Pareto dominance of deferred acceptance through early decision*

Somouaoga Bonkoungou†

August 19, 2016

An early decision market is governed by rules that allow each student to apply to (at most) one college and require the student to attend this college if admitted. This market is ubiquitous in college admissions in the United States. We model this market as an extensive-form game of perfect information and study a refinement of subgame perfect equilibrium (SPE) that induces undominated Nash equilibria in every subgame (SPUE). Our main result shows that this game can be used to define a decentralized matching mechanism that weakly Pareto dominates student-proposing deferred acceptance.

Keywords: Early decision, Pareto dominance, decentralized market, subgame perfect equilibrium, subgame perfect undominated Nash equilibrium, costly application.

JEL Classification: C78, C72, C73, I29.

1 Introduction

In early admissions programs, colleges admit students before the general application period. In the United States, these programs have become popular among colleges; and, a high percentage of freshmen are admitted to college each year through early admissions programs.

*I am very grateful to Lars Ehlers and Sean Horan for invaluable advice and support. I would also like to thank William Thomson, Vikram Munjunath, seminar participants at 2016 Canadian Economics Association meetings in Ottawa, at the 13th Meeting for the Society for Social Choice and Welfare in Lund for their helpful comments.

†Département de sciences économiques and CIREQ, Université de Montréal. Address: C.P.6128, succ. Centre-Ville, Montréal, QC, H3C 3J7, Canada. E-mail address: somouaoga.bonkoungou@umontreal.ca.

Early admission is an indication that the college admission market has failed.[1] The unraveling has taken different forms ranging from colleges that offer (non-binding) programs of early admission to ones that offer (binding) programs of early decision. In an *early decision* program, each applicant is required to (1) apply to exactly one college and (2) to commit to attend the college if admitted.[2,3] Due to this binding commitment to enrol, the debate on early admissions has largely focused on early decisions.

Formally, early decision can be viewed as a decentralized two-sided market that matches students to colleges. Decentralization is one way to organize a matching market by letting agents directly approach one another and propose partnerships. Matching theory is mainly concerned with the stability and the efficiency of the matchings that a market produces. A matching is said to be stable if it matches all agents to acceptable partners and no unmatched student-college prefers one another to their proposed partners. Data from organized matching markets have shown that many of the most successful matching markets produce stable matches; and, there is now a great deal of empirical evidence suggesting that stability played a critical role in the success of these markets.

The bulk of the literature on decentralized matching has focused on market mechanisms that implement stable matchings. However, stability is not a concern in early decision markets: the rules simply do not allow participants to seek new partners once the matching has been produced. This raises the question whether there exists an early decision market which produces matchings (according to an appropriate equilibrium concept) that Pareto dominate any stable matching from the students' perspective.

We show that, when students can coordinate, the matchings produced weakly Pareto dominate the outcome of the student-proposing deferred acceptance (DA) mechanism. We model the early decision market as an extensive form game with perfect information:

[1] Avery and Levin (2010) explained this unraveling by a signaling game where students signal their enthusiasms for colleges via early admissions.

[2] These informations are taken from Columbia undergraduate admissions website: https://undergrad.admissions.columbia.edu/apply/first-year/early-decision.

[3] The commitment is not legally enforceable. Some colleges enforce it by an assignment of agreement of enrolment. The early decision's agreement of Wake Forest University must be assigned by the applicant, a parent and his school counselor (http://static.wfu.edu/files/pdf/admissions/early.decision.pdf). Other colleges enforce the agreements via high school guidance counselors (Coles et al., 2011) by sharing information, refusing to admit students who applied more than one early decision.

first, students sequentially make at most one application and, then, colleges sequentially decide on which applicants to admit. Of course, an admitted student attends the college as reflected in the binding commitment.

While the appropriate equilibrium concept for a perfect information game is subgame perfect equilibrium (SPE), there are non-intuitive SPEs whose outcomes may be Pareto dominated by the outcome of the DA mechanism. They are equilibria where some students play strategies that are weakly dominated at some subgames. We propose a refinement that induces undominated Nash equilibria in every subgame (SPUE). We also consider costly applications and identify the SPEs of the corresponding game. We show that either of these solution concepts allow one to define a matching mechanism that weakly Pareto dominates the DA mechanism (Theorem 3).

Our analysis shows that information is important for coordination. In general, students prefer orders where they can observe other students' applications before they have to decide (Proposition 2). It is only for a particular subset of markets, where the DA mechanism is claims consistent, that this information is irrelevant. Arguably, unmodeled details of the market could introduce randomness into the order of applications. For this reason, we also consider an early decision matching game with random ordering. We show that every matching in the support of a random matching induced by an SPUE weakly Pareto dominates the outcome of the DA algorithm.

Currently, the DA mechanism is employed by many US cities (like New York and Boston) to assign students to public schools. While the DA mechanism has the advantage of being stable, it has been criticized for its lack of Pareto efficiency (Kesten, 2010). In general, Pareto efficiency is not compatible with stability (Roth, 1982; Ergin, 2002; Abdulkadiroǧlu et al., 2009). At the same time, there can be no dominant strategy incentive compatible mechanism (strategy-proof mechanisms) which Pareto dominates the DA mechanism (Kesten, 2010; Abdulkadiroǧlu et al., 2009; Alva and Munjunath, 2016). For this reason, the literature has recently begun to consider mechanisms that impose weaker incentive requirements. The main result of this paper can be viewed as a contribution to this literature.

Despite the interest in decentralized matching going back to Roth and Xing (1994), two-sided matching theory has devoted little attention to this issue. In part, this is likely due to the fact that, in practice, some of these markets have failed and centralized clearinghouses have emerged.[4] Nonetheless, a large number of matching markets are still decentralized — the early decision market being an important case in a point.

We discuss the relevant literature in Section 5 after we introduce our matching model in Section 2 and derive the results in Section 3 and Section 4. We collect all proofs in the appendices.

2 Model

2.1 Many-to-one matching market

Let S denote a finite set of students and C a finite set of colleges. Let $|S|:= n$, $C = \{c_1, ..., c_m\}$ and $S \cup C$ denote the set of agents with a generic agent denoted by v.[5] For each agent, remaining unmatched is denoted by $\emptyset$ (which also stands for the empty set).

Each student s has a strict preference relation P_s over $C \cup \{\emptyset\}$. Let $\mathcal{P}$ denote the set of preference relations and $\mathcal{P}^S$ the set of preference profiles $P = (P_s)_{s \in S}$ such that for each $s \in S$, $P_s \in \mathcal{P}$. Given $s \in S$ and $P_s \in \mathcal{P}$, let R_s denote the weak preference relation associated with P_s: for all $v, \widehat{v} \in C \cup \{\emptyset\}$, $v \; R_s \; \widehat{v}$ if $v \; P_s \; \widehat{v}$ or $v = \widehat{v}$. For $\widehat{S} \subsetneq S$, we often write $(P_{\widehat{S}}, P_{-\widehat{S}})$ instead of P.

