grounds of *wastefulness*.

[7] A *rank-based relaxation of stability* requires that there are no blocking pairs for which the respective school has a certain rank in the respective student's preference order.

The class of FC mechanisms includes all commonly employed variants of the Boston mechanism (e.g., with varying tie-breakers, varying limits on the length of preference lists, and where filled schools are or are not skipped in the application process). Such mechanisms are used in many school districts, including Minneapolis, Seattle, Lee County [24], San Diego, Amsterdam (until 2014) [14], Wake County (until 2015) [16], and in Nordrhein Westfalen and Freiburg (Germany) [8]. Asymmetric Chinese Parallel mechanisms with $e_0 = 1$ also belong to this class and are used for college admissions in the Chinese Beijing, Gansu, and Shandong provinces [11]. A plausible motivation for using FC mechanisms is the desire to achieve FCM and FCS. As market designers, we are of course aware of the fact that strategic misreporting by students may impede this objective. Nonetheless, administrators may find FC mechanisms appealing, for example, if they believe that students will report their preferences truthfully despite contrary incentives.[8] Moreover, even if students strategize, administrators may prefer FC mechanisms for cosmetic reasons (e.g., if they are driven by other considerations such as favorable media coverage).

For market designers, the question arises whether and to what extent *FC mechanisms actually achieve the intended desiderata* to match a maximal number of students to their *true* first choices and to avoid blocking pairs of students with their respective *true* first choices. This research question is the focus of our present paper. To answer it, we proceed in two steps:

Step 1. We identify the incentives for students under FC mechanisms (by comparing these mechanisms by their vulnerability to manipulation).
Step 2. We investigate how strategic reporting by students affects outcomes (by studying the Nash equilibria of the induced preference revelation games).

Regarding **incentives** (Step 1), our first insight is that FCM alone is already incompatible with strategyproofness, even without the additional restriction of FCS. Despite this incompatibility, some FC mechanisms may invite more strategic misreporting than others. Towards understanding these differences, we employ the concept of comparing mechanisms by their *vulnerability to manipulation*, introduced by Pathak and Sönmez [30].[9] We show that all Pareto efficient FC mechanisms are manipulable at exactly the same problems and are therefore equivalent in this sense. Moreover, we show that for any Pareto *in*efficient FC mechanism, there exists a Pareto efficient FC mechanism that Pareto dominates the original mechanism but is also at most as manipulable.

Our results have two significant consequences: First, the two most widely studied FC mechanisms are the classic *Boston mechanism (BM)* [4] and the *adaptive Boston mechanism (ABM)* [5, 15, 22, 26, 27]. One would intuitively suspect ABM to have better incentive properties than BM. However, since both mechanisms are Pareto efficient, they are manipulable at the same problems. Thus, surprisingly, the intuitive difference in their incentive properties cannot be formalized via the comparison by vulnerability to manipulation.[10] The second consequence of our results pertains to FC mechanisms used in practice that are not Pareto efficient, such as Asymmetric Chinese Parallel mechanisms [11]. Our results imply that these mechanisms are more manipulable and less efficient than necessary, even if administrators are restricted to only using FC mechanisms. Thus, the motivation for using Pareto *in*efficient FC mechanisms must rely on other considerations

[8] For example, the Teach for America (TfA) program uses a non-strategyproof mechanism to match teachers-to-be to teaching positions. However, TfA administrators feel that the use of this mechanism is justified because participants would find it hard to acquire the skills and information necessary to successfully manipulate the mechanism [19].

[9] A mechanism φ is *at most as manipulable as* another mechanism ψ if the set of problems where some student can benefit from misreporting under φ is a subset of the respective set under ψ.

[10] ABM differs from BM in that students automatically skip exhausted schools in the application process under ABM, while under BM they apply to exhausted schools (thereby wasting some rounds). More nuanced assumptions are required to establish a formal understanding of the different incentive properties of BM and ABM, e.g., when school may have zero capacity or find some students unacceptable [15] or when priorities are random [26].

besides incentives. Otherwise, unambiguous improvements to these mechanisms would be possible, even within the class of FC mechanisms.

Regarding the **impact of strategic behavior on outcomes** (Step 2), we show that the set of Nash equilibrium outcomes of any FC mechanism corresponds precisely to the set of matchings that are stable with respect to the true preferences. This means that the equilibrium outcomes of any FC mechanism are first-choice stable with respect to the true preferences, but the may not match a maximal number of students to their true first choices.[11] Our result generalizes the main result of Ergin and Sönmez [18], who showed this correspondence for BM. For market designers, the most important consequence of our result is that in markets where all students are strategic (and when the primary concerns are concerns stability and matching true first choices), there is no reason for using an FC mechanism instead of the *student-proposing deferred-acceptance (DA) mechanism* [4]: DA produces the *student-optimal* stable matching, which is the unique stable matching that Pareto dominates all other stable matchings, and DA is strategyproof. In contrast, FC mechanisms merely produce *some* stable matchings, and they do this only subject to a weaker solution concept.

Our observations so far pertain to the case when *all* students strategize. In practice, however, students may exhibit varying levels of strategic sophistication. For example, it may be cognitively challenging to determine a beneficial misreport, or acquiring the necessary information may be costly. Experimental results suggest that under BM, a significant share of the participants report their preferences truthfully despite incentives to misreport [12, 13], and a lack of information further increases this share [28].[12] We therefore follow Pathak and Sönmez [29], who considered *mixed* problems with two types of students: *Sincere* students report their preferences truthfully, independent of incentives, while *sophisticated* students recognize the strategic aspect of the matching process. For these mixed problems, we identify the Nash equilibrium outcomes of FC mechanisms. Specifically, we show that, from the perspective of the sophisticated students, the set of Nash equilibrium outcomes corresponds to the set of matchings that are stable with respect to the true preferences and certain *augmented priorities*.[13] Our result partially generalizes the result of Pathak and Sönmez [29], who showed this correspondence for BM.[14] Our result also implies the existence of equilibrium outcomes that are uniformly preferred by all sophisticated students.

With these results at our disposal, we can conclusively answer our main research question: To what extent do **FC mechanisms actually achieve FCM and FCS**? Towards FCS, we show that all equilibrium outcomes of any FC mechanism are FCS with respect to the true preferences. Thus, FC mechanisms actually implement FCS, independent of which students strategize and independent of which equilibrium is chosen. Towards FCM, we isolate the effect of any individual student's decision (say i') whether to strategize or to report truthfully by comparing the respective sophisticated-student-optimal equilibrium outcomes. We show that under any FC mechanism, i' prefers being sophisticated to being sincere, all other sophisticated students prefer i' being sincere, and for sincere students the difference is ambiguous with one notable exception: Any student who is matched to her true first choice when i' is sophisticated is still matched to her true first choice when i' is sincere. Thus, the number of true first choices matched in strategic-student-optimal

[11] This follows because stability implies FCS but is incompatible with FCM.

[12] In laboratory experiments conducted by Chen and Sönmez [13], this share was 13%, and it increased to 28% when priorities were random. Chen and Kesten [12] observed shares between 23% and 46% in similar experiments, and Pais and Pinter [28] found that withholding information significantly increased these shares from 47% to 87%.

[13] *Augmented priorities* arise as an adjustment of the original priorities. They account for the positions in which students rank schools and whether or not students are sophisticated (see Section 6.1).

[14] The correspondence proven by Pathak and Sönmez [29] holds for all students. Our restriction to the perspective of sophisticated students is necessary because FC mechanisms are identified by how they match students to their reported first choices but are unrestricted in how they match students to other choices (see Lemma 4.3). In equilibrium, this freedom only affects sincere students.

equilibrium outcomes is lowest if all students strategize, increases in mixed problems with more sincere students, and is maximal when all students are sincere. In this sense, FC mechanisms have the potential to match more true first choices in equilibrium than strategyproof alternatives like DA. This prediction is consistent with experimental findings that BM and ACPM both match significantly more students to their true first choices than DA (Result 4 of Chen and Kesten [12]).

