# COALITIONAL MANIPULABILITY IN SCHOOL CHOICE

EMMA MORENO-GARCÍA AND JUAN PABLO TORRES-MARTÍNEZ

Abstract. We compare the degree of coalitional manipulability across stable school-choice mechanisms. Although no stable mechanism is group strategy-proof for all school choice contexts, we show that the student-optimal stable mechanism is the least manipulable by coalitions, whereas the school-optimal stable mechanism is the most coalitionally manipulable. These results still hold when students can only report truncated preferences, when the set of coalitions is limited, or when schools cannot be declared unacceptable as long as there are more available seats than students. In contrast, if there is a shortage of seats and students have no outside options, the school-optimal stable mechanism can be less manipulable by coalitions than the student-optimal stable mechanism.

Keywords: School choice, coalitional manipulability, mechanism design.
JEL Classification: D47, C78.

## 1. Introduction

In recent years, centralized admission systems have been implemented in many countries to assign students to schools at different educational levels (Pathak, 2017; Neilson, 2024). Among the mechanisms used to place students in schools, the *student-optimal stable mechanism*—denoted by DA—is the only one that is stable and strategy-proof (Alcalde and Barberà, 1994). Stability ensures that, after assignments

*Date*: March, 2026

Emma Moreno-García (emmam@usal.es)
Departamento de Economía e Historia Económica, Universidad de Salamanca, Spain

Juan Pablo Torres-Martínez (juan.torres@fen.uchile.cl)
Department of Economics, Faculty of Economics and Business, University of Chile.

We thank the Editor Federico Echenique, two anonymous referees, Gabriela Denis, and Juan Pereyra for their helpful comments and suggestions. Emma Moreno-García acknowledges the support by the Research Grants PID2022-136718NB-I00 (Ministerio de Economía e Innovación) and SA094P24 (Junta de Castilla y León). Juan Pablo Torres-Martínez acknowledges the hospitality of the University of Salamanca where part of this work was carried out.

are made, no student prefers a school with available seats or where she has a higher priority than an admitted candidate. Strategy-proofness ensures that no student can benefit from misrepresenting preferences. Despite these properties, the outcome of DA may be inefficient, and this cannot be mitigated without compromising stability or strategy-proofness (Gale and Shapley, 1962; Abdulkadiroğlu, Pathak, and Roth, 2009; Kesten, 2010). Underlying the inefficiency of DA is the fact that it can be manipulated by groups of students (Ergin, 2002). Indeed, whenever the assignment determined by DA is inefficient, there are groups of students with incentives to misrepresent their preferences, improving at least one of their members without harming the others.

Hence, DA is immune to individual manipulations, but not to coalitional manipulations. In fact, no stable mechanism is immune to coalitional manipulation. Therefore, it is arguably more important to understand the degree of coalitional manipulability of stable mechanisms than it is to study the degree of individual manipulability. The literature has made progress on the latter, but our work is the first to study the former.

From the individual manipulability perspective, DA is the least manipulable stable mechanism, while the *school-optimal stable mechanism*—denoted by $\text{DA}^{\text{S}}$—is as manipulable as any other stable mechanism under many comparability criteria (Chen, Egesdal, Pycia, and Yenmez, 2016; Pathak and Sönmez, 2013). Similar results are obtained in frameworks where no stable mechanism is strategy-proof, such as when the number of alternatives that students can report is limited or when both market sides may act strategically (Bonkoungou and Nesterov, 2023; Cai, 2025; Sirguiado, 2025). However, the comparison of stable mechanisms from a coalitional manipulability viewpoint has remained unexplored. Notice that, to avoid coalitional manipulability, two properties are required: strategy-proofness and non-bossiness.[1] DA only satisfies the former (Dubins and Freedman, 1981; Roth, 1982), while $\text{DA}^{\text{S}}$ only satisfies the latter (Roth, 1982; Afacan and Dur, 2017). Thus, it is not obvious which of these mechanisms is less coalitionally manipulable.

[1] Pápai (2001) shows that group strategy-proofness is equivalent to the combination of strategy-proofness and non-bossiness. This last property ensures that no student can misrepresent her preferences to change the placement of others without modifying her own.

In this paper, we study coalitional manipulability of stable mechanisms in classical school choice contexts (Abdulkadiroğlu and Sönmez, 2003) and also in scenarios with truncated preferences or no outside options. Coalitional manipulability is measured by taking into account the groups of students who can misrepresent their preferences, the improvements that members of these groups can attain, and the preference profiles at which manipulations can be performed. A dozen comparability criteria inducing partial orders over the set of stable mechanisms are considered, many of which are extensions of criteria proposed in the literature on individual manipulability (see Tables 1 and 2). We indicate by $\Psi \geqslant_{\mathbf{d}} \Phi$ that a mechanism $\Psi$ is as coalitionally manipulable as another mechanism $\Phi$ under a comparability criterion $\geqslant_{\text{d}}$.

| **Criterion** | **Meaning of $\Psi \geqslant_{\mathbf{d}} \Phi$** |
|---|---|
| $\geqslant^*_{\text{PS}}$ | [Pathak and Sönmez, 2013] At each preference profile, every coalition that manipulates $\Phi$ also manipulates $\Psi$. |
| $\geqslant_{\text{PS}}$ | [Pathak and Sönmez, 2013] At every preference profile in which $\Phi$ is coalitionally manipulable so is $\Psi$. |
| $\geqslant_{\text{CEPY}}$ | [Chen, Egesdal, Pycia, and Yenmez, 2016] At each preference profile, every attainable improvement for a coalition under $\Phi$ is an attainable improvement for it under $\Psi$.[2] |
| $\geqslant_{\text{AM}}$ | [Arribillaga and Massó, 2016] For each coalition, every truthful dominant strategy under $\Psi$ is a truthful dominant strategy under $\Phi$.[3] |
| $\geqslant^*_{\text{BN}}$ | [Bonkoungou and Nesterov, 2021] Every attainable improvement for a coalition under $\Phi$ is an attainable improvement for it under $\Psi$. |
| $\geqslant_{\text{BN}}$ | [Bonkoungou and Nesterov, 2023; Imamura and Tomoeda, 2022] At each preference profile, as many coalitions can manipulate $\Psi$ as $\Phi$. |
| $\geqslant_{\text{NRR}}$ | [Nesterov, Rospuskova, and Rubtcova, 2024] Given a coalition $C$ and a preference profile for its members, every attainable improvement for $C$ under $\Phi$ is an attainable improvement for it under $\Psi$. |

Table 1. Comparison criteria based on the literature on individual manipulability.

[2]As pointed out by Chen, Egesdal, Pycia, and Yenmez (2016), $\geqslant_{\text{CEPY}}$ is the ordinal counterpart of the cardinal concept of *"intense and strong manipulability"* studied by Pathak and Sönmez (2013).

[3]Decerf and Van der Linden (2021) propose a criterion analogous to $\geqslant_{\text{AM}}$, but focusing on dominant strategies instead of truthful dominant strategies. They define manipulable mechanisms as those that *"[...] sometimes fail to provide students with dominant strategies"*. In our context, manipulability

| **Criterion** | **Meaning of $\Psi \geqslant_{\mathrm{d}} \Phi$** |
|---|---|
| $\geqslant_{\mathrm{D1}}$ | Given a coalition $C$ and preferences for its members, there are as many preference profiles at which $C$ can manipulate $\Psi$ as preference profiles at which it can manipulate $\Phi$. |
| $\geqslant_{\mathrm{D2}}$ | Given a coalition $C$, there are as many preference profiles at which $C$ can manipulate $\Psi$ as preference profiles at which it can manipulate $\Phi$. |
| $\geqslant_{\mathrm{D3}}$ | Every coalition that manipulates $\Phi$ also manipulates $\Psi$. |
| $\geqslant_{\mathrm{D4}}$ | There are as many pairs $(P, C)$ such that $C$ can manipulate $\Psi$ at $P$, as pairs $(P', C')$ such that $C'$ can manipulate $\Phi$ at $P'$. |
| $\geqslant_{\mathrm{D5}}$ | There are as many preference profiles at which $\Psi$ can be manipulated as preference profiles at which $\Phi$ can be manipulated. |

TABLE 2. New criteria to compare coalitional manipulability.

The relationships between the different comparability criteria are described in Figure 1 and can be easily verified using the formal definitions provided in Section 4. We denote by $>_{\mathrm{d}}$ the strict comparability criterion induced by $\geqslant_{\mathrm{d}}$. Hence, $\Psi >_{\mathrm{d}} \Phi$ indicates that $\Psi$ is more coalitionally manipulable than $\Phi$ under $\geqslant_{\mathrm{d}}$ (i.e., $\Psi >_{\mathrm{d}} \Phi$ if and only if $\Psi \geqslant_{\mathrm{d}} \Phi$ and $\Phi \not\geqslant_{\mathrm{d}} \Psi$).[4]

We consider constrained school choice contexts (Haeringer and Klijn, 2009), situations in which students are restricted to report a limited number of schools as acceptable. We focus on *$\kappa$-stable mechanisms*, centralized protocols that associate each preference profile with a stable matching of the problem where only the best $\kappa \geq 1$ alternatives of each student are taken into account. Evidently, when $\kappa$ is greater than the number of schools, a $\kappa$-stable mechanism is just a stable mechanism.

In constrained school choice contexts, the $\kappa$-stable mechanisms induced by DA and $\mathrm{DA}^{\mathrm{S}}$—denoted by $\mathrm{DA}^{\kappa}$ and $\mathrm{DA}^{\mathrm{S},\kappa}$, respectively—are extreme alternatives concerning coalitional manipulability. More precisely, for each comparability criterion, $\mathrm{DA}^{\kappa}$ minimizes coalitional manipulability within the $\kappa$-stable mechanisms, while $\mathrm{DA}^{\mathrm{S},\kappa}$

refers to incentives to misrepresent preferences. Therefore, only by considering *truthful* dominant strategies we can induces a clear criteria to compare manipulability.

[4]Notice that, $\Psi \geqslant_{\mathrm{d}} \Phi \Rightarrow \Psi \geqslant_{\mathrm{d}'} \Phi$ does not necessarily ensure that $\Psi >_{\mathrm{d}} \Phi \Rightarrow \Psi >_{\mathrm{d}'} \Phi$, because it may be easier to compare under $\geqslant_{\mathrm{d}'}$ than under $\geqslant_{\mathrm{d}}$. Since more demanding criteria make it easier to distinguish mechanisms that can be compared, if $\geqslant_{\mathrm{d}}$ is more demanding than $\geqslant_{\mathrm{d}'}$ and $\Psi \geqslant_{\mathrm{d}} \Phi$, then $\Psi >_{\mathrm{d}'} \Phi \Rightarrow \Psi >_{\mathrm{d}} \Phi$.