Each college c has a maximum number $q_c \in \mathbb{N}$ of students it may admit, its quota. Let $q := (q_c)_{c \in C}$ denote the (profile of) quotas. Then, each college c has a strict preference relation $\succ_c$ over the set 2^S of all subsets of S. Given $c \in C$, $\succeq_c$ is the weak preference relation associated with $\succ_c$ and $Ch_c : 2^S \to 2^S$ is c's choice function induced by $\succ_c$ such that $Ch_c(\widehat{S}) := \max_{\succ_c} 2^{\widehat{S}}$ for each $\widehat{S} \in 2^S$. Note that Ch_c is well-defined because for each $\widehat{S} \in 2^S$, $Ch_c(\widehat{S})$ is uniquely defined since $\succ_c$ is strict. We assume that c's preference $\succ_c$ satisfies the followings: (1) for each $\widehat{S} \subseteq S$ with $|\widehat{S}| > q_c$, $\emptyset \succ_c \widehat{S}$ and (2) $\succ_c$ is *substitutable* and *acceptant*.

Susbtitutability and acceptability. *If Ch_c is c's choice function induced by $\succ_c$, (i) $\succ_c$ is substitutable if for each S' and each S'' such that $S' \subseteq S'' \subseteq S$, we have $Ch_c(S'') \cap S' \subseteq Ch_c(S')$; (ii) $\succ_c$ is acceptant if for each $S' \subseteq S$, $|Ch_c(S')| = \min\{q_c, |S'|\}$.*

Let $Ch = (Ch_c)_{c \in C}$ denote the profile of choice functions satisfying (1) and (2).

[4] Popular examples include entry-level medical markets in the US and Great Britain, and students placement to public schools in some US cities.

[5] We just fix the number of students here; we later identify them with their orders.

A *market* is a list (S, C, P, Ch, q). Since S, C and q mostly remain fixed throughout the paper, we suppress them and denote a market (S, C, P, Ch, q) by $M = (P, Ch)$.

A *matching* is a function $\mu : S \cup C \to 2^{S \cup C}$ such that (1) for each $s \in S$, $|\mu(s)| \leq 1$ and $\mu(s) \in C \cup \{\emptyset\}$, (2) for each $c \in C$, $\mu(c) \subseteq S$ and (3) for each $s \in S$ and each $c \in C$, $\mu(s) = \{c\}$ if and only if $s \in \mu(c)$. By convention, for each $v \in S \cup C$, we write μ_v instead of $\mu(v)$ and for each $s \in S$ and $c \in C$, $\mu_s = c$ instead of $\mu_s = \{c\}$. Let $\mathcal{M}$ denote the set of matchings. We extend agents' preferences over the appropriate sets to the set $\mathcal{M}$ naturally as follows: for each agent v and $\mu, \widehat{\mu} \in \mathcal{M}$, v prefers (or weakly prefers) μ to $\widehat{\mu}$ if and only he or it prefers (or weakly prefers) μ_v to $\widehat{\mu}_v$. We write $\mu R \widehat{\mu}$ if $\mu_s \, R_s \, \widehat{\mu}_s$ for each $s \in S$.

If $\mu_v = \emptyset$, we say that agent v is unmatched under μ. We say that a matching μ is *Individually Rational (IR) at* $M = (P, Ch)$ if for each $s \in S$, $\mu_s \, R_s \, \emptyset$ and for each $c \in C$, $Ch_c(\mu_c) = \mu_c$. Next, we say that μ is *blocked by the pair* $(s, c) \in S \times C$ *at* $M = (P, Ch)$ if $c \, P_s \, \mu_s$ and $s \in Ch_c(\mu_c \cup \{s\})$. Finally, we say that μ is *stable at* M if it is *IR* at M and it is not blocked by any pair at M.

A *(matching) mechanism* $\varphi : \mathcal{P}^S \to \mathcal{M}$ selects a matching $\varphi(P) \in \mathcal{M}$ for each $P \in \mathcal{P}^S$. Given $P \in \mathcal{P}^S$ and $\mu, \widehat{\mu} \in \mathcal{M}$, we say that μ *Pareto dominates* $\widehat{\mu}$ *(for students) at* P if $\mu_s R_s \widehat{\mu}_s$ for each $s \in S$ and for some $s \in S$, $\mu_s P_s \widehat{\mu}_s$. We say that μ *weakly Pareto dominates* $\widehat{\mu}$ *at* P if μ Pareto dominates $\widehat{\mu}$ at P or $\mu = \widehat{\mu}$. A *mechanism* φ *Pareto dominates a mechanism* $\widehat{\varphi}$ if for each $P \in \mathcal{P}^S$, $\varphi(P)$ weakly Pareto dominates $\widehat{\varphi}(P)$ and for some $P \in \mathcal{P}^S$, $\varphi(P)$ Pareto dominates $\widehat{\varphi}(P)$. Finally, a mechanism φ *weakly Pareto dominates a mechanism* $\widehat{\varphi}$ if φ Pareto dominates $\widehat{\varphi}$ or $\varphi = \widehat{\varphi}$.

2.2 Early decision matching game

Let $\pi : \{1, 2, ..., n\} \to S$ be a bijection and $\mathcal{O}$ the set of all such bijections. Each bijection describes the order in which students apply. Given an order π, we index the students such that $s_t = \pi(t)$ for $t = 1, 2, ..., n$. Since in each statement, there will be one order involved, this convention will not create a confusion. Given an order π the game form is as follows:

Applications phase: student s_1 either applies to one college or chooses to remain unmatched. Let $a_1 \in C \cup \{\emptyset\}$ denote s_1's decision such that $a_1 = c$ if s_1 applies to c and $a_1 = \emptyset$ if he chooses to remain unmatched.[6] For each $t = 2, ..., n$, student s_t observes

[6]Similarly, for each $t \geq 2$, let $a_t \in C \cup \{\emptyset\}$ denote s_t's decision, such that $a_t = c$ if s_t applies to c and

any application decisions $a_1, ..., a_{t-1}$ and makes an application $a_t \in C \cup \{\emptyset\}$.

Admissions phase: colleges offer admissions following the order of their index numbers. Given any application decisions $a_1, ..., a_n$, college c_1 offers admissions to a subset $o(c_1) \subseteq \{s_t \in S | a_t = c_1\}$ of its applicants. For each $j = 2, ..., m$, college c_j observes any decisions $a_1, ..., a_n, o(c_1), ..., o(c_{j-1})$ and offers admissions to a subset $o(c_j) \subseteq \{s_t | a_t = c_j\}$ of its applicants.