Finally, we provide a new characterization of BM, the FC mechanism most widely used in practice. For this purpose, we consider an extended model where schools may have zero capacity initially. Since FCM and FCS can become void constraints in this extended model, we extend the definition of FCM to best-choice maximality by requiring that a mechanism maximizes the number of students assigned to their k^{th} choice such that in any matching no students can be assigned to a better choice than k. We define best-choice stability analogously. We then show that best-choice maximality and best-choice stability, in combination with consistency [24], characterize BM in the extended model.

For market designers, our main results imply that the use of FC mechanisms in practice may be justified if FCM and FCS are primary desiderata and if a sufficiently large share of the students can be expected to report their preferences truthfully. Our characterization result of BM shows that, within the class of FC mechanisms, BM stands out a little more than the others, which may be a potential rationale for its popularity in practice.

2 PRELIMINARIES

2.1 Formal Model

A *school choice problem* is a tuple $(I, S, q, P, \succ)$ with a finite set of *students* $I = \{i_1, \ldots, i_n\}$ and a finite set of *schools* $S = \{s_1, \ldots, s_m\}$. $q = (q_s)_{s \in S}$ is the vector of *school capacities* (i.e., q_s is the number of seats available at school s), $P = (P_i)_{i \in I}$ is the *preference profile* in which each P_i is the strict *preference order* of student i over the schools in S and the *outside option*, denoted by $\emptyset$. $\succ = (\succ_s)_{s \in S}$ is the *priority profile* in which each $\succ_s$ is the *priority order* of school s over students in I. We assume that priorities are strict, known, and fixed. In particular, this means that tie-breaking is not necessary, random or otherwise, and that schools do not strategize. $s\ P_i\ s'$ means that student i strictly prefers school s to school s', and $i \succ_s i'$ means that student i has priority over student i' at school s. We assume that there is at least one seat at each school (i.e., $q_s \geq 1$ for all $s \in S$) and that all students can be matched to their outside option (i.e., $q_\emptyset = n$). For a preference order P_i, the corresponding *weak preference order* is denoted by R_i (i.e., $s\ R_i\ s'$ if either $s\ P_i\ s'$ or $s = s'$). Unless stated otherwise, we fix I, S, q, and we use $(P, \succ)$ to denote a specific *problem* throughout the paper.

A *matching* is a function $\mu : I \to S \cup \{\emptyset\}$. For a given matching μ, $\mu(i)$ is the school to which student i is matched, and $\mu^{-1}(s)$ is the set of students who are matched to schools s. We focus on *feasible* matchings (i.e., $\left|\mu^{-1}(s)\right| \leq q_s$ for every $s \in S \cup \{\emptyset\}$), and we simplify notation by writing μ_i and μ_s for $\mu(i)$ and $\mu^{-1}(s)$, respectively.

For a problem $(P, \succ)$ and matchings μ and ν, we say that μ *weakly Pareto dominates* ν if $\mu_i\ R_i\ \nu_i$ for all students $i \in I$, and μ *Pareto dominates* ν if μ weakly Pareto dominates ν and $\mu_{i'}\ P_{i'}\ \nu_{i'}$ for at least one student $i' \in I$. The matching μ is *Pareto efficient* if it is not Pareto dominated by any other matching, μ is *individually rational* if $\mu_i\ R_i\ \emptyset$ for all students $i \in I$, μ is *wasteful* if there exists some student $i \in I$ and some school $s \in S$ such that $|\mu_s| < q_s$ but $s\ P_i\ \mu_i$, and μ is *non-wasteful* if it is not wasteful. A student $i \in I$ has *justified envy (under μ)* if there exists another student $i' \in I$ and a school $s \in S$ such that $s\ P_i\ \mu_i$, $\mu_{i'} = s$, and $i \succ_s i'$, and μ is *fair* if no student has justified envy. Finally, μ is *stable* if it is individually rational, non-wasteful, and fair. Observe that for any unstable matching μ, there exists at least one pair $(i, s) \in I \times (S \cup \{\emptyset\})$ such that $s\ P_i\ \mu_i$ and $|\mu_s| < q_s$ (if μ is wasteful), or $s = \emptyset$ (if μ is not individually rational), or there exists a student $i' \in I$ with $\mu_{i'} = s$ and $i \succ_s i'$ (if i has justified envy). Any such student-school pair is called a *blocking pair*. Obviously,

Mechanism	SP	PE	ST	FCM	FCS	Examples of use in practice
BM		✓		✓	✓	Minneapolis, Seattle, Lee County, San Diego
ABM		✓		✓	✓	Amsterdam (until 2014), Nordrhein Westfalen
DA	✓		✓		✓	New York, Boston, Mexico City
ACPM					✓	Various Chinese provinces
ACPM, $e_0 = 1$				✓	✓	Bejing, Gansu, and Shangdong provinces
TTC	✓	✓				New Orleans Recovery School District (2012)

Table 1. Mechanisms, properties, examples of use; strategyproof (*SP*), Pareto efficient (*PE*), stable (*ST*).[15]

the matching μ is stable if and only if there exist no blocking pairs. Throughout the paper, we employ this equivalent definition of stability to simplify definitions and proofs.

A *mechanism* φ is a mapping that receives a problem $(P, \succ)$ as input and selects a matching, denoted by $\varphi(P, \succ)$. We denote by $\varphi_i(P, \succ)$ the school to which student i is matched under $\varphi(P, \succ)$. A mechanism φ is called *Pareto efficient*/*individually rational*/*non-wasteful*/*fair*/*stable* if it selects matchings with the respective property for all problems. The mechanism φ *Pareto dominates* another mechanism ψ if the matching $\varphi(P, \succ)$ weakly Pareto dominates the matching $\psi(P, \succ)$ for all problems $(P, \succ)$ and this dominance is not weak for at least one problem. Observe that these properties are formulated in terms of how the mechanisms handle reported preferences. However, students may lie about their preferences so that the input to the mechanism may differ from the true problem. Regarding incentives, a mechanism φ is called *strategyproof* if, for all problems $(P, \succ)$, all students $i \in I$, and all preference orders P_i', we have $\varphi_i(P, \succ) \; R_i \; \varphi_i((P_i', P_{-i}), \succ)$ where $P_{-i} = (P_j)_{j \neq i}$ are the preferences of all students except i.

Finally, we introduce some auxiliary notation: Given a problem $(P, \succ)$, let $\text{choice}_{P_i}(k)$ be the k^{th} *choice* of student i according to P_i. For a matching μ, let $I(\mu, k, P)$ be the set of students who are matched to their k^{th} choice under the matching μ; formally, $I(\mu, k, P) = \left\{ i \in I : \text{choice}_{P_i}(k) = \mu_i \right\}$. Thus, $I(\varphi(P, \succ), k, P)$ is the set of students who are matched to their k^{th} choice (according to P) when the mechanism φ is applied to the problem $(P, \succ)$.

2.2 School Choice Mechanisms

In this section, we describe common school choice mechanisms. Table 1 provides an overview of their properties as well as examples of their use in practice.

The *Boston mechanism (BM)* [4] works in rounds. In the first round, all students apply to their respective first choices, and each school permanently accepts applications from students in order of priority until all applications are accepted or until all seats are filled. Students whose applications are not accepted enter the second round where they apply to their respective second choices. Again, schools accept applications into unfilled seats by priority and reject all remaining applications once all seats are filled. This process continues (i.e., students who were rejected in round $k - 1$ apply to their k^{th} choices in round k) until no school receives new applications.

The *adaptive Boston mechanism (ABM)* [5, 15, 22, 26, 27] is similar to BM, except that students who are rejected in round $k - 1$ apply to their respective *most-preferred school that still has at least one unfilled seat* in round k. Students thus automatically skip schools in the application process when applications to these schools are bound to be rejected, independent of priority.

[15]Sources: [2, 3, 8, 10, 11, 14, 24], and Andrew Vanacore (April 16, 2012). *The Times-Picayune*. Retrieved March 22, 2017: www.nola.com

Under the *Student-Proposing Deferred Acceptance (DA) mechanism* [4, 20], students also apply to schools in rounds, and priorities determine which applications are accepted. However, acceptances are *tentative* rather than permanent. This means that students can be rejected by a school where they were tentatively accepted in a previous round if other students apply in later rounds who have higher priority. Newly rejected students apply to their most preferred school that has not yet rejected them. The application process ends when no school receives any new applications.