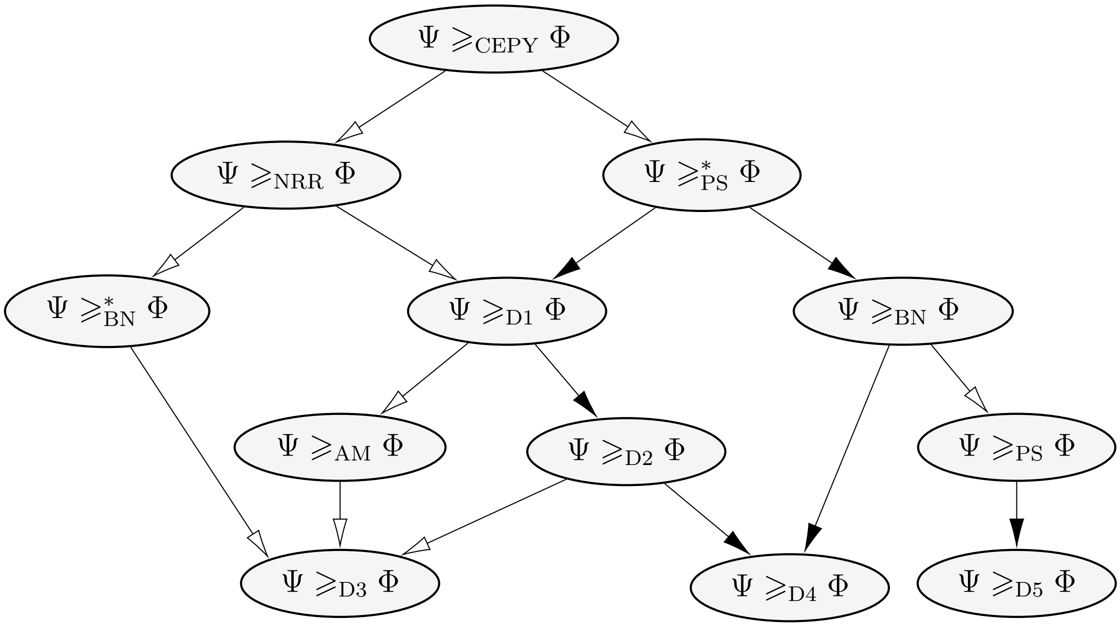


FIGURE 1. Relationships between the comparability criteria.

(The symbol $\longrightarrow$ indicates that the property $\Psi >_{\mathrm{d}} \Phi \Rightarrow \Psi >_{\mathrm{d}'} \Phi$ also holds)

maximizes it (Theorem 1). We deduce that each stable mechanism is as coalitionally manipulable as DA and at most as coalitionally manipulable as $\mathrm{DA}^{\mathrm{S}}$. This validates the robustness of DA concerning manipulability, even when facing group deviations.

Building on the previous findings, we refine the analysis to show that $\mathrm{DA}^{\kappa}$ and $\mathrm{DA}^{\mathrm{S},\kappa}$ are, respectively, the least and most coalitionally manipulable $\kappa$-stable mechanisms under the comparability criterion $\geqslant_{\mathrm{CEPY}}$ (Theorem 2). Furthermore, by appealing to the strategy-proofness of DA, we show that $\mathrm{DA}^{\kappa}$ is the unique $\kappa$-stable mechanism that minimizes coalitional manipulability under many of our comparability criteria (Theorem 3).

Unlike what happens for individual manipulability, a truncation of preferences does not necessarily increase coalitional manipulability when DA is implemented (cf., Pathak and Sönmez, 2013). Essentially, by truncating preferences, a student may begin manipulating to avoid exclusion from certain schools, but a coalition may stop manipulating when one of its members no longer acts as an *interrupter* (i.e., a student who prevents others from being assigned to a school that ultimately rejects her proposal when DA is implemented (cf., Kesten, 2010)). Depending on the preference profile, the first effect can outweigh the second, or vice versa. As a consequence, there are school choice contexts in which $\mathrm{DA}^1$ and DA are incomparable in terms of coalitional manipulability under most of our comparability criteria (see Propositions 1 and 2).

As a byproduct of the arguments in the proofs of Theorems 1 and 2 we obtain that, within the set of $\kappa$-stable mechanisms, $\Psi >_{\mathrm{CEPY}} \Phi$ if and only if $\Phi$ Pareto dominates $\Psi$, whereas $\Psi \geqslant_{\mathrm{CEPY}} \Phi$ if and only if $\Phi$ weakly Pareto dominates $\Psi$ (Theorem 4). Additionally, we show that these equivalences hold regardless of the set of feasible coalitions as long as every student can unilaterally misrepresent her preferences.

We also analyze the validity of our previous results in contexts where students have no outside options, and the mechanism designer is aware of this. The properties of coalitional manipulability that DA and $\mathrm{DA}^{\mathrm{S}}$ satisfy in classical school choice contexts still hold without outside options when there are more available seats than students (Theorem 5). However, when there is a shortage of seats, it is possible to find school choice contexts in which $\mathrm{DA}^{\mathrm{S}}$ is group strategy-proof and DA is coalitionally manipulable (Proposition 3).

The rest of the paper is organized as follows. Section 2 relates our results with the literature on individual manipulability. Section 3 describes classical and constrained school choice models. Section 4 provides formal definitions of the comparability criteria used in this work. Section 5 studies the coalitional manipulability of $\kappa$-stable mechanisms. Section 6 compares DA with $\mathrm{DA}^1$ in terms of coalitional manipulability. Section 7 focuses on the relationship between Pareto dominance and $\geqslant_{\mathrm{CEPY}}$. Section 8 studies coalitional manipulability in contexts with no outside options. Finally, Section 9 provides some remarks on topics for future research.

## 2. Related literature

There is an extensive literature on school choice comparing mechanisms in terms of their *individual manipulability*. The seminal work of Pathak and Sönmez (2013) on school admission reforms in Chicago and England, has given rise to numerous studies on manipulability measures to compare school admission systems with each other or with alternatives that satisfy interesting normative properties (cf., Arribillaga and Massó, 2016; Chen, Egesdal, Pycia, and Yenmez, 2016; Chen and Kesten, 2017; Dur, 2019; Dur, Hammond, and Morrill, 2019; Van der Linden, 2019; Bonkoungou and Nesterov, 2021, 2023; Decerf and Van der Linden, 2021; Dur, Pathak, Song, and Sönmez, 2022; Imamura and Tomoeda, 2022; Lomakin, Minibaev, and Nesterov, 2024; Nesterov, Rospuskova, and Rubtcova, 2024).

The work closest to ours is that of Chen, Egesdal, Pycia, and Yenmez (2016), who show that lower individual manipulability according to $\geqslant_{\mathrm{CEPY}}$ is equivalent to Pareto dominance. We extend this result to coalitional manipulability, including constrained school choice contexts and scenarios with no outside options. Furthermore, for the dozen comparability criteria considered in this work, we show that $\mathrm{DA}^{\kappa}$ and $\mathrm{DA}^{\mathrm{S},\kappa}$ are extreme alternatives within the set of $\kappa$-stable mechanisms in terms of coalitional manipulability. In particular, DA and $\mathrm{DA}^{\mathrm{S}}$ are, respectively, the stable mechanisms that minimize and maximize coalitional manipulability in classical school choice contexts. For individual manipulability, this last result follows from the strategy-proofness of DA, the fact that $\mathrm{DA}^{\mathrm{S}}$ is the student-pessimal stable mechanism, and the relationship between $\geqslant_{\mathrm{CEPY}}$ and Pareto dominance. Moreover, Pathak and Sönmez (2013) show that $\mathrm{DA}^{\mathrm{S}}$ is as individually manipulable as any other stable mechanism under $\geqslant^{*}_{\mathrm{PS}}$, while Bonkoungou and Nesterov (2023) show that $\mathrm{DA}^{\kappa}$ is at most as individually manipulable as any other $\kappa$-stable mechanism under $\geqslant^{*}_{\mathrm{PS}}$.[5]

We also show that, for every comparability criterion, DA minimizes coalitional manipulability among the stable mechanisms in scenarios without outside options when there are more available seats than students, while $\mathrm{DA}^{\mathrm{S}}$ maximizes it. Cai (2025) and Sirguiado (2025) show a similar result in two-sided one-to-one matching markets in which all agents may act strategically: without outside options, the

[5]Bonkoungou and Nesterov (2023) focus on the comparability criterion $\geqslant_{\mathrm{BN}}$ and allow schools to manipulate priorities and capacities. However, Step 1 in the proof of their Theorem 1 establishes that, at each preference profile, students who manipulate $\mathrm{DA}^{\kappa}$ also manipulate any other $\kappa$-stable mechanism.

man-proposing deferred acceptance mechanism is less individually manipulable than any other stable mechanism under $\geqslant_{\mathrm{BN}}$ if and only if there are fewer men than women.

Much of the previous literature on individual manipulability focuses on comparing $\mathrm{DA}^\kappa$ with truncated versions of the *Immediate Acceptance mechanism* (IA) and the *First Preference First mechanism* (FPF). The objective is to understand the effects on incentives of some regulatory changes in the United States and England, which led to the abandonment of IA and FPF in favor of DA, but keeping the number of alternatives that can be reported restricted (see Table 1 in Pathak and Sönmez, 2013). In this way, when there are at least $\kappa > 1$ schools, $\mathrm{IA}^\kappa$ is more individually manipulable than $\mathrm{DA}^\kappa$ under $\geqslant_{\mathrm{PS}}$ and $\geqslant_{\mathrm{BN}}$ (Pathak and Sönmez, 2013; Imamura and Tomoeda, 2022; Bonkoungou and Nesterov, 2023). These results imply that $\mathrm{IA}^\kappa >_{\mathrm{D4}} \mathrm{DA}^\kappa$ and $\mathrm{IA}^\kappa >_{\mathrm{D5}} \mathrm{DA}^\kappa$ for all $\kappa > 1$, because the comparability criteria for individual manipulability satisfy the relationships described in Figure 1. Furthermore, the mechanisms $\mathrm{IA}^\kappa$ and $\mathrm{FPF}^\kappa$ are more individually manipulable than $\mathrm{DA}^\kappa$ under $\geqslant^*_{\mathrm{BN}}$, $\geqslant_{\mathrm{AM}}$, and $\geqslant_{\mathrm{NRR}}$ (cf., Bonkoungou and Nesterov, 2021; Decerf and Van der Linden, 2021; Nesterov, Rospuskova, and Rubtkova, 2024).

It is important to remark that, in order to compare $\mathrm{DA}^\kappa$ with mechanisms as $\mathrm{IA}^\kappa$ and $\mathrm{FPF}^\kappa$, the literature of individual manipulability has used a property that is a direct consequence of the strategy-proofness of DA: the truncated mechanism $\mathrm{DA}^\kappa$ is only manipulated by those who would get a place in a school under DA, but are left alone when $\mathrm{DA}^\kappa$ is implemented.[6] From the coalitional manipulability perspective, it is not easy to identify with such precision the groups of students who have incentives to misrepresent their preferences. This limits our ability to compare the coalitional manipulability of $\mathrm{DA}^\kappa$ with that of other mechanisms that are not $\kappa$-stable.

In the literature of individual manipulability, there are also results comparing truncated versions of DA with each other. For instance, Pathak and Sönmez (2013) show that the more restricted the preferences, the greater the individual manipulation: $\mathrm{DA}^\rho >_{\mathrm{PS}} \mathrm{DA}^\kappa$ for all $\rho < \kappa$. The same result holds for the comparability criteria $\geqslant^*_{\mathrm{BN}}$, $\geqslant_{\mathrm{AM}}$, $\geqslant_{\mathrm{NRR}}$, and $\geqslant_{\mathrm{BN}}$ (Bonkoungou and Nesterov, 2021,

[6]Recently, Bonkoungou and Nesterov (2025) characterize the relationship between manipulating students and blocking students of the truncated mechanisms $\mathrm{DA}^\kappa$, $\mathrm{IA}^\kappa$, and $\mathrm{FPF}^\kappa$.

2023; Decerf and Van der Linden, 2021; Imamura and Tomoeda, 2022; Nesterov, Rospuskova, and Rubtkova, 2024). However, this result is not valid in our framework, since $\text{DA}^1$ and DA may be incomparable in terms of their coalitional manipulability under most comparability criteria.