Each student's action consists of applying to a college or choosing to remain unmatched. Each college's action consists of admitting some students from among its applicants. A *history* is an ordered collection of actions (Osborne and Rubinstein, 1994). Let h^0 denote the empty history. Given an order π and $t = 1, ..., n$, the ordered collection $(h^0, a_1, ..., a_t) := h^t$ of actions is a history.[7] Let $\mathcal{H}_\pi^t$ denote the set of all such histories. The outcome at any terminal history $(h^0, a_1, ..., a_n, o(c_1), ..., o(c_m))$ is the matching μ such that for each $t = 1, ..., n$, (i) if $a_t = \emptyset$, then $\mu_{s_t} = \emptyset$ and (ii) if $a_t = c$, then $\mu_{s_t} = c$ if $s_t \in o(c)$ and $\mu_{s_t} = \emptyset$ otherwise. The game form just described is a well-defined *finite perfect-information extensive-form game*. Given an order π and a market M, let $G[\pi, M]$ denote the game induced by π and M.

Given an order π, a market M and $t = 1, ..., n$, a *strategy* σ_{s_t} *for student* s_t in $G[\pi, M]$ is a function $\sigma_{s_t} : \mathcal{H}_\pi^{t-1} \to C \cup \{\emptyset\}$ specifying an application decision $\sigma_{s_t}(h^{t-1}) \in C \cup \{\emptyset\}$ for each $h^{t-1} \in \mathcal{H}_\pi^{t-1}$.

It is straightforward to see that given a history $h^n = (h^0, a_1, ..., a_n)$, the optimal admission of each $c \in C$ is $o(c) = Ch_c(\{s_t \in S | a_t = c\})$. This is because a college does not get any further applications and therefore can choose only from $\{s_t \in S | a_t = c\}$. Henceforth, we ignore colleges' strategy profile and abuse language and speak of students' strategy profile as equilibrium of $G[\pi, M]$.

In any game $G[\pi, M]$, one can use Kuhn's (1953) backwards-induction algorithm to find a strategy profile that induces a Nash equilibrium in every subgame of $G[\pi, M]$. Such a strategy profile is a *subgame perfect (Nash) equilibrium (SPE)* of $G[\pi, M]$. Given a game $G[\pi, M]$, let $\mathcal{M}^{SPE}(G[\pi, M])$ denote its *SPE* outcomes.

$a_t = \emptyset$ if he chooses to remain unmatched.

[7] The empty history h^0 is considered as nature's action.

3 Results

We first describe the student-proposing deferred acceptance (DA) algorithm by Gale and Shpley (1962) whose outcome is one our main interests. Given a market $M = (P, Ch)$, it works as follows:

Step 1. Every student proposes to his most preferred acceptable college under P (if any). Let $\hat{S}_c^1$ be the set of students proposing to college c. College c tentatively accepts the students in $S_c^1 = Ch_c(\hat{S}_c^1)$ and rejects the proposers in $\hat{S}_c^1 \setminus S_c^1$.

Step t. ($t \geq 2$). Every student who was rejected at Step $(t-1)$ proposes to his next preferred acceptable college under P (if any). Let $\hat{S}_c^t$ be the set of students proposing to college c at this step. College c tentatively accepts the students in $N_c^t = Ch_c(S_c^{t-1} \cup \hat{S}_c^t)$ and rejects the proposers in $(S_c^{t-1} \cup \hat{S}_c^t) \setminus S_c^t$.

The algorithm terminates when each student is either accepted or rejected by all of his acceptable colleges. The tentative acceptance becomes final when the algorithm terminates and each student rejected by all of his acceptable colleges is unmatched. The mechanism that selects the outcome of the DA algorithm at $M = (P, Ch)$ for each P is the DA mechanism. Let φ^{Ch} denote the DA mechanism induced by Ch.[8]

Next, we show that the outcome of an SPE may be Pareto dominated by the outcome of a DA algorithm.

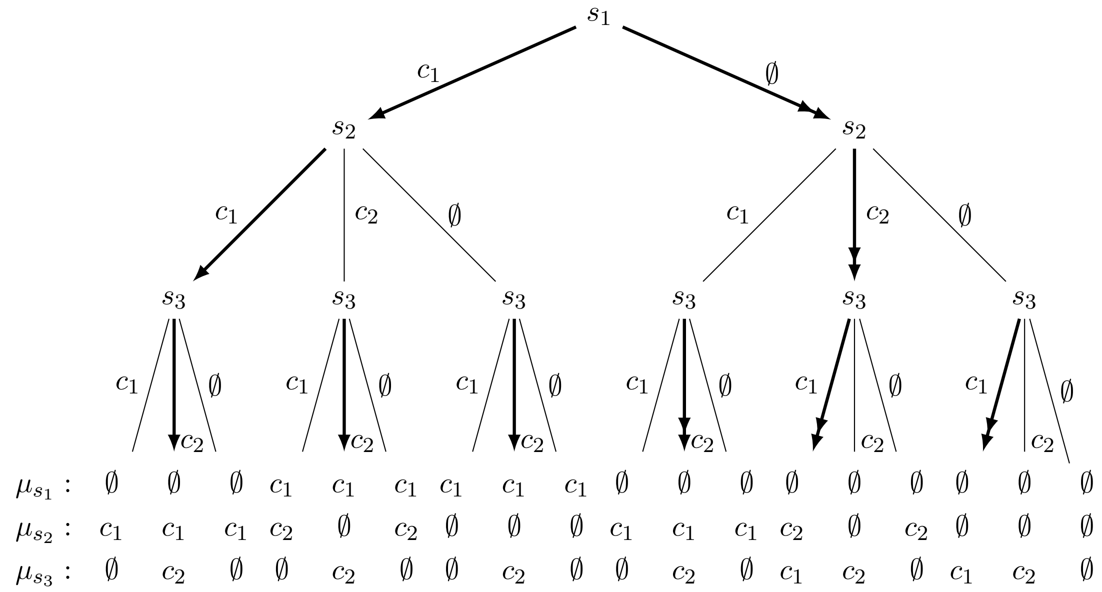
Example 1. ***An outcome of a DA algorithm Pareto dominating an SPE outcome.*** *Consider a market where* $S = \{s_1, s_2, s_3\}$ *and* $C = \{c_1, c_2\}$ *with quotas* $q_{c_1} = q_{c_2} = 1$ *and the preferences are as follows:*

P_{s_1}	P_{s_2}	P_{s_3}	$\succ_{c_1}$	$\succ_{c_2}$
$\emptyset$	c_2	c_1	s_2	s_1
	c_1	c_2	s_1	s_3
			s_3	s_2

Students move according to their index numbers. We represent only the relevant part of the tree. Specifically, we omit colleges' moves and the part where s_1 *applies to* c_2 *and the subsequent subgame since such application will be accepted and* c_2 *is not acceptable*

[8] This mechanism is due to Roth and Sotomayor (1990). However, the current version of the student-proposing DA algorithm is taken from Kojima and Manea (2010).

to s_1. We also represent the outcome in the terminal histories of the tree obtained when colleges choose optimally from among their applicants.