Asymmetric Chinese Parallel mechanisms (ACPM) [11] are mechanisms that combine elements of BM and DA. They are parametrized by a vector of integers $(e_0, e_1, \ldots)$ with $e_k \geq 1$. Initially, all matches are tentative as under DA. However, unlike under DA, students who are rejected by their ${e_0}^{\text{th}}$ choices pause in the application process. When all students are either matched tentatively or have been rejected by their ${e_0}^{\text{th}}$ choices, all tentative matches are finalized. In the next phase, the rejected students continue to apply but pause again when they have been rejected by their $(e_0 + e_1)^{\text{th}}$ choices. The tentative matches are again finalized if all students are either tentatively accepted at some school or pausing. This process continues until all students are matched or have been rejected by all schools on their preference list. Observe that ACPM specifies a class of mechanisms. This class subsumes BM (for $e_k = 1$ for all k) and DA (for $e_0 \geq |S|$).

The *Top Trade Cycles (TTC) mechanism* [4] works by forming a directed graph: Each student points to her most-preferred school with unfilled seats, each school points to the student who has highest priority at that school, and the outside option points to all students who are pointing to it. In each step, a cycle of this graph is selected and implemented (i.e., each student in the cycle is permanently matched to the school to which she is pointing, and the respective seats and students are removed from the mechanism). Students and schools then adjust where they are pointing and the process repeats. The process ends when all students have been removed.

3 SETTING THE STAGE: FIRST-CHOICE MAXIMALITY, FIRST-CHOICE STABILITY

As we have argued in the introduction, the number of students who are matched to their (reported) first choices receives a lot of attention. Following this observation, we define first-choice maximality.

Definition 3.1. Given a problem $(P, \succ)$, a matching μ is *first-choice maximal* if there exists no other matching ν such that $|I(\mu, 1, P)| < |I(\nu, 1, P)|$. A mechanism φ is *first-choice maximal (FCM)* if, for all problems $(P, \succ)$, the matching $\varphi(P, \succ)$ is first-choice maximal.[16]

Next, we define rank-based relaxations of stability.

Definition 3.2. Given a problem $(P, \succ)$ and an integer $k \in \{1, \ldots, m\}$, a matching μ is *k^{th}-choice stable* if there exists no blocking pair $(i, s) \in I \times (S \cup \{\emptyset\})$ where s is the k^{th} choice of i, and μ is *first-choice stable* if this holds for $k = 1$. A mechanism is *k^{th}-choice stable (first-choice stable (FCS))* if, for all problems $(P, \succ)$, the matching $\varphi(P, \succ)$ is k^{th}-choice stable (first-choice stable).

We observe that FCM and FCS are compatible: It is easy to see that BM satisfies both properties. On the other hand, FCM by itself is already a severe restriction: As the next example shows, it is incompatible with strategyproofness, with stability, or even with k^{th}-choice stability for $k \geq 2$.

Example 3.3. There are three students $I = \{1, 2, 3\}$, two schools $S = \{a, b\}$ with a single seat each. The preferences and priorities are

$$P_i \text{ for } i \in I: \qquad a \; P_i \; b \; P_i \; \emptyset,$$
$$\succ_s \text{ for } s \in S: \qquad 1 \succ_s 2 \succ_s 3.$$

[16] FCM is a strictly weaker requirement than the axiom that a mechanism *favors higher ranks* [24]: Specifically, a mechanism is FCM if and only if it favors the *first* rank.

Let φ be an FCM mechanism. By feasibility, φ must leave at least one student unmatched. Without loss of generality, suppose that this is student 3 (i.e., $\varphi_3(P, \succ) = \emptyset$). If student 3 reports $b\ P_3'\ a\ P_3'\ \emptyset$ instead of reporting P_3 truthfully, then φ must match student 3 to school b. Since $\varphi_3((P_3', P_{-3}), \succ) = b\ P_3\ \emptyset = \varphi_3(P, \succ)$, P_3' is a beneficial misreport for student 3. Thus, φ cannot be strategyproof.

Next, observe that at the problem $((P_3', P_{-3}), \succ)$, there exists a unique stable matching μ where $\mu_1 = a$, $\mu_2 = b$, and $\mu_3 = \emptyset$. However, any FCM mechanism φ must match student 3 to school b, so φ cannot be stable. A straightforward extension illustrates that φ also violates k^{th}-choice stability for any $k \geq 2$: Consider a problem with k schools $S = \{s_1, \ldots, s_k\}$ with a single seat each, and $k + 1$ students $I = \{1, 1', 2, \ldots, k\}$. The preferences and priorities are

$$\begin{aligned} P_i \text{ for } i \in \{1, 1'\} &: \quad s_1\ P_i\ s_2\ P_i\ \ldots\ P_i\ s_k\ P_i\ \emptyset, \\ P_i \text{ for } i \in \{2, \ldots, k\} &: \quad s_i\ P_i\ \ldots, \\ \succ_s \text{ for } s \in S &: \quad 1 \succ_s 1' \succ_s 2 \succ_s \ldots \succ_s k. \end{aligned}$$

FCM implies that φ matches each student $i \in \{2, \ldots, k\}$ to s_i. Thus, either student 1 or student 1′ is unmatched. Then that student forms a blocking pair with her k^{th} choice.

It follows from Example 3.3 that among all rank-based relaxations of stability, FCS is the only one compatible with FCM. For the remainder of this paper, we therefore focus on the class of mechanisms that are both FCM and FCS.

Definition 3.4. A mechanism is a *first-choice (FC) mechanism* if it is FCM and FCS.

For example, BM and ABM are both FCM and FCS and known to be Pareto efficient, but they violate strategyproofness and stability. ACPMs with $e_0 = 1$ are also FCM and FCS. Their first phase is the same as the first round of BM, and the matches made in this phase are not undone in subsequent phases. These mechanisms thus satisfy FCM and FCS (by Lemma 4.3 in Section 4). Moreover, as Chen and Kesten [11] pointed out, ACPMs are not Pareto efficient unless $e_k = 1$ for all k, in which case they are equivalent to BM, and they are not stable unless $e_0 > m$, in which case they are equivalent to DA. DA is stable and therefore FCS. However, FCM is incompatible with stability, so DA cannot be FCM. It is straightforward to see that TTC is neither FCM nor FCS, and it is known to be strategyproof and Pareto efficient but unstable.[17]

4 INCENTIVES UNDER FC MECHANISMS

As Example 3.3 illustrates, no FC mechanism is strategyproof. However, some FC mechanisms may be *more manipulable* than others. Towards understanding these differences, we employ the concept of *comparing mechanisms by their vulnerability to manipulation* introduced by Pathak and Sönmez [30], which we restate formally before presenting our results.

Definition 4.1. A mechanism φ is *manipulable at* $(P, \succ)$ if there exist a student $i \in I$ and a preference order P_i' such that $\varphi_i((P_i', P_{-i}), \succ)\ P_i\ \varphi_i(P, \succ)$. For two mechanisms φ and ψ, we say that φ is *at least as manipulable as* ψ if, for all problems $(P, \succ)$, manipulability of ψ at $(P, \succ)$ implies manipulability of φ at $(P, \succ)$. φ is *more manipulable than* ψ if in addition there exists a problem where φ is manipulable but ψ is not.

In words, the comparison by vulnerability to manipulation classifies mechanisms by the sets of problems at which they are manipulable. If φ is at least as manipulable as ψ, then the set of problems at which φ is manipulable is a superset of the set of problems where ψ is manipulable, and φ is more manipulable if it is a strict superset.

[17] TTC violates FCM at the problem $((P_3', P_{-3}), \succ)$ from Example 3.3, and it violates FCS at the same problem if the priorities of school a are changed to $3 \succ_a 2 \succ_a 1$.

Our next results, Theorem 4.2 and Proposition 4.5, reveal how FC mechanisms compare in terms of their vulnerability to manipulation. First, we show that all Pareto efficient FC mechanisms are manipulable at *exactly the same problems*. Second, we show that any Pareto *in*efficient FC mechanism is manipulable at a (possibly strict) *superset* of these problems. In this sense, the Pareto efficient FC mechanisms form a minimally manipulable subset within the class of FC mechanisms.