## 3. Model

In a **school choice context** $(I, S, \succ, q)$ there is a finite set $I$ of students and a finite set $S$ of schools, with $|I| = n$. Each school $s$ is characterized by a strict linear order $\succ_s$ representing its priority for students, and by a capacity $q_s \geq 1$. Let $\succ = (\succ_s)_{s\in S}$ be the profile of priority orders and $q = (q_s)_{s\in S}$ the vector of capacities. Each student $i \in I$ has a complete, transitive, and strict preference $P_i$ defined on $S \cup \{s_0\}$, where $s_0$ denotes an outside option. Hence, a school $s$ is acceptable for the student $i$ whenever $sP_is_0$. We denote by $R_i$ the weak preference associated with $P_i$. Let $\mathcal{P}$ be the set of strict linear orders defined on $S \cup \{s_0\}$ and $P = (P_i)_{i\in I} \in \mathcal{P}^n$ the profile of students' preferences. We refer to $(I, S, \succ, q, P)$ as a **school choice problem**.

A **matching** is a function $\mu : I \to S \cup \{s_0\}$ that determines an assignment $\mu(i)$ for each student $i$, and satisfies $|\mu^{-1}(s)| \leq q_s$ for all $s \in S$, where $\mu^{-1}(s) = \{i \in I : \mu(i) = s\}$. Let $\mathcal{M}$ denote the set of matchings. A matching $\mu$ is individually rational at $P$ when $\mu(i)R_is_0$ for every $i \in I$. Moreover, $\mu$ is stable at $P$ when it is individually rational, and there is no pair $(i, s) \in I \times S$ such that $sP_i\mu(i)$ and either $|\mu^{-1}(s)| < q_s$ or $i \succ_s j$ for some student $j$ assigned to $s$ under $\mu$.

In a school choice context $(I, S, \succ, q)$, a **mechanism** is a function $\Phi : \mathcal{P}^n \to \mathcal{M}$ that associates a matching with each preference profile. We denote by $\Phi_i(P)$ the assignment of student $i$ when preferences are $P$. Analogously, $\Phi_s(P)$ are the students assigned to $s \in S \cup \{s_0\}$ under $P$.

Given mechanisms $\Phi, \Psi : \mathcal{P}^n \to \mathcal{M}$, we say that $\Phi$ **Pareto dominates** $\Psi$ when for all $P \in \mathcal{P}^n$ and $i \in I$ we have that $\Phi_i(P)R_i\Psi_i(P)$, and there exist $P' \in \mathcal{P}^n$ and $j \in I$ such that $\Phi_j(P')P'_j\Psi_j(P')$. Moreover, $\Phi$ **weakly Pareto dominates** $\Psi$ when either $\Phi = \Psi$ or $\Phi$ Pareto dominates $\Psi$.

To capture models in which students can report at most $\kappa \geq 1$ schools as acceptable, it suffices to assume that preferences can be truncated before determining the final assignment. Given a preference profile $P = (P_i)_{i \in I}$, let $P_i^\kappa$ be the preference relation that respects the order of schools induced by $P_i$ but considers $s \in S$ acceptable only when $sP_is_0$ and $|\{s' \in S : s'R_is\}| \leq \kappa$. Notice that, since $P_i^\kappa$ is obtained from $P_i$ by repositioning the outside option $s_0$, for all $s, s' \in S$, we have that $sP_i^\kappa s'$ if and only if $sP_is'$. Let $P^\kappa = (P_i^\kappa)_{i \in I}$. The truncation policy does not induce frictions when $\kappa \geq |S|$, because in this case $P = P^\kappa$ for all $P \in \mathcal{P}^n$.

A mechanism $\Phi : \mathcal{P}^n \to \mathcal{M}$ is **$\kappa$-stable** when $\Phi(P)$ is a stable matching of $(I, S, \succ, q, P^\kappa)$ for all $P \in \mathcal{P}^n$. When $\kappa \geq |S|$, we refer to a $\kappa$-stable mechanism as **stable mechanism**.

Let $\mathrm{DA}^\kappa : \mathcal{P}^n \to \mathcal{M}$ be the **student-optimal $\kappa$-stable mechanism** and let $\mathrm{DA}^{\mathrm{S},\kappa} : \mathcal{P}^n \to \mathcal{M}$ be the **school-optimal $\kappa$-stable mechanism**. That is, the matchings $\mathrm{DA}^\kappa(P)$ and $\mathrm{DA}^{\mathrm{S},\kappa}(P)$ are obtained by applying the deferred acceptance algorithm to $(I, S, \succ, q, P^\kappa)$ assuming that students and schools make proposals, respectively (cf., Gale and Shapley, 1962). When $\kappa \geq |S|$, we denote these mechanisms by DA and $\mathrm{DA}^{\mathrm{S}}$.

A **student $i$ can manipulate** a mechanism $\Phi$ at the preference profile $P$ if there exists $\hat{P}_i \in \mathcal{P}$ such that $\Phi_i(\hat{P}_i, P_{-i})P_i\Phi_i(P)$, where $P_{-i} = (P_j)_{j \in I \setminus \{i\}}$. A mechanism $\Phi$ is **strategy-proof** if no student can manipulate it at any preference profile.

A coalition is a non-empty set of students. A **coalition $C$ can manipulate** a mechanism $\Phi$ at the preference profile $P$ if there exists $\hat{P}_C = (\hat{P}_i)_{i \in C} \in \mathcal{P}^{|C|}$ such that $\Phi_i(\hat{P}_C, P_{-C})R_i\Phi_i(P)$ for every $i \in C$, and $\Phi_j(\hat{P}_C, P_{-C})P_j\Phi_j(P)$ for some $j \in C$, where $P_{-C} = (P_i)_{i \in I \setminus C}$. A mechanism $\Phi$ is **group strategy-proof** if no coalition can manipulate it at any preference profile.

It is well known that no stable mechanism is group strategy-proof in all school choice contexts. In fact, Alcalde and Barberà (1996, Theorem 3) and Ergin (2002, Theorem 1) ensure that the following properties hold for each $(I, S, \succ, q)$:

- A mechanism $\Phi : \mathcal{P}^n \to \mathcal{M}$ is stable and strategy-proof if and only if $\Phi = \mathrm{DA}$.
- DA is group strategy-proof if and only if $(\succ, q)$ is Ergin-acyclic.

Motivated by the incompatibility between stability and group strategy-proofness, we aim to compare mechanisms by their vulnerability to manipulation by coalitions.

## 4. Measures of coalitional manipulability

In this section, we describe a dozen criteria to compare mechanisms based on their vulnerability to coalitional manipulations. As we pointed out earlier, most of the criteria considered in this paper are extensions to coalitions of those previously proposed in the literature for individual manipulation (see Tables 1 and 2).

Given a mechanism $\Phi : \mathcal{P}^n \to \mathcal{M}$ and a preference profile $P$, an assignment $(s_i)_{i\in C}$ is an **attainable improvement** for the coalition $C$ under $\Phi$ at $P$ when the members of $C$ can obtain $(s_i)_{i\in C}$ by manipulating $\Phi$ at $P$. In other words, there exists $\hat{P}_C = (\hat{P}_i)_{i\in C} \in \mathcal{P}^{|C|}$ such that $s_i = \Phi_i(\hat{P}_C, P_{-C}) R_i \Phi_i(P)$ for all $i \in C$, and $s_j P_j \Phi_j(P)$ for some $j \in C$. Let $\mathcal{A}_\Phi(C,P)$ be the set of attainable improvements for $C$ under $\Phi$ at the preference profile $P$.

Denote by $\mathcal{A}^+_\Phi(C,P_C)$ the family of attainable improvements for $C$ under $\Phi$ at preference profiles of the form $(P_C, P'_{-C})$, and by $\mathcal{A}^+_\Phi(C)$ the collection of attainable improvements for $C$ under $\Phi$ at the different preference profiles. More formally,

$$\mathcal{A}^+_\Phi(C,P_C) \equiv \bigcup_{P'\in\mathcal{P}^n : P'_C = P_C} \mathcal{A}_\Phi(C,P'), \qquad \mathcal{A}^+_\Phi(C) \equiv \bigcup_{P\in\mathcal{P}^n} \mathcal{A}_\Phi(C,P).$$

Let $\mathcal{C}_\Phi(P)$ be the coalitions that can manipulate $\Phi$ at $P$. That is, $C \in \mathcal{C}_\Phi(P)$ if and only if $\mathcal{A}_\Phi(C,P) \neq \emptyset$.

For each coalition $C$ and $P_C \in \mathcal{P}^{|C|}$, denote by $\mathcal{P}_\Phi(C,P_C)$ the collection of preferences profiles at which $C$ can manipulate $\Phi$ when its members have preferences given by $P_C$. That is, $\mathcal{P}_\Phi(C,P_C) = \{P' \in \mathcal{P}^n : C \in \mathcal{C}_\Phi(P'),\ P'_C = P_C\}$. Thus, it follows that $\mathcal{P}_\Phi(C,P_C) \neq \emptyset$ if and only if $\mathcal{A}^+_\Phi(C,P_C) \neq \emptyset$.

Consider the sets

$$\mathcal{P}^+_\Phi(C) = \bigcup_{P_C\in\mathcal{P}^{|C|}} \mathcal{P}_\Phi(C,P_C), \qquad \mathcal{P}^+_\Phi = \bigcup_{C\subseteq I,\, C\neq\emptyset} \mathcal{P}^+_\Phi(C).$$

The elements of $\mathcal{P}^+_\Phi(C)$ are the preference profiles at which $C$ can manipulate $\Phi$, while $\mathcal{P}^+_\Phi$ are the preference profiles at which some coalition can manipulate $\Phi$. Hence, $\mathcal{P}^+_\Phi(C) \neq \emptyset$ if and only if $\mathcal{A}^+_\Phi(C) \neq \emptyset$, and $P \in \mathcal{P}^+_\Phi$ if and only if $\mathcal{C}_\Phi(P) \neq \emptyset$.

We will consider a family of criteria for comparing mechanisms regarding their coalitional manipulability. In the following tables, $\mathbf{\Psi \geqslant_d \Phi}$ denotes that **$\Psi$ is as coalitionally manipulable as $\Phi$** under a criterion $\geqslant_{\rm d}$.

| **Symbol** | **Meaning** |
|---|---|
| $\Psi \geqslant_{\rm CEPY} \Phi$ | For each coalition $C$ and $P \in \mathcal{P}^n$, $\mathcal{A}_\Phi(C,P) \subseteq \mathcal{A}_\Psi(C,P)$. |
| $\Psi \geqslant_{\rm NRR} \Phi$ | For each coalition $C$ and $P_C \in \mathcal{P}^{\lvert C\rvert}$, $\mathcal{A}^+_\Phi(C,P_C) \subseteq \mathcal{A}^+_\Psi(C,P_C)$. |
| $\Psi \geqslant_{\rm D1} \Phi$ | For each coalition $C$ and $P_C \in \mathcal{P}^{\lvert C\rvert}$, $\lvert\mathcal{P}_\Phi(C,P_C)\rvert \leq \lvert\mathcal{P}_\Psi(C,P_C)\rvert$. |
| $\Psi \geqslant_{\rm AM} \Phi$ | For each coalition $C$ and $P_C \in \mathcal{P}^{\lvert C\rvert}$, $\mathcal{A}^+_\Psi(C,P_C) = \emptyset \Rightarrow \mathcal{A}^+_\Phi(C,P_C) = \emptyset$.[7] |

TABLE 3. Criteria for comparing manipulability at the coalition and preference levels.

| **Symbol** | **Meaning** |
|---|---|
| $\Psi \geqslant^*_{\rm BN} \Phi$ | For every coalition $C$, $\mathcal{A}^+_\Phi(C) \subseteq \mathcal{A}^+_\Psi(C)$. |
| $\Psi \geqslant_{\rm D2} \Phi$ | For every coalition $C$, $\lvert\mathcal{P}^+_\Phi(C)\rvert \leq \lvert\mathcal{P}^+_\Psi(C)\rvert$. |
| $\Psi \geqslant_{\rm D3} \Phi$ | For every coalition $C$, $\mathcal{A}^+_\Phi(C) \neq \emptyset \Rightarrow \mathcal{A}^+_\Psi(C) \neq \emptyset$.[8] |

TABLE 4. Criteria for comparing manipulability at the coalition level.