Applying Kuhn's algorithm identifies two SPEs represented by the connected arrows and double arrows. In both SPE outcomes, s_1 is unmatched. Clearly, the outcome of the first SPE, $\begin{pmatrix} s_1 & s_2 & s_3 \\ \emptyset & c_1 & c_2 \end{pmatrix}$, is Pareto dominated by the outcome of the DA algorithm, $\begin{pmatrix} s_1 & s_2 & s_3 \\ \emptyset & c_2 & c_1 \end{pmatrix}$.

As the example suggests, the underlying feature of the first SPE is that some students apply to unacceptable colleges; yet remain unmatched in the equilibrium. While such applications do not affect their welfare, it may induce subsequent students to apply to colleges that are less preferred to the ones where they would have applied otherwise.

Note that s_1's strategy of applying to the unacceptable college c_1 in the example is weakly dominated. Indeed, choosing to remain unmatched weakly dominates the strategy of applying to c_1: under c_1's strategy of admitting s_1, he strictly prefers choosing to remain unmatched to applying to c_1. The case of an SPE containing a weakly dominated strategy of Example 1 is representative of a more general issue: an SPE may contain strategies that are weakly dominated at some subgames, that is, their restrictions to some subgames may contain weakly dominated strategies. We deal with these non-intuitive strategies next.

3.1 Equilibrium characterization

3.1.1 Subgame perfect equilibrium with costly applications

One reasonable way of ruling out SPEs that are weakly dominated at some subgames is to assume that applications are costly. This way, no student is indifferent between the option of applying to a college and not being admitted, which we denote by $\emptyset_a$, and the option of not applying to any college and obviously remaining unmatched, denoted by $\emptyset_{na}$. We next make the necessary adjustments.

Each student s has now a strict preference P_s over $C \cup \{\emptyset_a, \emptyset_{na}\}$ such that $\emptyset_{na} \, P_s \, \emptyset_a$. Let $\widetilde{\mathcal{P}}$ denote the set of all such preferences. A college c is acceptable to s if $c \, P_s \, \emptyset_{na}$. Now a matching is a function $\mu : C \cup S \to 2^{C \cup S} \cup \{\emptyset_a, \emptyset_n\}$ such that (1) for each $s \in S$, $|\mu_s| \leq 1$, $\mu_s \in C \cup \{\emptyset_a, \emptyset_{na}\}$ and $\mu_s \neq \emptyset$, (2) for each $c \in C$, $\mu_c \subseteq S$ and (3) for each $s \in S$ and each $c \in C$, $\mu_s = c$ if and only if $s \in \mu_c$. Finally, the outcome attached to any terminal history $(h^0, a_1, ..., a_n, o(c_1), ..., o(c_m))$ is the matching μ such that for each $t = 1, ..., n$, (i) if $a_t = \emptyset_{na}$, then $\mu_{s_t} = \emptyset_{na}$ and (ii) if $a_t = c$ for some $c \in C$, then $\mu_{s_t} = c$ if $s_t \in o(c)$ and $\mu_s = \emptyset_a$ otherwise. With regard to the DA algorithm, its outcome matches each student who is rejected by all his acceptable colleges to $\emptyset_{na}$. Finally, given an order π and a market $M = (P, Ch)$ with $P \in \widetilde{\mathcal{P}}$, let $\widetilde{G}[\pi, M]$ denote the game induced by π and M when applications are costly. We now turn its SPEs.

In this setting, each game $\widetilde{G}[\pi, M]$ turns out to have a unique SPE, which we characterize. Given $s \in S$ and $c \in C$, let $P_s^c \in \widetilde{\mathcal{P}}$ denote s's preference relation in which c is his unique acceptable college and $P_s^{\emptyset_{na}} \in \widetilde{\mathcal{P}}$ his preference relation in which no college is acceptable.

Given an order π, a market $M = (P, Ch)$ with $P \in \widetilde{\mathcal{P}}^S$, define $\boldsymbol{P}(h^0) \equiv P$ and for each $t = 2, ..., n$ and each history $h^{t-1} = (h^0, a_1, ..., a_{t-1})$ of $\widetilde{G}[P, M]$, let $\boldsymbol{P}(h^{t-1}) \in \widetilde{\mathcal{P}}^S$ denote the following preference profile

$$\boldsymbol{P}(h^{t-1}) = (P_{s_1}^{a_1}, ..., P_{s_{t-1}}^{a_{t-1}}, P_{s_t}, ..., P_{s_n}).$$

It is obtained by simply replacing the preference relation $P_{s_{\widehat{t}}}$ of each student $s_{\widehat{t}}$ with $\widehat{t} < t$ and a decision $a_{\widehat{t}}$ according to h^{t-1} by $P_{s_{\widehat{t}}}^{a_{\widehat{t}}}$ and maintaining the preference relations of the remaining students as in P. Note that for each $\widehat{t} = t, ..., n$, $\boldsymbol{P}_{s_{\widehat{t}}}(h^{t-1}) = P_{s_{\widehat{t}}}$.

Then for each $t = 1, ..., n$, let $\sigma^*_{s_t} : \mathcal{H}^{t-1}_\pi \to C \cup \{\emptyset_{na}\}$ define the following strategy for s_t in $G[\pi, M]$. We dub it solidarity strategy.

Solidarity strategy. *Given an order π and a market $M = (P, Ch)$, s_t's strategy $\sigma^*_{s_t}$ in $\widetilde{G}[\pi, M]$ is a solidarity strategy if for each $h^{t-1} \in \mathcal{H}^{t-1}_\pi$,*

$$\sigma^*_{s_t}(h^{t-1}) = \varphi^{Ch}_{s_t}(\boldsymbol{P}(h^{t-1})).$$

At each history h^{t-1}, $\sigma^*_{s_t}$ recommends that (1) if $\varphi^{Ch}_{s_t}(\boldsymbol{P}(h^{t-1})) = c$ for some $c \in C$ then s_t applies to c and (2) if $\varphi^{Ch}_{s_t}(\boldsymbol{P}(h^{t-1})) = \emptyset_{na}$, then s_t chooses to remain unmatched. Let $\sigma^* = (\sigma^*_s)_{s \in S}$ denote the profile of solidarity strategies of $\widetilde{G}[\pi, M]$. Then we have,

Theorem 1. *Given an order π and a market M, the profile σ^* of solidarity strategies is the unique subgame perfect equilibrium of $\widetilde{G}[\pi, M]$.*

Appendix A contains the proof of Theorem 1.