THEOREM 4.2. *Let φ and ψ be two Pareto efficient FC mechanisms. Then φ is at least as manipulable as ψ and vice versa.*

PROOF. We state two lemmas, which we also use in other proofs. Recall that $\text{choice}_{P_i}(k)$ denotes the k^{th} choice school according to P_i and that $I(\mu, k, P)$ denotes the set of students who are matched to their k^{th} choice according to P under μ.

LEMMA 4.3. *φ is an FC mechanism if and only if $I(\varphi(P, \succ), 1, P) = I(BM(P, \succ), 1, P)$ for all $(P, \succ)$.*

LEMMA 4.4. *Given an FC mechanism φ, a problem $(P, \succ)$, and a subset of students $A \subseteq I$, let (P'_A, P_{-A}) be a preference profile such that $\text{choice}_{P'_i}(1) = \varphi_i(P, \succ)$ for all $i \in A$. Then, $\varphi_i((P'_A, P_{-A}), \succ) = \varphi_i(P, \succ)$ for all $i \in A$.*

The proofs of these Lemmas are given in Appendices A and B of the full version of this paper. Lemma 4.3 characterizes the set of FC mechanisms as those mechanisms that match exactly the same students to their reported first choices as BM; but otherwise they are free in how they match the remaining students. Lemma 4.4 shows that when some students change their reported preferences by ranking first the school to which they are matched under an FC mechanism, then this mechanism continues to match these students to the same schools.[18]

Next, let φ be manipulable at $(P, \succ)$. Then $\varphi_i((P'_i, P_{-i}), \succ) \; P_i \; \varphi_i(P, \succ)$ for some student $i \in I$ and some preference order P'_i. Without loss of generality, let $s = \varphi_i((P'_i, P_{-i}), \succ)$ be the most preferred school according to P_i at which i can obtain a seat by misreporting (and possibly $s = \emptyset$). By Lemma 4.4, we can choose P'_i such that s is ranked first. Observe that s cannot be the first choice under P_i (otherwise, P'_i would not be a strictly beneficial deviation for i). There are two cases:

Case 1: $\varphi(P, \succ) = \psi(P, \succ)$. Then $I(\varphi((P'_i, P_{-i}), \succ), 1, P) = I(\psi((P'_i, P_{-i}), \succ), 1, P)$ by Lemma 4.3. Since $i \in I(\varphi((P'_i, P_{-i}), \succ), 1, P)$ and s is i's first choice under P'_i, we have $i \in I(\psi((P'_i, P_{-i}), \succ), 1, P)$ and $\psi_i((P'_i, P_{-i}), \succ) = s$. Hence, $\psi_i((P'_i, P_{-i}), \succ) \; P_i \; \psi_i(P, \succ)$, that is, i can manipulate ψ at $(P, \succ)$.

Case 2: $\varphi(P, \succ) \neq \psi(P, \succ)$. Since both $\varphi(P, \succ)$ and $\psi(P, \succ)$ are FCS and Pareto efficient, there exists a student $i' \in I$ such that $\varphi_{i'}(P, \succ) \neq \psi_{i'}(P, \succ)$ and $\varphi_{i'}(P, \succ) \; P_{i'} \; \psi_{i'}(P, \succ)$. Let $s' = \varphi_{i'}(P, \succ)$ and let $P'_{i'}$ be a preference order in which s' is ranked first. Then $\varphi_{i'}((P'_{i'}, P_{-i'}), \succ) = s'$ by Lemma 4.3, and $I(\varphi((P'_{i'}, P_{-i'}), \succ), 1, (P'_{i'}, P_{-i'})) = I(\psi((P'_{i'}, P_{-i'}), \succ), 1, (P'_{i'}, P_{-i'}))$ by Lemma 4.4. Hence, $\psi_{i'}((P'_{i'}, P_{-i'}), \succ) = s' \; P_{i'} \; \psi_{i'}(P, \succ)$, that is, i' can manipulate ψ at $(P, \succ)$.

In conclusion, manipulability of φ at $(P, \succ)$ implies manipulability of ψ at $(P, \succ)$. Symmetrically, it follows that if ψ is manipulable at some problem, then so is φ. □

While Theorem 4.2 pertains to Pareto efficient FC mechanisms, the next proposition closes the remaining gap for Pareto *in*efficient FC mechanisms.

[18]This corresponds to a relaxed notion of Maskin monotonicity for FC mechanisms: The preference profile (P'_A, P_{-A}) is a monotonic transformation of P at $\varphi(P, \succ)$, and under any FC mechanism the matching of the students in A may not change. Observe that this property is independent of the *rank respecting invariance* property, a different relaxation of Maskin monotonicity that Kojima and Ünver [24] used for their axiomatic characterization of BM.

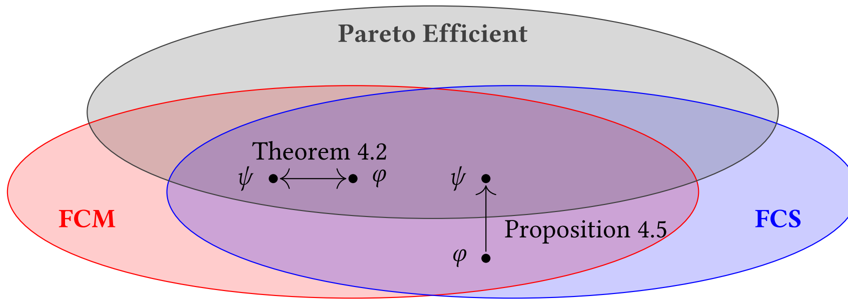

Fig. 1. Illustration of Theorem 4.2 and Proposition 4.5.

PROPOSITION 4.5. *Let φ be an FC mechanism that violates Pareto efficiency. Then there exists a mechanism ψ with the following properties:*

1. *ψ is a Pareto efficient FC mechanism,*
2. *ψ Pareto dominates φ,*
3. *φ is at least as manipulable as ψ.*

The proof is given in Appendix C of the full version of this paper. The Venn diagram in Figure 1 illustrates the results of Theorem 4.2 and Proposition 4.5. All Pareto efficient FC mechanisms (i.e., mechanisms in the intersection of all three areas) are equivalent when comparing them by their vulnerability to manipulation. The horizontal arrow on the left symbolizes this equivalence. Moreover, any Pareto *in*efficient FC mechanism (i.e., a mechanism in the intersection of the blue and red areas at the bottom) is at least as manipulable as some (and therefore any) Pareto efficient FC mechanism. The vertical arrow on the right symbolizes this relationship.

Theorem 4.2 and Proposition 4.5 have two implications for market design: First, recall that under BM, students apply to their k^{th} choices in the k^{th} round, even if these schools have no more unfilled seats. Students can therefore strategize by omitting full schools in their ranking. In contrast, under ABM, students automatically skip such schools and apply to their most preferred school with one or more unfilled seats. While students may still strategize in other ways, the above-mentioned manipulation becomes unnecessary. Thus, intuitively, we would expect ABM to have better incentive properties than BM. Surprisingly, this difference does not surface because both mechanism are equivalent in this sense by Theorem 4.2.[19]

Second, ACPMs with $e_0 = 1$ are Pareto *in*efficient (unless $e_k = 1$ for all k, in which case they are equivalent to BM), but they are used in practice (e.g., for college admission in the Chinese Beijing, Gansu, and Shandong provinces). Proposition 4.5 shows that the choice of such mechanisms cannot be justified by the desiderata FCM, FCS, and "good" incentives alone, because administrators could choose a different FC mechanism that simultaneously yields unambiguous improvements in terms of student welfare and robustness to manipulation.

This concludes Step 1, the discussion of incentives under FC mechanisms.

[19]The indistinguishability of BM and ABM is even more severe: It cannot be recovered via the as-*strongly*-manipulable-as relation, and for random priorities, even the weak distinction by vulnerability to manipulation is inconclusive (see Appendices G and H of the full version of this paper. More nuanced approaches are needed to obtain meaningful distinctions (see, e.g., [15, 26]).