[7]Equivalently, for each coalition $C$ and $P_C \in \mathcal{P}^{\lvert C\rvert}$, $\mathcal{P}_\Phi(C,P_C) \neq \emptyset \Rightarrow \mathcal{P}_\Psi(C,P_C) \neq \emptyset$.
[8]It holds that $\Psi \geqslant_{\rm D3} \Phi$ if and only if, for each coalition $C$, $\mathcal{P}^+_\Phi(C) \neq \emptyset \Rightarrow \mathcal{P}^+_\Psi(C) \neq \emptyset$.
[9]The property $\Psi \geqslant^*_{\rm PS} \Phi$ is equivalent to any of the following conditions:
(i) For each coalition $C$ and $P \in \mathcal{P}^n$, $\mathcal{A}_\Phi(C,P) \neq \emptyset \Rightarrow \mathcal{A}_\Psi(C,P) \neq \emptyset$.
(ii) For each coalition $C$, $\mathcal{P}^+_\Phi(C) \subseteq \mathcal{P}^+_\Psi(C)$.
(iii) For each coalition $C$ and $P_C \in \mathcal{P}^{\lvert C\rvert}$, $\mathcal{P}_\Phi(C,P_C) \subseteq \mathcal{P}_\Psi(C,P_C)$.

[10]Notice that, $\Psi \geqslant_{\rm PS} \Phi$ is equivalent to require that $\mathcal{P}^+_\Phi \subseteq \mathcal{P}^+_\Psi$.

| Symbol | Meaning |
|---|---|
| $\Psi \geqslant^{*}_{\text{PS}} \Phi$ | For all $P \in \mathcal{P}^n,\ \mathcal{C}_\Phi(P) \subseteq \mathcal{C}_\Psi(P)$.[9] |
| $\Psi \geqslant_{\text{BN}} \Phi$ | For all $P \in \mathcal{P}^n,\ \lvert\mathcal{C}_\Phi(P)\rvert \leq \lvert\mathcal{C}_\Psi(P)\rvert$. |
| $\Psi \geqslant_{\text{PS}} \Phi$ | For all $P \in \mathcal{P}^n,\ \mathcal{C}_\Phi(P) \neq \emptyset \Rightarrow \mathcal{C}_\Psi(P) \neq \emptyset$.[10] |

Table 5. Criteria for comparing manipulability at the preference level.

| Symbol | Meaning |
|---|---|
| $\Psi \geqslant_{\text{D4}} \Phi$ | $\sum_{P\in\mathcal{P}^n} \lvert\mathcal{C}_\Phi(P)\rvert \leq \sum_{P\in\mathcal{P}^n} \lvert\mathcal{C}_\Psi(P)\rvert$.[11] |
| $\Psi \geqslant_{\text{D5}} \Phi$ | $\lvert\mathcal{P}^+_\Phi\rvert \leq \lvert\mathcal{P}^+_\Psi\rvert$. |

Table 6. Criteria for comparing manipulability at a general level.

Let $\mathscr{S}^\kappa$ and $\mathscr{S}$ be the collections of $\kappa$-stable and stable mechanisms defined on $\mathcal{P}^n$, respectively. Given a school choice context $(I, S, \succ, q)$ and a $\kappa$-stable mechanism $\Phi : \mathcal{P}^n \to \mathcal{M}$, consider the following definitions:

- $\Phi$ **minimizes coalitional manipulability** on $\mathscr{S}^\kappa$ under the comparability criterion $\geqslant_{\text{d}}$ when $\Psi \geqslant_{\text{d}} \Phi$ for every $\Psi \in \mathscr{S}^\kappa$. Moreover, $\Phi$ is **optimal on $(\mathscr{S}^\kappa, \geqslant_{\text{d}})$** when it is the *unique* mechanism that minimizes coalitional manipulability on $\mathscr{S}^\kappa$ (i.e., $\Psi >_{\text{d}} \Phi$ for every $\Psi \in \mathscr{S}^\kappa$).
- $\Phi$ **maximizes coalitional manipulability** on $\mathscr{S}^\kappa$ under the comparability criterion $\geqslant_{\text{d}}$ when $\Phi \geqslant_{\text{d}} \Psi$ for every $\Psi \in \mathscr{S}^\kappa$. Moreover, $\Phi$ is **pessimal on $(\mathscr{S}^\kappa, \geqslant_{\text{d}})$** when it is the *unique* mechanism that maximizes coalitional manipulability on $\mathscr{S}^\kappa$ (i.e., $\Phi >_{\text{d}} \Psi$ for every $\Psi \in \mathscr{S}^\kappa$).

[11]This criterion seeks to measure coalitional manipulability by counting the group of students who can misrepresent their preferences. However, we consider each coalition $C$ as many times as there are preference profiles in $\mathcal{P}^+_\Phi(C)$. Notice that, $\sum_{P\in\mathcal{P}^n} \lvert\mathcal{C}_\Phi(P)\rvert = \sum_{C\subseteq I, C\neq\emptyset} \lvert\mathcal{P}^+_\Phi(C)\rvert$. In this way, we avoid considering a mechanism that is manipulable at all preference profiles by the same coalition as less manipulable than one that is manipulable only at one preference profile by two different coalitions.

Even though the order induced by $\geqslant_{\mathrm{d}}$ is not necessarily complete, minimizing/maximizing coalitional manipulability on $\mathscr{S}^{\kappa}$ requires being comparable to any other $\kappa$-stable mechanism.

## 5. Coalitional manipulability of $\kappa$-stable mechanisms

In this section, we study $\kappa$-stable mechanisms from the point of view of coalitional manipulability. Our first result shows that, under each comparability criterion, every mechanism $\Phi \in \mathscr{S}^{\kappa}$ is at least as coalitionally manipulable as $\mathrm{DA}^{\kappa}$ and at most as coalitionally manipulable as $\mathrm{DA}^{\mathrm{S},\kappa}$.

**Theorem 1.** *Given a school choice context $(I, S, \succ, q)$ and a criterion $\geqslant_{\mathrm{d}}$:*

- $\mathrm{DA}^{\kappa}$ *minimizes coalitional manipulability on $\mathscr{S}^{\kappa}$ under $\geqslant_{\mathrm{d}}$.*
- $\mathrm{DA}^{\mathrm{S},\kappa}$ *maximizes coalitional manipulability on $\mathscr{S}^{\kappa}$ under $\geqslant_{\mathrm{d}}$.*

*Proof.* Let $S_{\Phi}(C, P)$ denote the assignments that a coalition $C$ can achieve under a mechanism $\Phi$ when preferences are $P$. That is, $(s_i)_{i\in C} \in S_{\Phi}(C, P)$ if and only if there exists $\hat{P}_C \in \mathcal{P}^{|C|}$ such that $s_i = \Phi_i(\hat{P}_C, P_{-C})$ for all $i \in C$. The set $S_{\Phi}(C, P)$ contains not only the attainable *improvements* for $C$ at $P$ but every assignment that $C$ can achieve. In particular, $(\Phi_i(P))_{i\in C} \in S_{\Phi}(C, P)$.

Given $\kappa$-stable mechanisms $\Phi, \Psi : \mathcal{P}^n \to \mathcal{M}$, we have that:

(1a) *$S_{\Phi}(C, P) = S_{\Psi}(C, P)$ for every coalition $C$ and preference profile $P \in \mathcal{P}^n$.*

By symmetry, it is sufficient to show that $S_{\Phi}(C, P) \subseteq S_{\Psi}(C, P)$. Given $(s_i)_{i\in C}$ in $S_{\Phi}(C, P)$, there exists $\hat{P}_C$ such that $s_i = \Phi_i(\hat{P}_C, P_{-C}) \in S \cup \{s_0\}$ for each student $i \in C$. Let $\tilde{P}_C = (\tilde{P}_i)_{i\in C}$ be such that, for each $i \in C$, $s_i$ is the best alternative under $\tilde{P}_i$ and $s_0 \tilde{P}_i s$ for all $s \in S \setminus \{s_i\}$. Hence, when $s_i$ is a school, it is the only acceptable alternative under $\tilde{P}_i$. Since $\Phi(\hat{P}_C, P_{-C})$ is stable under $(\hat{P}^{\kappa}_C, P^{\kappa}_{-C})$ and $\tilde{P}^{\kappa}_i = \tilde{P}_i$ for all $i \in C$, $\Phi(\hat{P}_C, P_{-C})$ is also stable under $(\tilde{P}^{\kappa}_C, P^{\kappa}_{-C})$. This last property, the Rural Hospital Theorem (Roth, 1986), and the fact that $\Psi$ is $\kappa$-stable ensure that $\Psi_i(\tilde{P}_C, P_{-C}) = s_i$, for every $i \in C$. Therefore, $(s_i)_{i\in C} \in S_{\Psi}(C, P)$.

(1b) *Given $P \in \mathcal{P}^n$ such that $\Psi_i(P) R_i \Phi_i(P)$ for all $i \in I$, we have that $\mathcal{A}_\Psi(C,P) \subseteq \mathcal{A}_\Phi(C,P)$ for every coalition $C$. Therefore, $\Phi \geqslant_{\text{CEPY}} \Psi$ as long as $\Psi$ weakly Pareto dominates $\Phi$.*

Given $P \in \mathcal{P}^n$ and $(s_i)_{i \in C} \in \mathcal{A}_\Psi(C,P)$, it follows that $s_i R_i \Psi_i(P)$ for all $i \in C$, and $s_j P_j \Psi_j(P)$ for some $j \in C$. Moreover, since $(s_i)_{i\in C} \in S_\Psi(C,P)$, the property (1a) implies that $(s_i)_{i\in C} \in S_\Phi(C,P)$. That is, there is $\hat{P}_C \in \mathcal{P}^{|C|}$ such that $\Phi_i(\hat{P}_C, P_{-C}) = s_i$ for all $i \in C$. Hence, $\Phi_i(\hat{P}_C, P_{-C}) = s_i R_i \Psi_i(P) R_i \Phi_i(P)$ for all $i \in C$ and $\Phi_j(\hat{P}_C, P_{-C}) = s_j P_j \Psi_j(P) R_j \Phi_j(P)$ for some $j \in C$. Therefore, $(s_i)_{i\in C} \in \mathcal{A}_\Phi(C,P)$.

For every $P \in \mathcal{P}^n$ and $i \in I$, $\text{DA}^\kappa_i(P)$ is the best assignment that $i$ can achieve in a stable matching of $(I, S, \succ, q, P^\kappa)$. That is, $\text{DA}^\kappa_i(P) R^k_i \Phi_i(P)$ for every $\Phi \in \mathscr{S}^\kappa$. By definition, $s R^\kappa_i s'$ implies that $s R_i s'$, for all $s, s' \in S$. Since $\text{DA}^\kappa_i(P) = s_0$ implies that $\Phi_i(P) = s_0$ for all $\Phi \in \mathscr{S}^\kappa$, property (1b) ensures that $\text{DA}^\kappa$ minimizes coalitional manipulability on $\mathscr{S}^\kappa$ under $\geqslant_{\text{CEPY}}$, which implies that it also minimizes manipulability under any other criterion (see Figure 1).