3.1.2 Subgame perfect undominated Nash equilibria

Given an order π, a market M, a history $h \in \bigcup_{t=1}^{n-1} \mathcal{H}^t_\pi$ of $G[\pi, M]$, let $G[\pi, M|_h]$ denote the subgame of $G[\pi, M]$ that starts at h. Moreover, given a student s and his strategy σ_s, $\sigma_s|_h$ is the restriction of σ_s to $G[\pi, M|_h]$. Let $\sigma|_h := (\sigma_s|_h)_{s \in S}$. The matching attached to the terminal history reached when σ is executed starting at h is denoted by $\mu(\sigma|_h)$. Given an order π, a market $M = (P, Ch)$ and $t = 1, ..., n$, s_t's strategy σ_{s_t} *weakly dominates* his strategy $\widehat{\sigma}_{s_t}$ at the subgame $G[\pi, M|_{h^{t-1}}]$, if for each strategy profile σ_{-s_t}, he weakly prefers $\mu(\sigma|_{h^{t-1}})$ to $\mu(\widehat{\sigma}_{s_t}, \sigma_{-s_t}|_{h^{t-1}})$ under P_{s_t} with at least one strict preference, that is, $\mu_{s_t}(\sigma|_{h^{t-1}})\, R_{s_t}\, \mu_{s_t}(\widehat{\sigma}_{s_t}, \sigma_{-s_t}|_{h^{t-1}})$ and $\mu_{s_t}(\sigma_{s_t}, \widehat{\sigma}_{-s_t}|_{h^{t-1}}) P_{s_t} \mu_{s_t}(\widehat{\sigma}|_{h^{t-1}})$ for some $\widehat{\sigma}_{-s_t}$. A *strategy is undominated at some subgame* if it is not weakly dominated at that subgame.

Remark. *Given an order π, a market $M = (P, Ch)$, $t = 1, ..., n$ and a history $h^{t-1} \in \mathcal{H}^{t-1}_\pi$, s_t's strategy of applying to an unacceptable college c under P_{s_t} is is weakly dominated at the subgame $G[\pi, M|_{h^{t-1}}]$ by any strategy where he chooses to remain unmatched at h^{t-1}. Indeed, the application to c may lead to two outcomes for s_t. Either c admits him in which case he prefers choosing to remain unmatched to applying to c at h^{t-1} or c does not admit him, the same outcome as choosing to remain unmatched at h^{t-1}.*

Although weakly dominated strategies may be best responses, the literature on game theory has a long tradition of eliminating them. The dominance solvable concept

(Moulin, 1979) widely applied in game theory is based on the idea that successive elimination of weakly dominated strategies yields a unique strategy profile. Strategy-proofness of mechanisms is based on the idea that truthful reporting of preferences by every agent weakly dominates any untruthful reporting of preferences. In all these well-known applications, weakly dominated strategies are not expected. We follow this tradition and reject strategy profiles that contain weakly dominated strategies at some subgame.

Given an order π, $t = 1, ..., n$, a market M and a history $h^{t-1} \in \mathcal{H}_{\pi}^{t-1}$ in $G[\pi, M]$, s_t's strategy in the subgame of $G[\pi, M]$ that starts at h^{t-1} becomes a decision at h^{t-1}. Thus, in this subgame choosing to remain unmatched weakly dominates the strategy of applying to an unacceptable college. The elimination of strategies that are weakly dominated in some subgames yields the following refinement of SPE.

Subgame perfect undominated Nash equilibrium (SPUE). *Given an order π and a market M, a strategy profile in $G[\pi, M]$ is a subgame perfect undominated Nash equilibrium of $G[\pi, M]$ if it induces an undominated Nash equilibrium in every subgame of $G[\pi, M]$.*

Given an order π and a market M, let $\mathcal{M}^{SPUE}(G[\pi, M])$ denote the outcomes of the SPUEs of $G[\pi, M]$. This mild refinement of SPE produces outcomes that weakly Pareto dominate the outcome of the DA algorithm. Wasted applications to unacceptable colleges are ruled out in any SPUE. However, there are other wasted applications. Any application to an acceptable college which is not admitted is also wasteful. Such application does not affect the applicant's outcome of not being admitted in the SPUE but may affect others. We then dub such strategy a *bossy strategy.*[9]

Example 2 (continued)**. *A bossy strategy.*** *Consider Example 1 where student s_1's preference relation becomes $P_{s_1}^{c_1}$ in which c_1 is the unique acceptable college to him. Applying Kuhn's backwards-induction algorithm again yields the same equilibria identified in the tree of this example. The outcomes of the SPEs are now the outcome of the DA algorithm $\begin{pmatrix} s_1 & s_2 & s_3 \\ \emptyset & c_1 & c_2 \end{pmatrix}$ and the matching $\begin{pmatrix} s_1 & s_2 & s_3 \\ \emptyset & c_2 & c_1 \end{pmatrix}$. Student s_1 applied to c_1 in an SPE but is not admitted. The other students prefer the outcome of the SPE where he chose to remain unmatched to the one where he applied to c_1. It is a bossy strategy.*

[9]We thank William Thomson for making a parallel between *non-bossiness*, a social choice theory concept, and the property embodied in the strategy mentioned; which led to the choice of bossy strategy. A social choice function is bossy when some agent can change the outcome for others without changing his own.

Bossy strategy. *Given an order π, a market $M = (P, Ch)$ and $t = 1, ..., n$, s_t's strategy $\sigma^{**}_{s_t}$ in $G[\pi, M]$ is a bossy strategy if for each history $h^{t-1} \in \mathcal{H}^{t-1}_{\pi}$, letting $\varphi^{Ch}_{s_t}(\boldsymbol{P}(h^{t-1})) := v$*

(i) $\sigma^{**}_{s_t}(h^{t-1}) = v$ if $v = c$ for some $c \in C$ and

(ii) $\sigma^{**}_{s_t}(h^{t-1}) = \widehat{v}$ with $\widehat{v}\, R_{s_t}\, \emptyset$ if $v = \emptyset$ and $\widehat{v} \neq \emptyset$ for at least one history.

In a bossy strategy $\sigma^{**}_{s_t}$, (i) for each history h^{t-1} with $\varphi^{Ch}_{s_t}(\boldsymbol{P}(h^{t-1})) = c$ for some $c \in C$, s_t applies to c and (ii) for each history h^{t-1} with $\varphi^{Ch}_{s_t}(\boldsymbol{P}(h^{t-1})) = \emptyset$, s_t either chooses to remain unmatched or applies to an acceptable college ; with at least one application.