5 EQUILIBRIUM UNDER FC MECHANISMS

Recall that no FC mechanism is strategyproof. We must therefore be concerned about the impact of strategic reporting by students. Following prior work, we identify the equilibrium outcomes of the preference revelation game induced by the mechanisms and use *Nash equilibrium under complete information* as the main solution concept [18, 21, 23, 29].

Definition 5.1. Given a mechanism φ and a problem $(P, \succ)$, a preference profile P^* is a *Nash equilibrium* (*of* φ *at* $(P, \succ)$) if, for all students $i \in I$ and all preference orders $P'_i \neq P^*_i$, we have that $\varphi_i(P^*, \succ)\ R_i\ \varphi_i((P'_i, P^*_{-i}), \succ)$.

In words, no student can be matched to a school that she strictly prefers by unilaterally deviating from the equilibrium profile.[20] For the special case of BM, Ergin and Sönmez [18] showed that the Nash equilibrium outcomes of BM correspond precisely to the matchings that are stable with respect to the true preferences and priorities.

Fact 1 (Theorem 1 of Ergin and Sönmez [18])**.** *Given a problem* $(P, \succ)$*, a matching* μ *is stable if and only if there exists a Nash equilibrium* P^* *of BM at* $(P, \succ)$ *with* $\mu = BM(P^*, \succ)$.

Our next theorem shows that the same result holds for all FC mechanisms.

THEOREM 5.2. *Given a problem* $(P, \succ)$ *and an FC mechanism* φ*, a matching* μ *is stable at* $(P, \succ)$ *if and only if there exists a Nash equilibrium* P^* *of* φ *at* $(P, \succ)$ *with* $\mu = \varphi(P^*, \succ)$.

PROOF. *Necessity.* Let P^* be a Nash equilibrium. Assume towards contradiction that the matching $\varphi(P^*, \succ)$ is not stable wrt. $(P, \succ)$. Then there exist $i \in I$ and $s \in S \cup \{\emptyset\}$ such that $s\ P_i\ \varphi_i(P^*, \succ)$ and either $|\varphi_s(P^*, \succ)| < q_s$, or there exists $i' \in I$ such that $\varphi_{i'}(P^*, \succ) = s$ and $i \succ_s i'$. Let P'_i be a preference order with $\text{choice}_{P'_i}(1) = s$. When BM is applied to the preference profile (P'_i, P^*_{-i}), i is matched to s in the first round. By Lemma 4.3 we get $\varphi_i((P'_i, P^*_{-i}), \succ) = \text{BM}_i((P'_i, P^*_{-i}), \succ) = s$. Therefore, P^*_i is not a best response to P^*_{-i} for i, a contradiction.

Sufficiency. Let μ be stable wrt. $(P, \succ)$. Consider a preference profile P^* where each student i ranks μ_i first. Since μ is feasible, φ produces the matching μ in the first step when applied to $(P^*, \succ)$ by Lemma 4.3. Moreover, P^* is a Nash equilibrium of φ: Assume towards contradiction that some student i can improve her match by deviating from P^* (i.e., get matched to a school s). By Lemma 4.4, she can do so by ranking s first. Then i either displaces another student with lower priority at s, or s has unfilled seats under μ. In both cases, the pair (i, s) blocks μ, a contradiction. □

Remark 1. Theorem 5.2 shows that stability (with respect to the true preferences and priorities) is the characterizing feature of the Nash equilibrium outcomes of FC mechanisms. This generalizes Fact 1 from BM to all FC mechanisms. Ergin and Sönmez [18] also showed that all monotonic rank-priority mechanisms have this property. Relatedly, within the class of rank-priority mechanisms, Jaramillo et al. [23] characterized those that Nash implement the set of stable matchings. Our Theorem 5.2 is independent of both of these results because the set of FC mechanisms is neither a subset nor a superset of the rank-priority mechanisms (see Appendix D of the full version of this paper). An interesting subject for future research would be a characterization of all direct-revelation mechanisms that Nash implement the set of stable matchings that unifies these results.

[20] The *preference revelation game induced by* φ *at* $(P, \succ)$ is a simultaneous move game (N, O, τ), where the students are the agents (i.e., $N = I$), the outcomes are the matchings that are possible under φ (i.e., $O = \{\varphi(\widehat{P}, \succ) \mid \widehat{P} \text{ preference profile}\}$), and for each agent i, the weak preference order τ_i over outcomes is induced by the respective student's weak preference order over schools (i.e., for outcomes $x, y \in O$, $x\ \tau_i\ y$ if and only if $x_i\ R_i\ y_i$).

Observe that we simplify the formal exposition of Nash equilibrium in two ways: First, we use the term *Nash equilibrium of* φ *at* $(P, \succ)$ to mean a Nash equilibrium of the induced preference revelation game. Second, we consider *preference profiles* instead of strategy profiles. This is without loss of generality because we consider only direct revelation mechanisms.

DA is known to implement the *student-optimal* stable matchings in weakly-dominant strategies.[21] In contrast, by Theorem 5.2, FC mechanisms implement *all* stable matchings in Nash equilibrium, not just the student-optimal ones. Thus, DA produces (weakly) Pareto dominant matchings, and it does so subject to a more robust solution concept. This provides a partial answer to our main research question whether FC mechanisms actually achieve the desiderata FCM and FCS with respect to the true preferences: They achieve FCS because they lead to stable matchings (in equilibrium when all students strategize), but they fail to achieve FCM (which is incompatible with stability).

6 EQUILIBRIUM UNDER FC MECHANISMS WHEN SOME STUDENTS ARE SINCERE

In practice, students exhibit varying levels of strategic sophistication. Some students may report their preferences truthfully, e.g., because they lack the information that is necessary to determine beneficial misreports. Following Pathak and Sönmez [29], we consider *mixed problems* with two groups of students: *Sincere students* simply report their preferences truthfully, independent of incentives, while *sophisticated students* recognize the strategic nature of the preference revelation game and play best responses. In this section, we identify the Nash equilibrium outcomes of FC mechanisms (Section 6.1) and study the implications on student welfare (Section 6.2).

6.1 Identification of Equilibrium Outcomes

We first extend our definition of Nash equilibrium to mixed problems with both sophisticated and sincere students.

Definition 6.1. Given a problem $(P, \succ)$, a set of sophisticated students $A \subseteq I$, and a mechanism φ, a preference profile P^* is an *A-Nash equilibrium (of φ at $(P, \succ)$)* if $P^*_i = P_i$ for all sincere students $i \in I \backslash A$ and $\varphi_i(P^*, \succ)\ R_i\ \varphi_i((P'_i, P^*_{-i}), \succ)$ for all sophisticated students $i \in A$ and all preference orders P'_i.

In words, in an A-Nash equilibrium, all sincere students report truthfully and no sophisticated student can benefit by unilaterally deviating from the equilibrium profile. Pathak and Sönmez [29] showed that the A-Nash equilibrium outcomes of BM correspond to the matchings that are stable with respect to the true preferences and an *augmented priority profile*. These augmented priorities capture three intuitive aspects of BM: First, students who have been accepted keep their seats, even if they have lower priority than another student who applies in a later round of BM. The fact that a student ranks a school higher thus overrules the fact that her priority at that school may be lower. Augmented priorities capture this by giving higher priority to students who rank a school higher. Second, while sincere students report their preferences truthfully, sophisticated students can misreport their preferences. In particular, they can rank any school first. Augmented priorities reflect this advantage by treating sophisticated students as if they ranked every school first. Third, BM breaks ties according to the original priority profile, and augmented priorities do the same. The next definition formalizes these aspects (where I^s_k denotes the set of students who rank s in k^{th} position according to P).

Definition 6.2. Given a problem $(P, \succ)$ and a set of sophisticated students $A \subseteq I$, for each school $s \in S$, the *augmented priority order* $\widehat{\succ}_s$ is constructed as follows:

- $i \mathrel{\widehat{\succ}_s} j$ if $i \in A \cup I^s_1$ and $j \in I^s_2 \backslash A$
- $i \mathrel{\widehat{\succ}_s} j$ if $i \in I^s_k \backslash A$ and $j \in I^s_{k+1} \backslash A$ for any $k \geq 2$
- $i \mathrel{\widehat{\succ}_s} j$ if $i \succ_s j$ and either $i, j \in A \cup I^s_1$ or $i, j \in I^s_k \backslash A$ for any $k \geq 2$
- All undefined priorities are implied by transitivity

[21] A stable matching is *student-optimal* if all students prefer it to any other stable matching. These matchings are unique for the problems we consider [31].