Analogously, since $\text{DA}^{\text{S},\kappa}(P)$ determines the worst assignments that students can achieve in a stable matching of $(I, S, \succ, q, P^\kappa)$, we have that $\Phi_i(P) R^k_i \text{DA}^{\text{S},\kappa}_i(P)$ for each $i \in I$, $P \in \mathcal{P}^n$, and $\Phi \in \mathscr{S}^\kappa$. Therefore, combining the implication that $s R^\kappa_i s' \implies s R_i s'$ with the Rural Hospital Theorem and property (1b) ensure that $\text{DA}^{\text{S},\kappa}$ maximizes coalitional manipulability on the set of $\kappa$-stable mechanisms under every comparability criterion. □

Theorem 1 still holds when the coalitions that can be formed to manipulate are limited. In particular, when coalitions are restricted to $\{\{i\} : i \in I\}$, it follows that:

> *Under every comparability criterion, each $\kappa$-stable mechanism is at least as* ***individually manipulable*** *as* $\text{DA}^\kappa$ *and at most as individually manipulable as* $\text{DA}^{\text{S},\kappa}$.

The related literature has already discussed this assertion. For stable mechanisms, it follows from the strategy-proofness of DA, the fact that $\text{DA}^{\text{S}}$ is the student-pessimal stable mechanism, and Chen, Egesdal, Pycia, and Yenmez (2016, Theorem 1). For the comparability criterion $\geqslant^*_{\mathrm{PS}}$, Pathak and Sönmez (2013, Theorem 2) show that $\mathrm{DA}^{\mathrm{S}}$ is as individually manipulable as any other stable mechanism, while Bonkoungou and Nesterov (2023) ensure that $\mathrm{DA}^{\kappa}$ is at most as manipulable by students as any other $\kappa$-stable mechanism.

**Remark 1** *(On coalitional strategies to manipulate a $\kappa$-stable mechanism)*

When a $\kappa$-stable mechanism is implemented, a coalition can achieve an attainable improvement simply by misrepresenting some schools as unacceptable. More precisely, given $\Phi \in \mathscr{S}^{\kappa}$ and $(s_i)_{i\in C}$ in $\mathcal{A}_{\Phi}(C,P)$, for each $i \in C$ denote by $\tilde{P}_i$ the preference relation obtained from $P_i$ by declaring all schools in $S \setminus \{s_i\}$ unacceptable. Then, the $\kappa$-stability of $\Phi$ and the Rural Hospital Theorem guarantee that $s_i = \Phi_i(\tilde{P}_C, P_{-C})$ for all $i \in C$, where $\tilde{P}_C = (\tilde{P}_i)_{i\in C}$. This result follows from the same arguments in the proof of Theorem 1—property (1a)—by taking $\Psi = \Phi$. □

Next, we prove that, under the most demanding comparability criterion, $\mathrm{DA}^{\kappa}$ is less coalitionally manipulable than any other $\kappa$-stable mechanism, while $\mathrm{DA}^{\mathrm{S},\kappa}$ is more coalitionally manipulable than any other $\kappa$-stable mechanism.

**Theorem 2.** *Given a school choice context $(I, S, \succ, q)$, we have that:*

- $\mathrm{DA}^{\kappa}$ *is optimal on* $(\mathscr{S}^{\kappa}, \geqslant_{\mathrm{CEPY}})$.
- $\mathrm{DA}^{\mathrm{S},\kappa}$ *is pessimal on* $(\mathscr{S}^{\kappa}, \geqslant_{\mathrm{CEPY}})$.

*Proof.* Given mechanisms $\Phi, \Psi \in \mathscr{S}^{\kappa}$, the following properties hold:

(2a) *If $\Phi_i(P) P_i \Psi_i(P)$ for some $i \in I$ and $P \in \mathcal{P}^n$, then $\Phi_i(P) = \Psi_i(\tilde{P}_i, P_{-i})$ for some $\tilde{P}_i \in \mathcal{P}$.*

Since $s^* \equiv \Phi_i(P) \in S_{\Phi}(\{i\}, P)$, and the property (1a) in the proof of Theorem 1 ensures that $S_{\Phi}(\{i\}, P) = S_{\Psi}(\{i\}, P)$, we have that $s^* = \Psi_i(\tilde{P}_i, P_{-i})$ for some $\tilde{P}_i \in \mathcal{P}$.

(2b) *Given $P \in \mathcal{P}^n$, if $\mathcal{A}_{\Psi}(C,P) \subseteq \mathcal{A}_{\Phi}(C,P)$ for every coalition $C$, then $\Psi_i(P) R_i \Phi_i(P)$ for all $i \in I$. Therefore, $\Phi \geqslant_{\mathrm{CEPY}} \Psi$ guarantees that $\Psi$ weakly Pareto dominates $\Phi$.*

Given $P \in \mathcal{P}^n$, suppose that $\mathcal{A}_\Psi(C,P) \subseteq \mathcal{A}_\Phi(C,P)$ for all coalition $C$ and $\Phi_i(P) P_i \Psi_i(P)$ for some $i \in I$. Let $s^* \equiv \Phi_i(P)$. The property (2a) implies that $s^* = \Psi_i(\tilde{P}_i, P_{-i}) P_i \Psi_i(P)$ for some $\tilde{P}_i \in \mathcal{P}$. Hence, $s^*$ belongs to $\mathcal{A}_\Psi(\{i\},P)$. Since $\mathcal{A}_\Psi(\{i\},P) \subseteq \mathcal{A}_\Phi(\{i\},P)$, we have that $s^* P_i \Phi_i(P)$. A contradiction because $s^* \equiv \Phi_i(P)$.

If $\mathrm{DA}^\kappa \geqslant_{\mathrm{CEPY}} \Psi$ for some $\Psi \in \mathscr{S}^\kappa$, the property (2b) ensures that $\Psi_i(P) R_i \mathrm{DA}^\kappa_i(P)$ for every $i \in I$ and $P \in \mathcal{P}^n$. Since $s R_i s'$ implies that $s R^\kappa_i s'$ for all $s, s' \in S$, and the Rural Hospital Theorem ensures that $\Psi_i(P) = s_0$ if and only if $\mathrm{DA}^\kappa_i(P) = s_0$, it follows that $\Psi_i(P) R^\kappa_i \mathrm{DA}^\kappa_i(P)$ for every $i \in I$ and $P \in \mathcal{P}^n$. As $\mathrm{DA}^\kappa$ is the student-optimal $\kappa$-stable mechanism, $\Psi = \mathrm{DA}^\kappa$. Since $\Phi \geqslant_{\mathrm{CEPY}} \mathrm{DA}^\kappa$ for all $\Phi \in \mathscr{S}^\kappa$ (see Theorem 1), it follows that $\mathrm{DA}^\kappa$ is less coalitionally manipulable than any other mechanism on $\mathscr{S}^\kappa$ under $\geqslant_{\mathrm{CEPY}}$. Analogous arguments ensure that $\mathrm{DA}^{\mathrm{S},\kappa}$ is the unique mechanism that maximizes coalitional manipulability on $\mathscr{S}^\kappa$ under $\geqslant_{\mathrm{CEPY}}$.□

To prove Theorem 2, we only require that $\{\{i\} : i \in I\}$ be admissible coalitions. Consequently, the arguments in the proof of Theorem 2 also guarantee that:

*The mechanisms* $\mathrm{DA}^\kappa$ *and* $\mathrm{DA}^{\mathrm{S},\kappa}$ *are, respectively, the least and most* ***individually manipulable*** *$\kappa$-stable mechanisms under* $\geqslant_{\mathrm{CEPY}}$.

For stable mechanisms, this conclusion follows from Chen, Egesdal, Pycia, and Yenmez (2016, Theorem 2) and the facts that DA is the only stable and strategy-proof mechanism defined on $\mathcal{P}^n$ and $\mathrm{DA}^{\mathrm{S}}$ is the student-pessimal stable mechanism.

The strategy-proofness of DA (cf., Dubins and Freedman, 1981; Roth, 1982) will allow us to show that—under several comparability criteria—$\mathrm{DA}^\kappa$ is the only mechanism that minimizes coalitional manipulability on $\mathscr{S}^\kappa$.[12]

**Theorem 3.** *Given a school choice context* $(I, S, \succ, q)$*, we have that:*

[12]Notice that, since the Rural Hospital Theorem ensures that $\mathrm{DA}^1(P) = \mathrm{DA}^{\mathrm{S},1}(P)$ for all $P \in \mathcal{P}^n$, $\mathrm{DA}^1$ is the only 1-stable mechanism. Therefore, $\mathrm{DA}^1$ is trivially optimal on $\mathscr{S}^1$ under every comparability criterion.

- $\mathrm{DA}^\kappa$ *is optimal on* $(\mathscr{S}^\kappa, \geqslant_{\mathrm{d}})$ *for every comparability criterion in the family* $\{\geqslant^*_{\mathrm{PS}}, \geqslant_{\mathrm{BN}}, \geqslant_{\mathrm{D1}}, \geqslant_{\mathrm{D2}}, \geqslant_{\mathrm{D4}}, \geqslant_{\mathrm{NRR}}\}$.
- DA *is optimal on* $(\mathscr{S}, \geqslant_{\mathrm{d}})$ *for every* $\geqslant_{\mathrm{d}}$ *other than* $\geqslant_{\mathrm{PS}}$ *and* $\geqslant_{\mathrm{D5}}$.

*Proof.* Given $\Phi \in \mathscr{S}^\kappa \setminus \{\mathrm{DA}^\kappa\}$, let $P \in \mathcal{P}^n$ be such that $\Phi(P) \neq \mathrm{DA}^\kappa(P)$. Since $\mathrm{DA}^\kappa(P)$ is the student-optimal stable matching of $(I, S, \succ, q, P^\kappa)$, we have that $\mathrm{DA}^\kappa_i(P) P^\kappa_i \Phi_i(P)$ for some $i \in I$. The Rural Hospital Theorem guarantee that $\mathrm{DA}^\kappa_i(P)$ and $\Phi_i(P)$ are schools, which ensures that $\tilde{s} \equiv \mathrm{DA}^\kappa_i(P) P_i \Phi_i(P)$. Moreover, property (2a) in the proof of Theorem 2 implies that $\tilde{s} = \Phi_i(\widetilde{P}_i, P_{-i}) P_i \Phi_i(P)$ for some $\tilde{P}_i \in \mathcal{P}$. Thus, $\{i\} \in \mathcal{C}_\Phi(P)$. On the other hand, the fact that $\mathrm{DA}^\kappa_i(P) \in S$ implies that, if $i$ can manipulate $\mathrm{DA}^\kappa$ at $P$, then she can manipulate DA at $P^\kappa$. Hence, the strategy-proofness of DA guarantees that $\{i\} \notin \mathcal{C}_{\mathrm{DA}^\kappa}(P)$.