We can adapt the definition of solidarity strategy to the present setting as follows: given an order π and a market $M = (P, Ch)$, s_t's strategy $\sigma^*_{s_t}$ in $G[\pi, M]$ is a solidarity strategy if for each $h^{t-1} \in \mathcal{H}^{t-1}_{\pi}$,

$$\sigma^*_{s_t}(h^{t-1}) = \varphi^{Ch}_{s_t}(\boldsymbol{P}(h^{t-1})).$$

Theorem 2. *Given an order π and a market M, the subgame perfect undominated Nash equilibria of $G[\pi, M]$ correspond to the strategy profiles $\sigma = (\sigma_s)_{s \in S}$ in $G[\pi, M]$ such that for each student s, σ_s is either a solidarity strategy or a bossy strategy in $G[\pi, M]$.*

Appendix B contains the proof of Theorem 2.

3.2 Properties of the equilibrium matchings

Whether we consider costly applications or SPUE concept, an early decision matching mechanism recommends the outcome of an intuitive equilibrium for every preference profile.

Early decision matching mechanisms. *Given an order π and a choice profile Ch, the early decision matching mechanisms induced by Ch are:*

(i) the mechanism $\varphi(., Ch) : \widetilde{\mathcal{P}} \to \mathcal{M}$ such that for each $P \in \widetilde{\mathcal{P}}^S$; $\varphi(P, Ch) = \mathcal{M}^{SPE}(\widetilde{G}[\pi, M])$, where $M = (P, Ch)$ and

(ii) each mechanism $\varphi(., Ch) : \mathcal{P}^S \to \mathcal{M}$ such that for each $P \in \mathcal{P}^S$; $\varphi(P, Ch) \in \mathcal{M}^{SPUE}(G[\pi, M])$, where $M = (P, Ch)$.

Then, our main result is as follows:

Theorem 3. *Each early decision matching mechanism weakly Pareto dominates the DA mechanism.*

Appendix C contains the proof of Theorem 3.

Remark. *As a corollary of the theorem, the outcome of the DA mechanism is the unique stable matching that can arise at an intuitive equilibrium.*

Since the order matters for the SPE outcome under costly applications, it is important to understand how often this is the case. We later characterize the markets for which the outcome is order independent. But first we show that the the outcome of an SPE or SPUE is in certain sense order independent. A student who is matched at some order will remain matched for any other order. This holds for any unmatched student. This result resembles a feature of the set of stable matchings (Roth 1984, 1986),[10] known there as *rural hospital theorem.* A student who is matched at some stable matching remains matched in any other stable matching and this holds for any unmatched student. Furthermore, a college with unfilled seats at some stable matching is matched to the same set of students at each stable matching.

The rural hospital theorem owes its name from the following mal-distribution problem in the American medical labor market. It is empirically reported that hospitals in rural areas experience a shortage of doctors while hospitals in other areas succeed in filling their available positions with resident students. The problem persisted when the National Resident Matching Program centralized its matching procedure and used a stable matching mechanism. The insight from the rural hospital theorem suggests that no other stable matching mechanism can address this issue.

Rural hospital theorem. *A non-empty subset $\mathcal{M}' \subseteq \mathcal{M}$ of matchings satisfies the rural hospital theorem if for each $\mu, \mu' \in \mathcal{M}'$, (i) for each $v \in C \cup S$, $|\mu(v)| = |\mu'(v)|$ and (ii) for each $c \in C$, $|\mu(c)| < q_c$ implies $\mu(c) = \mu'(c)$.*

Given a market $M = (P, Ch)$ with $P \in \mathcal{P}^S$ and a market $M' = (P', Ch')$ with $P' \in \widetilde{\mathcal{P}}^S$, we consider the set $\bigcup_{\pi \in \mathcal{O}} \mathcal{M}^{SPE}(\widetilde{G}[\pi, M'])$ and the set $\bigcup_{\pi \in \mathcal{O}} \mathcal{M}^{SPUE}(G[\pi, M])$ obtained at various orders. These sets satisfy the rural hospital theorem.

[10] Hatfield and Milgrom (2005) partially obtained the rural hospital theorem on the set of stable matchings, when colleges have substitutable preferences coupled with another condition.

Proposition 1. *Given a market* $M = (P, Ch)$ *with* $P \in \mathcal{P}$ *and* $M' = (P', Ch)$ *with* $P' \in \widetilde{\mathcal{P}}$,

(i) the set $\bigcup_{\pi \in \mathcal{O}} \mathcal{M}^{SPE}(\widetilde{G}[\pi, M'])$ *satisfies the rural hospital theorem.*

(ii) the set $\bigcup_{\pi \in \mathcal{O}} \mathcal{M}^{SPUE}(G[\pi, M])$ *satisfies the rural hospital theorem.*

Appendix F contains the proof of Proposition 1. We note that Proposition 1 is not a negative result as in the medical labor market. It is a positive result in the sense that the order does not change a matched student to unmatched or an unmatched to matched. Unfortunately, when we insist on order independence in the sense that the order does not affect the outcome of an SPE, we obtain that only on a subset of markets.

The idea of order independence of a solution's outcomes is first modeled as an equilibrium concept by Moldovanu and Winter (1992). We rather regard it as a property of a market. We search for these markets only under costly applications for simplicity.

Order independence. *A market* M *induces an order independent* $\mathcal{G}$*-outcome if for all orders* π *and* π', $\mathcal{M}^{SPE}(\widetilde{G}[\pi, M]) = \mathcal{M}^{SPE}(\widetilde{G}[\pi', M])$.

Corollary 1. *If a market* M *induces an order independent* $\mathcal{G}$*-outcome, then for each order* π*, the subgame perfect equilibrium outcome of* $\widetilde{G}[\pi, M]$ *is the outcome of the DA algorithm.*

Appendix D contains the proof of Corollary 1. Given a mechanism φ and a market $M = (P, Ch)$, let $P^{\varphi} \in \mathcal{P}^S$ denote a profile such that for each student s, $\varphi_s(P)$ is s's unique acceptable college if any. Let $P^{\varphi}_{S'}$ denote the restriction of P^{φ} to S'. We study the robustness of a mechanism φ when some students, say S', deviate from their original preference profile $P_{S'}$ and claim each only his outcome at $\varphi(P)$. Then, *claims consistency* requires that whenever such deviation is observed then the mechanism still selects the same matching.

Claims consistency. [11] *A mechanism* φ *is claims consistent at* P *if for each subset* $S' \subseteq S$,

$$\varphi(P^{\varphi}_{S'}, P_{-S'}) = \varphi(P).$$

[11] Variants of consistency have been explored, ranging from consistency, bilateral consistency to average consistency (Thomson, 2011).