A matching μ is *augmented stable* if it is stable with respect to the problem $(P, \widehat{\succ})$.

With the notions of A-Nash equilibrium and augmented stability, we can now formally restate the main result of Pathak and Sönmez [29].

Fact 2 (Proposition 1 of Pathak and Sönmez [29])**.** *Given a problem $(P, \succ)$ and a set of sophisticated students $A \subseteq I$, a matching μ is augmented stable if and only if there exists an A-Nash equilibrium P^* of BM at $(P, \succ)$ with $\mu = BM(P^*, \succ)$.*

In words, Fact 2 means that the set of A-Nash equilibrium outcomes of BM corresponds precisely to the set of augmented-stable matchings. Our next result extends this characterization to the entire class of FC mechanisms, albeit with one limitation: We identify the equilibrium outcomes of FC mechanisms only up to equivalence from the perspective of the sophisticated students. Formally, we say that two matchings μ and ν are *A-equivalent* if $\mu_i = \nu_i$ for all students $i \in A$, denoted $\mu =_A \nu$.

Theorem 6.3. *Given a problem $(P, \succ)$, a set of sophisticated students $A \subseteq I$, and an FC mechanism φ, a matching μ is A-equivalent to some augmented-stable matching if and only if there exists an A-Nash equilibrium P^* of φ at $(P, \succ)$ with $\mu =_A \varphi(P^*, \succ)$.*

We give the proof in Appendix E of the full version of this paper. In words, Theorem 6.3 shows that augmented stability describes the equilibrium outcomes of FC mechanisms in mixed problems from the perspective of the sophisticated students. To see why the limitation to sophisticated students is needed, recall that FC mechanisms are restricted in how they handle first choices but they are free in how they to handle other choices (Lemma 4.3). In equilibrium, this freedom only affects sincere students; sophisticated students can always get their equilibrium school by ranking it first (Lemma 4.4).

If all students are sincere, then the equilibrium outcomes of FC mechanisms trivially satisfy FCM and FCS with respect to the true preferences. With sophisticated students, this may no longer be true. Example 6.4 illustrates that FCM (with respect to the true preferences) can be violated even if there is just one sophisticated student.

Example 6.4. There are three students $I = \{1, 2, 3\}$ and two schools $S = \{a, b\}$ with one seat each, and only student 2 is sophisticated (i.e., $A = \{2\}$). The preferences and priorities are

$$\begin{aligned} P_i \text{ for } i \in \{1,2\}: &\quad a\ P_i\ b\ P_i\ \emptyset, \\ P_3: &\quad b\ P_3\ a\ P_3\ \emptyset, \\ \succ_a: &\quad 3 \succ_a 1 \succ_a 2, \\ \succ_b: &\quad 1 \succ_b 2 \succ_b 3. \end{aligned}$$

Notice that at most two students can be matched to their true first choices. But with $A = \{2\}$, the unique augmented-stable matching is μ with $\mu_1 = a$, $\mu_2 = b$, and $\mu_3 = \emptyset$. By Thm. 6.3, student 2 is matched to school b in any $\{2\}$-Nash equilibrium outcome of any FC mechanism. Thus, at most one student is matched to her first choice which violates FCM (with respect to the true preferences).

While FC mechanisms may not match a maximal number of true first choices in equilibrium, the next corollary shows that all equilibrium outcomes are FCS with respect to the true preferences.

Corollary 6.5. *Given a problem $(P, \succ)$, a set of sophisticated students $A \subseteq I$, and an FC mechanism φ, let P^* be an A-Nash equilibrium of φ at $(P, \succ)$. Then $\varphi(P^*, \succ)$ is FCS with respect to $(P, \succ)$.*

Proof. Assume towards contradiction that $\varphi(P^*, \succ)$ is not FCS with respect to the true problem $(P, \succ)$. Then there exists a pair $(i, s) \in I \times (S \cup \{\emptyset\})$ that blocks $\varphi(P^*, \succ)$ with $s = \text{choice}_{P_i}(1)$. If $i \in A$, then i could get matched to s by ranking it first, a contradiction to the assumption that P^* is an A-Nash equilibrium. If $i \notin A$, then i already ranks s first. However, first choices are matched respecting priorities under any FC mechanism (by Lemma 4.3). Thus, all students matched to s have priority over i at s, again a contradiction. □

In summary, we have found that the equilibrium outcomes of FC mechanisms in mixed problems always satisfy FCS with respect to the true preferences, but they may violate FCM.

6.2 Student Welfare in Equilibrium

So far, we have analyzed student welfare under FC mechanisms in two extreme cases: When *all students are sophisticated*, then FC mechanisms match at most as many students to their true first choices in equilibrium as DA;[22] when *all students are sincere*, then FC mechanisms are trivially FCM. But what happens in intermediate cases, when some but not all students are sincere?

Towards this question, we consider specific equilibrium outcomes that are unanimously preferred by all sophisticated students.[23] To make this formal, fix an FC mechanism φ, a set of sophisticated students $A \subseteq I$, and a problem $(P, \succ)$. By the theory of stable matchings [32], there exists a *student-optimal augmented-stable matching*, say $\widehat{\mu}$. By Theorem 6.3, there exists an A-Nash equilibrium P^* (of φ at $(P, \succ)$) such that the outcome $\varphi(P^*, \succ)$ is A-equivalent to $\widehat{\mu}$, and all sophisticated students (weakly) prefer the matching $\varphi(P^*, \succ)$ to all other A-Nash equilibrium matchings. Because of this unanimous preference by the sophisticated students for $\varphi(P^*, \succ)$, we now focus on these *A-optimal Nash equilibrium matchings.*

For our analysis of student welfare in intermediate cases (i.e., with some sincere students), we identify how an individual student's behavior impacts student welfare by comparing the respective A-optimal Nash equilibrium matchings. First, we discuss an example to build intuition.

Example 6.6. There are six students $I = \{1, \dots, 6\}$ and four schools $S = \{a, b, c, d\}$ with one seat each. The preferences and priorities are

$$\begin{aligned}
P_i \text{ for } i \in \{1,2,3\}: &\quad a\ P_i\ b\ P_i\ c\ P_i\ \emptyset,\\
P_4: &\quad a\ P_4\ b\ P_4\ c\ P_4\ d\ P_4\ \emptyset,\\
P_5: &\quad a\ P_5\ b\ P_5\ d\ P_5\ \emptyset,\\
P_6: &\quad b\ P_6\ c\ P_6\ \emptyset,\\
\succ_s \text{ for } s \in S: &\quad 1 \succ_s \dots \succ_s 6.
\end{aligned}$$

Suppose that the FC mechanism ABM is used to match students to schools. For the sets of sophisticated students $A = \{6\}$ and $A' = \{3, 6\}$, the equilibrium outcomes in each case are unique and given in the following table.

Sophisticated students	**Unique equilibrium matchings under ABM**					
$A = \{6\}$	$\mu_1 = a$,	$\mu_2 = c$,	$\mu_3 = \emptyset$,	$\mu_4 = \emptyset$,	$\mu_5 = d$,	$\mu_6 = b$
$A' = \{3, 6\}$	$\nu_1 = a$,	$\nu_2 = \emptyset$,	$\nu_3 = b$,	$\nu_4 = d$,	$\nu_5 = \emptyset$,	$\nu_6 = c$

This example illustrates four aspects of the relationship between μ and ν.

(1) The sophisticated student 6 prefers the outcome when student 3 is sincere because she strictly prefers $\mu_6 = b$ to $\nu_6 = c$.
(2) Student 3, who is either sophisticated or sincere, prefers the outcome when she is sophisticated because she strictly prefers $\nu_3 = b$ to $\mu_3 = \emptyset$.