These properties and $\Phi \geqslant^*_{\mathrm{PS}} \mathrm{DA}^\kappa$ (see Theorem 1) ensure that $\Phi >^*_{\mathrm{PS}} \mathrm{DA}^\kappa$. Since $>^*_{\mathrm{PS}}$ is more demanding than $>_{\mathrm{BN}}$, $>_{\mathrm{D1}}$, $>_{\mathrm{D2}}$, and $>_{\mathrm{D4}}$ (see Figure 1), $\Phi$ is more coalitionally manipulable than $\mathrm{DA}^\kappa$ under any of these comparability criteria. Moreover, $\Phi \geqslant_{\mathrm{NRR}} \mathrm{DA}^\kappa$ (see Theorem 1) and $\Phi >_{\mathrm{D1}} \mathrm{DA}^\kappa$ guarantee that $\Phi >_{\mathrm{NRR}} \mathrm{DA}^\kappa$ (recall that, if $\geqslant_{\mathrm{d}}$ is more demanding than $\geqslant_{\mathrm{d}'}$, then $\Phi \geqslant_{\mathrm{d}} \Psi$ and $\Phi >_{\mathrm{d}'} \Psi$ implies that $\Phi >_{\mathrm{d}} \Psi$). Therefore, as $\Phi$ is generic, we conclude that $\mathrm{DA}^\kappa$ is optimal on $\mathscr{S}^\kappa$ under $\geqslant^*_{\mathrm{PS}}$, $\geqslant_{\mathrm{BN}}$, $\geqslant_{\mathrm{D1}}$, $\geqslant_{\mathrm{D2}}$, $\geqslant_{\mathrm{D4}}$, and $\geqslant_{\mathrm{NRR}}$.

Given $\Phi \in \mathscr{S} \setminus \{\mathrm{DA}\}$, there are $P \in \mathcal{P}^n$ and $i \in I$ such that $\hat{s} \equiv \mathrm{DA}_i(P) P_i \Phi_i(P)$. Hence, property (2a) implies that $\hat{s} \in \mathcal{A}^+_\Phi(\{i\})$. Since the strategy-proofness of DA ensures that $\mathcal{A}^+_{\mathrm{DA}}(\{i\}) = \emptyset$, $\Phi \geqslant_{\mathrm{D3}} \mathrm{DA}$ (see Theorem 1) implies that $\Phi >_{\mathrm{D3}} \mathrm{DA}$. Hence, $\Phi >_{\mathrm{d}} \mathrm{DA}$ for every comparability criterion $\geqslant_{\mathrm{d}}$ other than $\geqslant_{\mathrm{BN}}$, $\geqslant_{\mathrm{PS}}$, $\geqslant_{\mathrm{D4}}$ or $\geqslant_{\mathrm{D5}}$. Moreover, as $>^*_{\mathrm{PS}}$ is more demanding than $>_{\mathrm{BN}}$ and $>_{\mathrm{D4}}$, we have that $\Phi >_{\mathrm{BN}} \mathrm{DA}$ and $\Phi >_{\mathrm{D4}} \mathrm{DA}$. Therefore, as $\Phi$ is generic, we conclude that DA is optimal on $\mathscr{S}$ under each criterion other than $\geqslant_{\mathrm{PS}}$ and $\geqslant_{\mathrm{D5}}$. □

Our results remain true when preferences are truncated in a *non-homogeneous* way. Indeed, if student $i$ can report at most $\kappa_i \geq 1$ schools as acceptable, we have that:

- For all $s, s' \in S$, $s P^{\kappa_i}_i s'$ if and only if $s P_i s'$.

- If $s \in S \cup \{s_0\}$ is the unique acceptable alternative under $P_i$, then $P_i^{\kappa_i} = P_i$.

Hence, properties of the truncation policy required in Theorems 1, 2, and 3 hold.

## 6. Comparing $\mathrm{DA}^1$ with DA in terms of coalitional manipulability

Unlike what happens for individual manipulability, a truncation of preferences does not necessarily lead to an increment of coalitional manipulability. More precisely, the 1-stable mechanism $\mathrm{DA}^1$ can be incomparable with DA when coalitional manipulability is considered.

**Proposition 1.** *There are school choice contexts in which* $\mathrm{DA}^1$ *and* DA *cannot be compared in terms of their coalitional manipulability under a criterion different from* $\geqslant_{\mathrm{PS}}$*,* $\geqslant_{\mathrm{D4}}$*, and* $\geqslant_{\mathrm{D5}}$*.*

*Proof.* Let $I = \{1, 2, 3\}$, $S = \{s_1, s_2\}$, $\succ_{s_1}$: $1, 2, 3$, $\succ_{s_2}$: $3, 1, 2$, and $q_{s_1} = q_{s_2} = 1$. If $P = (P_1, P_2, P_3)$ is given by $P_1 : s_2, s_1, s_0$, $P_2 : s_1, s_0, s_2$, and $P_3 : s_1, s_2, s_0$, then

$$\mathrm{DA}(P) = [(1, s_1), (2, s_0), (3, s_2)], \qquad \mathrm{DA}^1(P) = [(1, s_2), (2, s_1), (3, s_0)].$$

Despite $\mathrm{DA}_2(P) \neq s_1$, student 2 prevents school $s_1$ from accepting student 3. Hence, every coalition $C \in \{\{1, 2\}, \{2, 3\}, \{1, 2, 3\}\}$ can manipulate DA by reporting $\tilde{P}_C = (\tilde{P}_2, (P_i)_{i \in C \setminus \{2\}})$ such that $\tilde{P}_2 : s_0, s_1, s_2$. Since DA is strategy-proof, it is not difficult to verify that these are the only coalitions that can manipulate DA at $P$. On the other hand, as 1 and 2 are assigned to their best schools in $\mathrm{DA}^1(P)$, and 3 has the best priority at $s_2$ and the worst priority at $s_1$, it follows that $\{3\}$ and $\{2, 3\}$ are the only coalitions that can manipulate $\mathrm{DA}^1$ at $P$ (any of these coalitions $C$ can manipulate by reporting $(\tilde{P}_3, (P_i)_{i \in C \setminus \{3\}})$ such that $\tilde{P}_3 : s_2, s_1, s_0$). We conclude that $|\mathcal{C}_{\mathrm{DA}^1}(P)| < |\mathcal{C}_{\mathrm{DA}}(P)|$. Moreover, $\mathcal{A}^+_{\mathrm{DA}}(\{1, 2, 3\}) \neq \emptyset$, and the scarcity of seats ensures that $\mathcal{A}^+_{\mathrm{DA}^1}(\{1, 2, 3\}) = \emptyset$.

If $\hat{P} = (\hat{P}_1, \hat{P}_2, \hat{P}_3)$ is given by $\hat{P}_1 : s_1, s_2, s_0$, $\hat{P}_2 : s_2, s_1, s_0$, and $\hat{P}_3 : s_1, s_2, s_0$, then

$$\mathrm{DA}(\hat{P}) = [(1, s_1), (2, s_0), (3, s_2)], \qquad \mathrm{DA}^1(\hat{P}) = [(1, s_1), (2, s_2), (3, s_0)].$$

Since $\mathrm{DA}(\hat{P})$ is Pareto efficient, Ergin (2002, Theorem 1) ensures that no coalition manipulates DA at $\hat{P}$. Moreover, the student 3 can manipulate $\mathrm{DA}^1$ at $\hat{P}$ by reporting $\tilde{P}_3 : s_2, s_1, s_0$. Thus, $|\mathcal{C}_{\mathrm{DA}^1}(\hat{P})| > |\mathcal{C}_{\mathrm{DA}}(\hat{P})|$. Moreover, $\mathcal{A}^+_{\mathrm{DA}^1}(\{3\}) \neq \emptyset$, and the strategy-proofness of DA guarantees that $\mathcal{A}^+_{\mathrm{DA}}(\{3\}) = \emptyset$.

Hence, $\mathrm{DA}^1$ and DA cannot be compared using $\geqslant_{\mathrm{BN}}$ or $\geqslant_{\mathrm{D3}}$. Therefore, these mechanisms are incomparable under all criteria different from $\geqslant_{\mathrm{PS}}$, $\geqslant_{\mathrm{D4}}$, and $\geqslant_{\mathrm{D5}}$ (see Figure 1). □

Proposition 1 does not hold for *individual manipulability*, because DA is strategy-proof (Dubins and Freedman, 1981; Roth, 1982) and $\mathrm{DA}^1$ is manipulable. Indeed, given $s_1, s_2 \in S$, if $q_{s_1} < n$ and $P_i : s_1, s_2, \dots$ for all $i \in I$, then each student left without a school in $\mathrm{DA}^1(P)$ has incentives to misrepresent her preferences by announcing $s_2$ as her preferred school.

By truncating preferences before implementing DA, a coalition may stop manipulating when one of its members ceases to be an *interrupter* (i.e., a student who prevents others from being assigned to a school that ultimately rejects her proposal (cf., Kesten, 2010)). [13] However, a student may begin manipulating to avoid exclusion from certain schools. The proof of Proposition 1 shows that—depending on the preference profile—the first effect can outweigh the second, or vice versa. For this reason, $\mathrm{DA}^1$ and DA are non-comparable by any criterion that takes into account the characteristics or the number of coalitions that can manipulate, or which schools can be attained after a manipulation.

This is not the case for the comparability criteria $\geqslant_{\mathrm{PS}}$ and $\geqslant_{\mathrm{D5}}$, which only take into account if mechanisms are coalitionally manipulable or not.

**Proposition 2.** $\mathrm{DA}^1$ *is as coalitionally manipulable as* DA *under* $\geqslant_{\mathrm{PS}}$ *and* $\geqslant_{\mathrm{D5}}$. *Moreover, if* $|S| \geq 2$ *and* $\min_{s\in S} q_s < n$, *then* $\mathrm{DA}^1 >_{\mathrm{PS}} \mathrm{DA}$ *and* $\mathrm{DA}^1 >_{\mathrm{D5}} \mathrm{DA}$.

*Proof.* Given $P \in \mathcal{P}^n$, let $I_{\otimes}(P)$ be the (possible empty) set of students who have *all* proposals immediately rejected when DA is implemented and preferences are given by

[13] For instance, in the proof of Proposition 1, student 2 is an interrupter of DA under $P$ and ceases to be one when $\mathrm{DA}^1$ is implemented.

$P$. That is, when the deferred acceptance algorithm underlying DA is implemented, a student in $I_{\otimes}(P)$ is not only assigned to $s_0$ but she is never temporarily accepted by a school. Notice that $I_{\otimes}(P) \subseteq \mathrm{DA}^1_{s_0}(P)$ and each $i \in \mathrm{DA}^1_{s_0}(P) \setminus I_{\otimes}(P)$ can manipulate $\mathrm{DA}^1$ by reporting as the unique acceptable alternative a school that (temporally or permanently) accepts her when DA is implemented.

Assume that $\mathcal{C}_{\mathrm{DA}^1}(P) = \emptyset$. It follows from the arguments above that $\mathrm{DA}^1_{s_0}(P) = I_{\otimes}(P)$, which implies that $\mathrm{DA}(P) = \mathrm{DA}^1(P)$. Therefore, $\mathcal{C}_{\mathrm{DA}}(P) = \emptyset$, because $\mathrm{DA}(P)$ is Pareto efficient (cf., Ergin, 2002). Since $P$ is arbitrary, we conclude that $\mathrm{DA}^1 \geqslant_{\mathrm{PS}} \mathrm{DA}$. This implies that $\mathrm{DA}^1 \geqslant_{\mathrm{D5}} \mathrm{DA}$.