Theorem 4. *A market $M = (P, Ch)$ induces an order independent $\mathcal{G}$-outcome if and only if the DA mechanism φ^{Ch} is claims consistent at P.*

Appendix E contains the proof of Theorem 4. It is worth connecting our definition to the popular version of consistency. Fix a choice profile Ch. Given a market (S, P, Ch, q), a subset $\widehat{S} \subseteq S$ of students and quotas q' such that $q'_c \leq q_c$ for all $c \in C$, call the market $(\widehat{S}, P_{\widehat{S}}, Ch|_{2^{\widehat{S}}}, q')$ a sub-market of (S, P, Ch, q) where $Ch|_{2^{\widehat{S}}}$ is the restriction of Ch to $2^{\widehat{S}}$. Next, the *extended DA mechanism* $\widetilde{\varphi}^{Ch}$ selects the outcome of the DA algorithm at $(\widehat{S}, P_{\widehat{S}}, Ch|_{2^{\widehat{S}}}, q')$ for any $P \in \mathcal{P}^S$ and any sub-market $(\widehat{S}, P_{\widehat{S}}, Ch|_{2^{\widehat{S}}}, q')$ of (S, P, Ch, q).

Given a matching μ, consider the sub-market $(\widehat{S}, P_{\widehat{S}}, Ch|_{2^{\widehat{S}}}, q^{\mu})$ with respect to $\widehat{S}$ and μ that results from a departure of the students in $S \setminus \widehat{S}$ with their outcomes at μ; that is, $q^{\mu}_c = q_c - |\mu_c \setminus \widehat{S}|$ for each $c \in C$. Then, given a domain $\overline{\mathcal{P}}$ of preferences, we say that the DA mechanism is consistent on $\overline{\mathcal{P}}$ if for all profile $P \in \overline{\mathcal{P}}^S$ and all subset $\widehat{S} \subseteq S$, letting $\mu = \widetilde{\varphi}^{Ch}(S, P, Ch, q)$, we have

$$\mu|_{\widehat{S}} = \widetilde{\varphi}^{Ch}(\widehat{S}, P_{\widehat{S}}, Ch|_{2^{\widehat{S}}}, q^{\mu}).$$

The matching literature has focused on the strict preference domain ($\overline{\mathcal{P}} = \mathcal{P}$) and searched for choice profiles Ch for which the DA mechanisms φ^{Ch} are consistent on $\mathcal{P}$. The following acyclical condition provided by Ergin (2002) for responsive preferences and later generalized to substitutable preferences by Kumano (2009) is the answer. A cycle consists of three distinct students i, j, k and two distinct colleges $c, \widehat{c}$ such that there exist two disjoint subsets $S_c, S_{\widehat{c}} \subset S \setminus \{i, j, k\}$ verifying (C) $j \notin Ch_c(S_c \cup \{i, j\})$, $k \notin Ch_c(S_c \cup \{j, k\})$ and $i \notin Ch_{\widehat{c}}(S_{\widehat{c}} \cup \{i, k\})$ and (S) $|S_c| = q_c - 1$ and $|S_{\widehat{c}}| = q_{\widehat{c}} - 1$. A choice profile is *acyclical* if there is no cycle. Given a choice profile Ch, the DA mechanism φ^{Ch} is consistent on $\mathcal{P}$ if and only if Ch is acyclical (Klijn, 2011).

Given an acyclical choice profile Ch, for any profile $P \in \mathcal{P}^S$, φ^{Ch} is claiming consistent at P.[12] However, claiming consistency additionally characterizes markets (P, Ch) for which $P \in \mathcal{P}^S$ and Ch has a cycle. For a simple example, if Ch is a choice profile that has a cycle and P any profile then, φ^{Ch} is claiming consistent at P^{φ}.

Remark. *If a choice profile Ch is acyclical then, the mechanism $\varphi(., Ch) : \widetilde{\mathcal{P}} \to \mathcal{M}$*

[12]Since Ch is acyclical, for each $P \in \mathcal{P}^S$, $\varphi^{Ch}(P)$ is Pareto efficient (Kumano, 2009). Fix $\widehat{S} \subseteq S$ and let $\widehat{P} := (P^{\varphi}_{\widehat{S}}, P_{-\widehat{S}})$. It is proven (Kojima and Manea, 2010) using an (IR monotonicity) property of the DA mechanism that $\varphi^{Ch}(\widehat{P}) \, \widehat{R} \, \varphi^{Ch}(P)$. Thus, for each $s \in \widehat{S}$, $\varphi^{Ch}_s(\widehat{P}) = \varphi^{Ch}_s(P)$. Therefore, if $\varphi^{Ch}(\widehat{P}) \neq \varphi^{Ch}(P)$ then, $\varphi^{Ch}(\widehat{P})$ Pareto dominates $\varphi^{Ch}(P)$ at P.

such that for each $P \in \widetilde{\mathcal{P}}$, $\varphi(P, Ch) = \mathcal{M}^{SPE}(\widetilde{G}[\pi, M])$, *where* $M = (P, Ch)$ *is the DA mechanism, that is,* $\varphi(., Ch) = \varphi^{Ch}$.

When the outcome of an SPE is order independent no student cares about his order of application. Otherwise, students have preferences over orders. We next model such preferences. We say that a non-empty subset $\widehat{S}$ of students have the same relative ranking under π and $\widehat{\pi}$ if for each $s, \widehat{s} \in \widehat{S}$, s is ordered before $\widehat{s}$ in π if and only if s is ordered before $\widehat{s}$ in $\widehat{\pi}$. Formally, for each $s, \widehat{s} \in \widehat{S}$, $\pi^{-1}(s) < \pi^{-1}(\widehat{s})$ if and only if $\widehat{\pi}^{-1}(s) < \widehat{\pi}^{-1}(\widehat{s})$. The following proposition says that essentially students prefer orders where they move later.

Proposition 2. *Given a market* M *and a student* s, *let* π *and* $\widehat{\pi}$ *be two orders such that (i) the set* $S \setminus \{s\}$ *has the same relative ranking under* π *and* $\widehat{\pi}$ *and (ii)* $\widehat{\pi}^{-1}(s) < \pi^{-1}(s)$. *Then,* s *weakly prefers the outcome of the subgame perfect equilibrium of* $\widetilde{G}[\pi, M]$ *to the outcome of the subgame perfect equilibrium of* $\widetilde{G}[\widehat{\pi}, M]$.

Appendix C contains the proof of Proposition 2. A fixed (deterministic) order of applications favors students moving late. This order is not observed. Some unmodeled details of the market may introduce randomness to how it can be determined. We turn to random orderings in the next section.

4 Extension to random ordering

So far we considered deterministic orderings of students. Since these orders are not known, we can assume that they are randomly determined. We introduce “nature" in the game whose role is to randomly determine the order of applications. Nature first randomly chooses the student who must move first and following each application decision, randomly chooses the student who must move next among the remaining ones.