[22] Assuming truthful reporting, which is a weakly dominant strategy under DA.
[23] Note that a comparison between equilibrium outcomes may, in general, be inconclusive. We focus on the equilibrium outcomes that are unanimously preferred by all sophisticated students because they are well motivated: the strategic students would select these equilibria if they were given a choice.

(3) Whether student 3 is sophisticated or sincere has ambiguous effects for sincere students: Students 2 and 5 prefer μ, student 4 prefers ν, and student 1 is indifferent between μ and ν.
(4) When more students are sincere, more student are matched to their true first choices: If students 3 and 6 are both sophisticated, then only student 1 is matched to her true first choice. However, if only student 6 is sophisticated, then students 1 and 6 are both matched to their true first choices.

The following theorem establishes that our observations about the relationship between μ and ν in Example 6.6 hold in general for all FC mechanisms and all mixed problems.

THEOREM 6.7. *Given a problem $(P, \succ)$, an FC mechanism φ, sophisticated students $A \subseteq I$, and a sincere student $i' \notin A$, let μ be an A-optimal Nash equilibrium matching and let ν be an A'-optimal Nash equilibrium matching where $A' = A \cup \{i'\}$. Then:*

1. *For all $i \in A$: $\mu_i \; R_i \; \nu_i$, i.e., all sophisticated students prefer μ to ν.*
2. *For i': $\nu_{i'} \; R_{i'} \; \mu_{i'}$, i.e., i' prefers ν to μ.*
3. *For $i \notin A \cup \{i'\}$: $\nu_i \; P_i \; \mu_i$, $\mu_i \; P_i \; \nu_i$, or $\nu_i = \mu_i$ are all possible, i.e., for all sincere students except i' the impact of the behaior of i' is ambiguous.*
4. *For all $i \in I$: If $\nu_i = choice_{P_i}(1)$, then $\mu_i = \nu_i$, i.e., any student who is matched to her true first choice under ν is also matched to her true first choice under μ.*

A formal proof is given in Appendix F of the full version of this paper. Our result generalizes the corresponding Proposition 4 of Pathak and Sönmez [29] in two ways: First, we extend its scope from BM to all FC mechanisms. Second, beyond the impact for sophisticated students and i', we also identify the impact for sincere students and for those students who are matched to their true first choices.

Remark 2. It is intuitive that strategizing improves the outcome for i' and harms the other sophisticated students. Thus, Statements 1 and 2 in Theorem 6.7 may appear trivial. However, this intuition is deceptive: By misreporting her preferences, i' changes the outcome for herself and others. But more importantly, she also changes the game for the other sophisticated students. They in turn respond by changing their own preference reports, which could deteriorate the outcome for i' in general. Showing that FC mechanisms do not induce such dynamics is precisely the contribution of Theorem 6.7, and the proof critically relies on the two properties FCM and FCS.

For market designers, Theorem 6.7 answers the question whether FC mechanisms can bring us closer to achieving first-choice maximality with respect to the true preferences. In the respective optimal equilibrium outcomes, the number of students who are matched to their true first choices is lowest if all students are sophisticated, increases when more students are sincere, and peaks when all students are sincere. This provides a potential justification for the use of FC mechanisms in practice: If administrators wish to match many students to their true first choices, if they care about a minimal fairness guarantee in terms of FCS, and if they expect some share of the students to report their preferences truthfully, then FC mechanisms may be an attractive design alternative.

Remark 3. This insight comes with a caveat: Robust predictions of equilibrium play are not the only reason for the appeal of strategyproofness. By simplifying participation for students, strategyproofness also alleviates the cognitive cost that sophisticated students incur when they strategize [6]. On top of that, strategyproofness yields another form of fairness because it levels the playing field between students with different levels of strategic sophistication [29]. Our finding that the use of non-strategyproof FC mechanisms may be justified by the fact that they may match more students to their true first choices is agnostic to this broader role of strategyproofness as a desideratum for market design.

7 A CHARACTERIZATION OF THE BOSTON MECHANISM

In this section, we present a characterization of BM, the FC mechanism most commonly used in practice. Our characterization is based on extensions of the properties FCM and FCS and the well-known axiom of consistency.

Consistency requires that removing any student and her assigned seat from a problem and applying the mechanism to this reduced problem leads to the same matchings for the remaining students to the remaining seats [24]. It is easy to see that removing students with their seats may lead to problems where some school has no capacity, a case that our formal model explicitly excludes (see Section 2.1). To properly define consistency, we therefore extend the model by allowing $q_s = 0$ for some schools $s \in S$.

Definition 7.1. A mechanism φ is *consistent* if, for all school choice problems $(I, S, q, P, \succ)$ and all students $i' \in I$, the following holds: Let $\mu = \varphi(I, S, q, P, \succ)$, and let $(I^-, S^-, q^-, P^-, \succ^-)$ be the reduced problem where $I^- = I \backslash \{i'\}$, $S^- = S$, $q^-_{\mu_{i'}} = q_{\mu_{i'}} - 1$, $q^-_s = q_s$ for all $s \in S \backslash \{\mu_{i'}\}$, $P^- = (P_i)_{i \in I^-}$, and $\succ^- = (\succ^-_s)_{s \in S^-}$, where $i \succ^-_s j$ whenever $i \succ_s j$ for all $i, j \in I^-$; then $\varphi_i(I, S, q, P, \succ) = \varphi_i(I^-, S^-, q^-, P^-, \succ^-)$ for all $i \in I^-$.

FCM and FCS can still be defined in the extended model. However, they become void if no student can be matched to their first choice, and void constraints are not useful in characterizing mechanisms. We therefore extend these properties in a different way: Instead of maximizing the number of matched first choices (as with FCM), we require that a mechanism maximizes the number of students matched to their k^{th} choice where k is the best rank for which this maximization is not trivial, and we take the same approach to extend FCS.

Definition 7.2. Given a problem $(P, \succ)$, let k^{best} be the minimal rank for which there exist a student $i \in I$ and a school $s \in S$ such that $\text{choice}_{P_i}\left(k^{\text{best}}\right) = s$ and $q_s \geq 1$. A matching μ is *best-choice maximal* if there exists no other matching ν such that $|I\left(\mu, k^{\text{best}}, P\right)| < |I\left(\nu, k^{\text{best}}, P\right)|$; and μ is *best-choice stable* if it is k^{best}*-choice stable* (see Definition 3.2). A mechanism φ is *best-choice maximal* (*best-choice stable*) if, for all problems $(P, \succ)$, the matching $\varphi(P, \succ)$ is best-choice maximal (best-choice stable).

Observe that FCM and best-choice maximality coincide for problems where all schools have at least one seat; the same is true for FCS and best-choice stability. Both properties can thus be understood as extensions of FCM and FCS, respectively. With these definitions and the extended model (where schools can have zero capacity) we can state our characterization result.

Theorem 7.3. *A mechanism φ is best-choice maximal, best-choice stable, and consistent if and only if it is the Boston mechanism.*

Proof. Best-choice stability and best-choice maximality of BM are an immediate consequence of its definition; and Kojima and Ünver [24] showed consistency of BM. Conversely, assume towards contradiction that a mechanism ψ is best-choice stable, best-choice maximal, and consistent but $\psi(I, S, P, \succ, q) \neq BM(I, S, q, P, \succ)$ for some problem $(I, S, q, P, \succ)$. First, observe that any best-choice maximal and best-choice stable mechanism assigns the same set of students to their k^{best} choices for any problem. To see this, note that if $k^{\text{best}} > 1$, then we can update the preferences of all students in I by truncating the first $k - 1$ choices and apply Lemma 4.3. We can remove these students with their assignments and define the reduced problem $(I^-, S^-, q^-, P^-, \succ^-)$ (as in Definition 7.1). By consistency of ψ, it matches the remaining students to the same schools under both $(I, S, q, P, \succ)$ and $(I^-, S^-, q^-, P^-, \succ^-)$; and the same is true for BM. Repeating this argument, we eventually arrive at a trivial problem that admits only one feasible matching (where no student is matched to any school).

The matchings from both mechanisms must coincide on this problem. Then, by construction, they must also coincide on all prior problems, a contradiction. □

Theorem 7.3 establishes the extraordinary role of BM in the set of FC mechanisms: While it is not the only FCM and FCS mechanism, it is essentially the only such mechanism that is also consistent (albeit in an extended model and with a special extension of the axioms to this model).