Let $(I, S, \succ, q)$ be a school choice context such that $|S| \geq 2$ and $q_{s_1} < n$ for some $s_1 \in S$. Given $s_2 \in S \setminus \{s_1\}$, if $P \in \mathcal{P}^n$ satisfies $P_i : s_1, s_2, s_0, \ldots$ for all $i \in I$, then $\mathcal{C}_{\mathrm{DA}}(P) = \emptyset$, because $\mathrm{DA}(P)$ is Pareto efficient. However, each $i \in \mathrm{DA}^1_{s_0}(P)$ can manipulate $\mathrm{DA}^1$ at $P$ by reporting $s_2$ as her best alternative. Therefore, as $q_{s_1} < n$ implies that $\mathrm{DA}^1_{s_0}(P)$ is non-empty, it follows that $\mathcal{C}_{\mathrm{DA}}(P) = \emptyset$ and $\mathcal{C}_{\mathrm{DA}^1}(P) \neq \emptyset$. Since $\mathrm{DA}^1 \geqslant_{\mathrm{PS}} \mathrm{DA}$, we conclude that $\mathrm{DA}^1 >_{\mathrm{PS}} \mathrm{DA}$. Hence, $\mathrm{DA}^1 >_{\mathrm{D5}} \mathrm{DA}$. □

## 7. Coalitional manipulability versus Pareto dominance

For stable mechanisms, Chen, Egesdal, Pycia, and Yenmez (2016, Theorems 1 and 2) show that lower individual manipulability under $\geqslant_{\mathrm{CEPY}}$ is equivalent to Pareto dominance. Thus, given $\Phi, \Psi \in \mathscr{S}$ and a comparability criterion $\geqslant_{\mathrm{d}}$, to ensure that $\Psi$ is as individually manipulable as $\Phi$ under $\geqslant_{\mathrm{d}}$ it is sufficient that $\Phi$ weakly Pareto dominates $\Psi$ (see Figure 1).[14]

In our context, properties (1b) and (2b) (see the proofs of Theorems 1 and 2) guarantee that analogous results hold for coalitional manipulability and $\kappa$-stable mechanisms.

**Theorem 4.** *Given a school choice context $(I, S, \succ, q)$ and $\Phi, \Psi \in \mathscr{S}^{\kappa}$:*

- $\Psi \geqslant_{\mathrm{CEPY}} \Phi$ *if and only if* $\Phi$ *weakly Pareto dominates* $\Psi$.
- $\Psi >_{\mathrm{CEPY}} \Phi$ *if and only if* $\Phi$ *Pareto dominates* $\Psi$.

[14]Pathak and Sönmez (2013, Lemma 1) show that weak Pareto dominance of $\Phi$ over $\Psi$ ensures that $\Psi$ is as individually manipulable as $\Phi$ under $\geqslant^*_{\mathrm{PS}}$.

*In particular, under each comparability criterion, $\Psi$ is as coalitionally manipulable as $\Phi$ as long as all students consider $\Phi$ as good as $\Psi$.*

It is important to remark that this result is satisfied regardless of students' ability to coordinate among themselves, because properties (1b) and (2b) remain valid as long as $\{\{i\} : i \in I\}$ are included in the set of coalitions that can be formed to manipulate. Furthermore, since the concept of (weakly) Pareto dominance only requires comparing students' well-being at the individual level, given $\Phi, \Psi \in \mathscr{S}^\kappa$ and $P \in \mathcal{P}^n$, properties (1b) and (2b) ensure that, for each family of coalitions $\mathcal{C}$ such that $\{\{i\} : i \in I\} \subseteq \mathcal{C}$,

$$\mathcal{A}_\Phi(C, P) \subseteq \mathcal{A}_\Psi(C, P), \ \ \forall C \in \mathcal{C} \qquad \Longleftrightarrow \qquad \mathcal{A}_\Phi(\{i\}, P) \subseteq \mathcal{A}_\Psi(\{i\}, P), \ \ \forall i \in I.$$

Hence, to compare manipulability at the coalitional level using $\geqslant_{\text{CEPY}}$ it is sufficient to do it at the individual level. This result recalls the equivalence between *pairwise strategy-proofness* and group strategy-proofness shown by Alva (2017, Corollary 3).

## 8. School choice without outside options

In this section, we focus on school choice contexts in which students have no outside options. To do so, let $\mathcal{Q} \subseteq \mathcal{P}$ denote the collection of preference relations for which all schools are acceptable. Thus, preference profiles belong to $\mathcal{Q}^n$ and a coalition $C$ can manipulate a mechanism only by reporting preferences in $\mathcal{Q}^{|C|}$. Let $\mathscr{S}(\mathcal{Q}^n)$ be the set of stable mechanisms defined on $\mathcal{Q}^n$.

**Theorem 5.** *In a school choice context $(I, S, \succ, q)$ where $n < \sum_{s\in S} q_s$, we have that:*

- DA *is optimal on $(\mathscr{S}(\mathcal{Q}^n), \geqslant_{\text{d}})$ for each $\geqslant_{\text{d}}$ other than $\geqslant_{\text{PS}}$ and $\geqslant_{\text{D5}}$.*
- DA *minimizes coalitional manipulability on $\mathscr{S}(\mathcal{Q}^n)$ under each $\geqslant_{\text{d}}$.*
- DA$^{\text{S}}$ *is pessimal on $(\mathscr{S}(\mathcal{Q}^n), \geqslant_{\text{CEPY}})$.*
- DA$^{\text{S}}$ *maximizes coalitional manipulability on $\mathscr{S}(\mathcal{Q}^n)$ under each $\geqslant_{\text{d}}$.*

*Proof.* In our previous results, the fact that students can declare some schools as unacceptable is only used to ensure property (1a) (see the proof of Theorem 1).

Hence, it is sufficient to prove that this property still holds when the preference domain is $\mathcal{Q}^n$ and the number of available seats exceeds the number of students.

Given $\Phi, \Psi \in \mathscr{S}(\mathcal{Q}^n)$ and $P \in \mathcal{Q}^n$, we claim that $S_\Phi(C,P) = S_\Psi(C,P)$ for all coalition $C$. By symmetry, it is enough to show that $S_\Phi(C,P) \subseteq S_\Psi(C,P)$. Given $(s_i)_{i\in C} \in S_\Phi(C,P)$, there exists $\hat{P}_C \in \mathcal{Q}^{|C|}$ such that $s_i = \Phi_i(\hat{P}_C, P_{-C})$ for each $i \in C$. Since $n < \sum_{s\in S} q_s$, there is a school $s^*$ such that $|\Phi_{s^*}(\hat{P}_C, P_{-C})| < q_{s^*}$. The fact that $(\hat{P}_C, P_{-C}) \in \mathcal{Q}^n$ and the stability of $\Phi$ ensure that $s_i \in S$ for all $i \in C$. Let $\tilde{P}_C = (\tilde{P}_i)_{i\in C} \in \mathcal{Q}^{|C|}$ be such that, for each $i \in C$, the schools $s_i$ and $s^*$ are the best alternatives, in this order. Note that it is possible that $s_i = s^*$ for some $i \in C$. As $\Phi(\hat{P}_C, P_{-C})$ is stable under $(\hat{P}_C, P_{-C})$ and for each student $i \in C$ the best alternative under $\tilde{P}_i$ is $s_i$, $\Phi(\hat{P}_C, P_{-C})$ is also stable under $(\tilde{P}_C, P_{-C})$. This last property and the Rural Hospital Theorem ensure that $s^*$ does not fill its quota in $\Phi(\tilde{P}_C, P_{-C})$ and $\Phi_{s^*}(\hat{P}_C, P_{-C}) = \Phi_{s^*}(\tilde{P}_C, P_{-C})$. The stability of $\Psi$ and the Rural Hospital Theorem guarantee that $s^*$ does not fill its quota in $\Psi(\tilde{P}_C, P_{-C})$ and $\Psi_{s^*}(\tilde{P}_C, P_{-C}) = \Phi_{s^*}(\tilde{P}_C, P_{-C}) = \Phi_{s^*}(\hat{P}_C, P_{-C}) = \Psi_{s^*}(\hat{P}_C, P_{-C})$.

Consequently, for each student $i \in C$, the following properties hold:

- When $s_i \neq s^*$, we claim that $\Psi_i(\tilde{P}_C, P_{-C}) = s_i$. Indeed, if $\Psi_i(\tilde{P}_C, P_{-C}) \neq s_i$, the stability of $\Psi(\tilde{P}_C, P_{-C})$ and $|\Psi_{s^*}(\tilde{P}_C, P_{-C})| < q_{s^*}$ ensure that $\Psi_i(\tilde{P}_C, P_{-C}) = s^*$, which implies that $\Phi_i(\hat{P}_C, P_{-C}) = s^*$, a contradiction.
- When $s_i = s^*$, $\Psi_{s^*}(\tilde{P}_C, P_{-C}) = \Phi_{s^*}(\hat{P}_C, P_{-C})$ implies that $\Psi_i(\tilde{P}_C, P_{-C}) = s_i$.

Hence, $\Psi_i(\tilde{P}_C, P_{-C}) = s_i$, for every $i \in C$. Therefore, $(s_i)_{i\in C} \in S_\Psi(C,P)$. □

Analogous to what was shown in Theorem 4, when there are more seats available than students, within the stable mechanisms defined on $\mathcal{Q}^n$, $\Phi$ is less coalitionally manipulable than $\Psi$ under $\geqslant_{\text{CEPY}}$ if and only if $\Phi$ Pareto dominates $\Psi$. Moreover, weak Pareto dominance of $\Phi$ over $\Psi$ ensures that $\Psi$ is as coalitionally manipulable as $\Phi$ under each comparability criterion.

These results still hold when *coalitional* manipulability is replaced by *individual manipulability*, because the proof of Theorem 5 remains valid when $\{\{i\} : i \in I\}$ are the only coalitions that can be formed to manipulate. Therefore, although Chen, Egesdal, Pycia, and Yenmez (2016) require the preference domain to be *closed* to obtain the equivalence between Pareto dominance and less individual manipulability under $\geqslant_{\text{CEPY}}$, when $n < \sum_{s\in S} q_s$ this property holds in $\mathcal{Q}^n$, which is not closed.[15]

The next result shows that Theorem 5 does not hold when $n \geq \sum_{s\in S} q_s$.

**Proposition 3.** *There are school choice contexts* $(I, S, \succ, q)$ *with* $n \geq \sum_{s\in S} q_s$ *in which* DA *is coalitionally manipulable and* $\text{DA}^{\text{S}}$ *is group strategy-proof on* $\mathcal{Q}^n$*.*

*Proof.* Let $(I, S, \succ, q)$ be a school choice context satisfying the following properties:

- There are as many students as seats, $n \geq \sum_{s\in S} q_s$.
- If $\tau_s$ denotes the set of $q_s$-best students under $\succ_s$, we have that $\tau_s \cap \tau_{s'} = \emptyset$ for all $s, s' \in S$.
- $(\succ, q)$ has a cycle in the sense of Han and Zhang (2025, Definition 2).[16]

It follows from Han and Zhang (2025, Theorem 1) that DA is coalitionally manipulable on $\mathcal{Q}^n$. On the other hand, as schools have different best students and there are no outside options, $\text{DA}^{\text{S}}_s(P) = \tau_s$ for all $s \in S$ and $P \in \mathcal{Q}^n$. Notice that each student receives only one proposal when the school-optimal deferred acceptance mechanism is implemented. Since the absence of outside options is publicly known, $\text{DA}^{\text{S}}$ is group strategy-proof on $\mathcal{Q}^n$. □

Since $\text{DA}^{\text{S}}$ is the student-pessimal stable mechanism, Proposition 3 implies that the equivalence between Pareto dominance and less coalitional manipulability under $\geqslant_{\text{CEPY}}$ is not valid for stable mechanisms when there is a shortage of seats. Previously, focusing on individual manipulations, Sirguiado (2025, Theorem 3) showed this result for one-to-one two-sided matching markets.

[15] In a school choice context $(I, S, \succ, q)$, a preference domain $\tilde{\mathcal{D}}$ is *closed* when for all $P \in \tilde{\mathcal{D}}$ and for all stable matching $\mu$ of $(I, S, \succ, q, P)$, the preference relation $\tilde{P}_i$ that ranks the alternatives in $S \cup \{s_0\}$ in the same way as $P_i$ except that only $\mu(i)$ is acceptable satisfies $(\tilde{P}_i, P_{-i}) \in \tilde{\mathcal{D}}$.

[16] Example 2 in Han and Zhang (2025) describes a school choice context satisfying these three requirements.

## 9. Extensions and open questions

In our framework, schools have *responsive preferences*, as they use strict priorities to choose students up to their capacity (Roth and Sotomayor, 1990). However, our results can be extended to more general models.

To illustrate this possibility, assume that each school $s \in S$ has a complete, transitive, and strict preference relation $P_s$ defined on $2^I$, the family of subsets of $I$. This preference relation determines a *choice function* $\text{Ch}_s : 2^I \to 2^I$, which associates to every $I' \subseteq I$ the set of students preferred by $s$ when candidates are restricted to $I'$. Moreover, $P_s$ induces a *quota* $q_s$, which is the smallest positive integer such that $\emptyset P_s I'$ for all $I' \subseteq I$ with $|I'| > q_s$. Given preferences $(P_s)_{s\in S}$, a matching $\mu$ is *individually rational* at $P \in \mathcal{P}^n$ when both $\mu(i) R_i s_0$ for all $i \in I$ and $\text{Ch}_s(\mu^{-1}(s)) = \mu^{-1}(s)$ for all $s \in S$. Moreover, $\mu$ is *stable* at $P$ when it is individually rational and there is no $(i, s) \in I \times S$ such that $s P_i \mu(i)$ and $i \in \text{Ch}_s(\mu^{-1}(s) \cup \{i\})$.

Assume that each $P_s$ satisfies the following properties:

**Substitutability**: Given $I' \subseteq I'' \subseteq I$, if $i \in I' \cap \text{Ch}_s(I'')$, then $i \in \text{Ch}_s(I')$.
**Law of aggregate demand**: Given $I' \subseteq I'' \subseteq I$, $|\text{Ch}_s(I')| \leq |\text{Ch}_s(I'')|$.

In this context, the student-optimal stable mechanism and the school-optimal stable mechanism are well-defined (Roth, 1984; Hatfield and Milgrom, 2005). Moreover, the student-optimal stable mechanism is strategy-proof, and the weak version of the Rural Hospital Theorem holds: given a school choice problem, both the set of unassigned students and the *number* of students assigned to a school are equal in every stable matching (Hatfield and Milgrom, 2005, Theorems 8 and 11). Therefore, the arguments made in the proofs of Theorems 1, 2, 3, and 4 still hold when schools' preferences are substitutable and satisfy the law of aggregate demand.

Nevertheless, in the proof of Theorem 5 it is crucial that every school that does not fill its quota has the *same students* in every stable matching. Martínez, Massó, Neme, and Oviedo (2000, Corollary 3) and Klijn and Yazızı (2014, Theorem 2) provide sufficient conditions to ensure this strong version of the Rural Hospital Theorem. For instance, it suffices that schools' preferences are substitutable and satisfy:

**Weak separability**: Given $I' \subseteq I$ with $|I'| < q_s$ and $\mathrm{Ch}_s(I') = I'$, for each $i \in I \setminus I'$ it holds that $[\{i\}\, P_s\, \emptyset \;\Rightarrow\; (I' \cup \{i\})\, P_s\, I']$.

Therefore, Theorem 5 remains true when schools have substitutable and weakly separable preferences that satisfy the law of aggregate demand.

We conclude with some open questions that may be of interest for future research:

- Which of our results can be extended to a framework without strict preferences? Note that, since the Rural Hospital Theorem does not hold in a context with ties (Roth and Sotomayor, 1990), our techniques cannot be adapted.
- What is the effect on coalitional manipulability of a change in the number of alternatives that students can report? In other words, under what conditions $\mathrm{DA}^{\kappa}$ and $\mathrm{DA}^{\kappa+1}$ are comparable in terms of coalitional manipulability?
- In college admissions problems (Balinski and Sönmez, 1999), institutions can misrepresent preferences and capacities. Is it possible to compare stable mechanisms in terms of coalitional manipulability? What happens when groups composed of colleges and students have incentives to manipulate?
- In many-to-many matching markets (Echenique and Oviedo, 2006; Sönmez and Ünver, 2010),[17] group strategy-proofness and strategy-proofness coincide within the class of stable mechanisms (Romero-Medina and Triossi, 2020). However, strong assumptions are needed to ensure that stable and strategy-proof mechanism exists (Kojima, 2013). How do stable mechanisms compare in terms of coalitional manipulability in these markets?

## References

[1] Abdulkadiroğlu, Atila, Parag A. Pathak, and Alvin E. Roth (2009): "Strategy-proofness versus efficiency in matching with indifferences: redesigning the NYC high school match," *American Economic Review*, 99, 1954-1978.

[17]The general framework of Chen, Egesdal, Pycia, and Yenmez (2016) includes many-to-many matching markets. Hence, within the class of stable mechanisms, less individual manipulability under $\geqslant_{\mathrm{CEPY}}$ is equivalent to Pareto dominance even when students have multi-unit demand.

[2] Abdulkadiroğlu, Atila, and Tayfun Sönmez (2003): "School choice: a mechanism design approach," *American Economic Review*, 93, 729-747.

[3] Afacan, Mustafa O., and Umut Mert Dur (2017): "When preference misreporting is Harm[less]ful?," *Journal of Mathematical Economics*, 72, 16-24.

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

[5] Alva, Samson (2017): "When is manipulation all about the ones and twos?," working paper. Available at https://samsonalva.com/PWSP.pdf

[6] Arribillaga, R. Pablo, and Jordi Massó (2016): "Comparing generalized median voter schemes according to their manipulability," *Theoretical Economics*, 11, 547-586.

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

[8] Bonkoungou, Somouaoga, and Alexander Nesterov (2021): "Comparing school choice and college admissions mechanisms by their strategic accessibility," *Theoretical Economics*, 16, 881-909.

[9] Bonkoungou, Somouaoga, and Alexander Nesterov (2023): "Incentives in matching markets: counting and comparing manipulating agents," *Theoretical Economics*, 18, 965-991.

[10] Bonkoungou, Somouaoga, and Alexander Nesterov (2025): "When do reforms meet fairness concerns in school admissions?," *Social Choice and Welfare*, 65, 449-473.

[11] Cai, Zhengyang (2025): "The manipulability of deferred acceptance algorithm without outside options," *Economics Letters*, 254, 112485.

[12] Chen, Peter, Michael Egesdal, Marek Pycia, and M. Bumin Yenmez (2016): "Manipulability of stable mechanisms," *American Economic Journal: Microeconomics*, 8, 202-214.

[13] Chen, Yan, and Onur Kesten (2017): "Chinese college admissions and school choice reforms: a theoretical analysis," *Journal of Political Economy*, 125, 99-139.

[14] Decerf, Benoit, and Martin Van der Linden (2021): "Manipulability in school choice," *Journal of Economic Theory*, 197, 105313.

[15] Dubins, Lester E., and David A. Freedman (1981): "Machiavelli and the Gale-Shapley algorithm," *American Mathematical Monthly* 88, 485-494.

[16] Dur, Umut Mert (2019): "The modified Boston mechanism," *Mathematical Social Sciences*, 101, 31-40.

[17] Dur, Umut, Robert G. Hammond, and Thayer Morrill (2019): "The Secure Boston Mechanism: theory and experiments," *Experimental Economics*, 22, 918-953.

[18] Dur, Umut, Parag A. Pathak, Fei Song, and Tayfun Sönmez (2022): "Deduction dilemmas: the Taiwan assignment mechanism," *American Economic Journal: Microeconomics*, 14, 164-185

[19] Echenique, Federico, and Jorge Oviedo (2006): "A theory of stability in many-to-many matching markets," *Theoretical Economics*, 1, 233-273.

[20] Ergin, Haluk I. (2002): "Efficient resource allocation on the basis of priorities," *Econometrica*, 70, 2489-2497.

[21] Gale, David, and Lloyd Shapley (1962): "College admissions and the stability of marriage," *The American Mathematical Monthly* 69, 9-15.

[22] Haeringer, Guillaume, and Flip Klijn (2009): "Constrained school choice," *Journal of Economic Theory*, 144, 1921-1947.

[23] Han, Xiang, and Junxiao Zhang (2025): "Characterizing priorities for deferred acceptance with or without outside options," *Economic Theory*, 79, 497-517.

[24] Hatfield, John William, and Paul R. Milgrom (2005): "Matching with contracts," *American Economic Review*, 95, 913-935.

[25] Imamura, Kenzo, and Kentaro Tomoeda (2022): "Measuring manipulability of matching mechanisms," working paper. Available at SSRN: https://ssrn.com/abstract=4000419

[26] Kesten, Onur (2010): "School choice with consent," *The Quarterly Journal of Economics,* 125, 1297-1348.

[27] Klijn, Flip, and Ayşe Yazıcı (2014): "A many-to-many 'rural hospital theorem'," *Journal of Mathematical Economics*, 54, 63-73.

[28] Kojima, Fuhito (2013): "Efficient resource allocation under multi-unit demand," *Games and Economic Behavior*, 82, 1-14.

[29] Lomakin, Artemii, Kamil Minivaeb, and Alexander Nesterov (2024): "Modifications of Boston, Taiwanese and Chinese mechanisms are not comparable via counting manipulating students," *Economics Letters*, 237, 111647.

[30] 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.

[31] Neilson, Christopher (2024): "The rise of coordinated choice and assignment systems in education markets around the world," background paper prepared for World Development Report 2024, World Bank, Washington, DC.

[32] Nesterov, Alexander, Olga Rospuskova, and Sofia Rubtcova (2024): "Robustness to manipulations in school choice," *Social Choice and Welfare*, 62, 519-548.

[33] Pápai, Szilvia (2000): "Strategyproof assignment by hierarchical exchange," *Econometrica*, 68, 1403-1433.

[34] Pathak, Parag A. (2017): "What really matters in designing school choice mechanisms," in *Advances in Economics and Econometrics*, edited by B. Honoré, A. Pakes, M. Piazzesi, and L. Samuelson. Cambridge University Press, Chapter 6, 176-214.

[35] 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.

[36] Romero-Medina, Antonio, and Matteo Triossi (2020): "Strategy-proof and group strategy-proof stable mechanisms: An equivalence," *International Journal of Economic Theory*, 16, 349-354.

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

[38] Roth, Alvin E. (1984): "Stability and polarization of interests in job matching," *Econometrica*, 52, 47-58.

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

[40] Roth, Alvin E., and Marilda Sotomayor (1990): "Two-sided matching: A study in game-theoretic modeling and analysis," Cambridge University Press, Cambridge, U.K.

[41] Sönmez, Tayfun, and M. Utku Ünver (2010): "Course bidding at business schools," *International Economic Review*, 51, 99-123.

[42] Sirguiado, Camilo J. (2025) : "Minimizing manipulation in matching markets without outside options," working paper. Available at SSRN: https://ssrn.com/abstract=4911281

[43] Van der Linden, Martin (2019): "Deferred acceptance is minimally manipulable," *International Journal of Game Theory*, 48, 609-645.