8 CONCLUSION

In school choice, FC mechanisms arise naturally, and many school choice mechanisms used in practice belong to this class. Understanding this class is therefore important for researchers and practitioners alike. The class of FC mechanisms is fairly large because FC mechanisms are only restricted in how they handle (reported) first choices. Nonetheless, we were able to analyze the incentives and equilibria of all mechanisms in this class purely based on the defining properties FCM and FCS. This suggests FCM and FCS as useful concepts analyzing school choice markets.

Our findings contribute to an ongoing debate about the respective merits and shortcomings of BM and DA. On the one hand, DA has the obvious advantage of being strategyproof [4], and its outcome weakly Pareto dominates the Nash equilibrium outcomes of BM in settings with strict priorities and when all students strategize optimally [18]. However, the comparison becomes less clear when either of the assumptions are relaxed. When priorities are weak, the equilibrium outcomes of BM can dominate those of DA from an *ex-ante* perspective [1, 17, 27]. Furthermore, if all students are sincere, BM *rank dominates* DA whenever they are comparable by rank dominance [22, 26], and BM satisfies the welfare property of *favoring higher ranks* [24]. Our present paper adds to these insights: If first choices matter and if a share of the students is expected to be sincere, then any FC mechanism (including BM) may yield more appealing matchings in equilibrium than DA. Our characterization result further shows that BM takes an elevated position among the FC mechanisms. Nevertheless, we refrain from recommending any specific mechanism in general; instead, our findings highlight the implicit trade-offs one must make when choosing between any of the mechanisms we have studied.

The prevalence of non-strategyproof mechanisms in practice has inspired new ways of thinking about incentives. For example, the concept for comparing mechanisms by their vulnerability to manipulation, put forward by [30], was instrumental in identifying a trend towards better incentive properties in the USA, UK, and Ghana. Interestingly, our results yield a criticism of this concept because it does not identify the intuitive differences between BM and ABM. To distinguish these two mechanisms by their incentive properties, more nuanced approaches are necessary (e.g., [15] for the case when schools find some students unacceptable or [26] when priorities are random).

Finally, our results give rise to promising directions for future research: Whether or not FC mechanisms can outperform DA hinges on the share of sophisticated students and their ability to coordinate on equilibrium. Some recent work has already considered field data to study student behavior and outcomes in school choice markets [9, 14, 16]. However, further research focusing specifically on sincerity and coordination would constitute an important contribution to the debate about school choice mechanisms, and we are aware of ongoing efforts in this direction.

REFERENCES

[1] Atila Abdulkadiroğlu, Yeon-Koo Che, and Yosuke Yasuda. 2011. Resolving Conflicting Preferences in School Choice: The Boston Mechanism Reconsidered. *American Economic Review* 101, 1 (2011), 399--410.

[2] Atila Abdulkadiroğlu, Parag Pathak, and Alvin Roth. 2005. The New York City High School Match. *American Economic Review* 95, 2 (May 2005), 364--367.

[3] Atila Abdulkadiroğlu, Alvin Roth, Parag Pathak, and Tayfun Sönmez. 2006. Changing the Boston School Choice Mechanism: Strategy-proofness as Equal Access. (2006). Working Paper.

[4] Atila Abdulkadiroğlu and Tayfun Sönmez. 2003. School Choice: A Mechanism Design Approach. *American Economic Review* 93, 3 (2003), 729--747.
[5] Jose Alcalde. 1996. Implementation of Stable Solutions to Marriage Problems. *Journal of Economic Theory* 69, 1 (1996), 240--254.
[6] Eduardo Azevedo and Eric Budish. 2015. Strategy-proofness in the Large. (2015). Working Paper.
[7] Michel Balinski and Tayfun Sönmez. 1999. A Tale of Two Mechanisms: Student Placement. *Journal of Economic Theory* 84, 1 (1999), 73--94.
[8] Christian Basteck, Katharina Huesmann, and Heinrich Nax. 2015. Matching Practices for Secondary Schools - Germany. (2015). matching-in-practice.eu matching-in-practice.eu, accessed May 16, 2015.
[9] Caterina Calsamiglia and Maia Güell. 2014. The Illusion of School Choice: Empirical Evidence from Barcelona. (2014). Working Paper.
[10] Li Chen and Juan Sebastián Pereyra. 2015. Self-selection in School Choice. (2015). Working Paper.
[11] Yan Chen and Onur Kesten. 2016. Chinese College Admissions and School Choice Reforms: A Theoretical Analysis. (2016). Working Paper.
[12] Yan Chen and Onur Kesten. 2016. Chinese College Admissions and School Choice Reforms: An Experimental Study. (2016). Working Paper.
[13] Yan Chen and Tayfun Sönmez. 2006. School Choice: an Experimental Study. *Journal of Economic Theory* 127, 1 (2006), 202--231.
[14] Monique de Haan, Pieter Gautier, Hessel Oosterbeek, and Bas van der Klaauw. 2015. The Performance of School Assignment Mechanisms in Practice. (2015). Working Paper.
[15] Umut Dur. 2015. The Modified Boston Mechanism. (2015). Working Paper.
[16] Umut Dur, Robert Hammond, and Thayer Morrill. 2016. The Secure Boston Mechanism: Theory and Experiments. (2016). Working Paper.
[17] Aytek Erdil and Haluk Ergin. 2008. What's the Matter with Tie-Breaking? Improving Efficiency in School Choice. *American Economic Review* 98, 3 (2008), 669--89.
[18] Haluk Ergin and Tayfun Sönmez. 2006. Games of School Choice Under the Boston Mechanism. *Journal of Public Economics* 90, 1-2 (2006), 215--237.
[19] Clayton Featherstone. 2015. Rank Efficiency : Investigating a Widespread Ordinal Welfare Criterion. (2015). Working Paper.
[20] David Gale and Lloyd Shapley. 1962. College Admissions and the Stability of Marriage. *The American Mathematical Monthly* 69, 1 (1962), 9--15.
[21] Guillaume Haeringer and Flip Klijn. 2009. Constrained School Choice. *Journal of Economic Theory* 144, 5 (2009), 1921--1947.
[22] Patrick Harless. 2015. A School Choice Compromise: Between Immediate and Deferred Acceptance. (2015). Working Paper.
[23] Paula Jaramillo, Cağatay Kayi, and Flip Klijn. 2016. School Choice: Nash Implementation of Stable Matchings Through Rank-Priority Mechanisms. (2016). Working Paper.
[24] Fuhito Kojima and Utku Ünver. 2014. The Boston School-Choice Mechanism: an Axiomatic Approach. *Economic Theory* 55, 3 (2014), 515--544.
[25] Timo Mennle and Sven Seuken. 2017. Partial Strategyproofness: Relaxing Strategyproofness for the Random Assignment Problem. (2017). Working Paper.
[26] Timo Mennle and Sven Seuken. 2017. Trade-offs in School Choice: Comparing Deferred Acceptance, the Naïve and the Adaptive Boston Mechanism. (2017). Working Paper.
[27] Antonio Miralles. 2008. School Choice: the Case for the Boston Mechanism. (2008). Working Paper.
[28] Joana Pais and Agnes Pinter. 2008. School Choice and Information: an Experimental Study on Matching Mechanisms. *Games and Economic Behavior* 64 (2008), 303--328.
[29] Parag Pathak and Tayfun Sönmez. 2008. Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism. *American Economic Review* 98, 4 (2008), 1636--52.
[30] Parag Pathak and Tayfun Sönmez. 2013. School Admissions Reform in Chicago and England: Comparing Mechanisms by Their Vulnerability to Manipulation. *American Economic Review* 103(1) (2013), 80--106.
[31] Alvin Roth. 1982. The Economics of Matching: Stability and Incentives. *Mathematical Operations Research* 7 (1982), 617--628.
[32] Alvin Roth and Marilda Sotomayor. 1990. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Cambridge University Press.
[33] Alexander Westkamp. 2013. An Analysis of the German University Admissions System. *Economic Theory* 53, 3 (2013), 561--589.