Given a non empty subset $\widehat{S} \subseteq S$, a *probability distribution* $F_{\widehat{S}}$ over $\widehat{S}$ is a function $F_{\widehat{S}} : \widehat{S} \to [0, 1]$ such that $\sum_{s \in \widehat{S}} F_{\widehat{S}}(s) = 1$. A *device* is a collection $F = (F_{\widehat{S}})_{\widehat{S} \subseteq S}$ of probability distributions such that for each $\widehat{S} \subseteq S$, $F_{\widehat{S}}$ is a probability distribution over $\widehat{S}$. A game form is now described by a device F as follows:

Applications phase: nature chooses each student s with probability $F_S(s)$ to move first. Let s_1 be the student chosen and $a_1 \in C \cup \{\emptyset\}$ his application decision such that $a_1 = c$ if he applies to college c and $a_1 = \emptyset$ if he chooses to remain unmatched. For each $t = 1, ..., n-1$, let $s_1, ..., s_t$ be the students chosen by the device respectively for the first move to the tth move and let $a_1, ..., a_t$ be their application decisions such that $a_{\widehat{t}}$ is the application decision of $s_{\widehat{t}}$. Let $\widehat{S} := \{s_1, ..., s_t\}$. Then, nature chooses each remaining student $s \in S \setminus \widehat{S}$ with probability $F_{S \setminus \widehat{S}}(s)$ to move next. Let s_{t+1} be the student chosen. Then, s_{t+1} observes the applications $a_1, ..., a_t$ and makes an application decision $a_{t+1} \in C \cup \{\emptyset\}$.

Admissions phase: colleges offer admissions following the order of their index numbers. Given application decisions $a_1, ..., a_n$ such for each t, a_t is the application decision of s_t, each college c observes all the decisions up to its turn and offers admissions to a subset $o(c) \subseteq \{s_t \in S | a_t = c\}$ of its applicants.

Given a device F and a market M, let $G[F, M]$ be the game induced by F and M. An action for nature after any application decision consists of choosing a student among the remaining ones. Let h^0 be the empty history. Given a student s such that $F_S(s) \neq 0$, (h^0, s) is the history in which the device F chooses s first. Let a_1 be his decision. Then, (h^0, s, a_1) is the history in which the device chooses s first and he takes the decision a_1. Again, given a student $\widehat{s} \in S \setminus \{s\}$ such that $F_{S \setminus s}(\widehat{s}) \neq 0$, $(h^0, s, a_1, \widehat{s})$ is the history following (h^0, s, a_1) in which the device chooses $\widehat{s}$ next. More generally, every such history alternates a student with an application decision.[13] Let $\mathcal{H}$ denote the histories. Then for each student s, let $\mathcal{H}_s = \{(h, s) \in \mathcal{H} | h \in \mathcal{H}\}$ denote the histories at which s is chosen to move next.

Given a device F and a market M, s's strategy in $G[F, M]$ is a function $\sigma_s : \mathcal{H}_s \to C \cup \{\emptyset\}$. Given a game $G[F, M]$, every strategy profile in $G[F, M]$ induces a probability distribution over the set of matchings attached to the terminal histories. This introduces the notion of random matching.

A *random matching* η is a probability distribution over $\mathcal{M}$. Given a random matching η and a student s let η_s denote the probability distribution which η induces over the set $C \cup \{\emptyset\}$ of outcomes for s. We assume that students evaluate random matchings according to first-order stochastic dominance. The random matching η *first-order stochastically* P_s*-dominates* the random matching η', in notation $\eta P_s^{sd} \eta'$, if for each $v \in C \cup \{\emptyset\}$,

$$\sum_{v' \in C \cup \{\emptyset\} : v' R_s v} \eta_s(v') \geq \sum_{v' \in C \cup \{\emptyset\} : v' R_s v} \eta'_s(v').$$

[13] Note that we only model histories before any college' decision. This is again an abuse to simplify the analysis since colleges optimal decisions are known.

Given a subset $\widehat{S}$ of students let $\pi : \{1, ..., |\widehat{S}|\} \to \widehat{S}$ be a bijection. It describes a possible order in which a device might choose the set of students in $\widehat{S}$ if they are the remaining ones. Let $\mathcal{O}(\widehat{S})$ be the set of all such bijections. Then any device F induces a probability distribution $\widetilde{\pi}_{\widehat{S}}$ over $\mathcal{O}(\widehat{S})$. For each $\pi \in \mathcal{O}(\widehat{S})$, let $Pr(\widetilde{\pi}_{\widehat{S}} = \pi)$ denote the probability that $\widetilde{\pi}_{\widehat{S}}$ places on π.

Fix $\widehat{S} \subset S$, $s \notin \widehat{S}$ and let $h \in \mathcal{H}_s$ be a history such that at h, $\widehat{S}$ is the set of remaining students. Then, an action $a \in C \cup \{\emptyset\}$ taken by s at h, a strategy profile $\sigma_{\widehat{S}}$ for the remaining students and an order $\pi \in \mathcal{O}(\widehat{S})$ induce a specific matching $\mu^{\pi}(a, \sigma_{\widehat{S}}|_h)$. Therefore, considering all possible orders of $\widehat{S}$, the action a and the strategy profile $\sigma_{\widehat{S}}$ induce a random matching $\eta(a, \sigma_{-s}|_h)$ as follows:

$$Pr(\eta(a, \sigma_{\widehat{S}}|_h) = \mu) = \sum_{\pi \in \mathcal{O}(\widehat{S}):\, \mu^{\pi}(a, \sigma_{\widehat{S}}|_h) = \mu} Pr(\widetilde{\pi}_{\widehat{S}} = \pi).$$

Given a device F, a market $M = (P, Ch)$, student s and a history $h \in \mathcal{H}_s$, s's strategy σ_s *first-order stochastically dominates* his strategy σ'_s in the subgame $G[F, M|_h]$ if for each strategy profile σ_{-s},

$$\eta(\sigma|_h)\, P_s^{sd}\, \eta(\sigma'_s, \sigma_{-s}|_h).$$

A *student strategy is first-order stochastically undominated in some subgame* if it is not first-order stochastically dominated in that subgame.

Subgame perfect first-order stochastically undominated Nash equilibrium (sd-SPUE). *Given a device F and a market M, we say that a strategy profile σ in $G[F, M]$ is a subgame perfect first-order stochastically undominated Nash equilibrium (sd-SPUE) of $G[F, M]$ if it induces a first-order stochastically undominated Nash equilibrium in every subgame of $G[F, M]$.*

The set of sd-SPUEs of $G[F, M]$ corresponds to the set of strategy profiles σ in which every agent uses either a solidarity strategy or a bossy strategy. A solidarity strategy simply adapts to the present set-up as follows: given a student s, for each $h \in \mathcal{H}_s$, $\sigma^*_s(h) = \varphi_s^{Ch}(\boldsymbol{P}(h))$. A Bossy strategy also adapts similarly. Given a device F and a market M, let $\Delta^{sd-SPUE}(G[F, M])$ denote the set of all random matchings induced by the sd-SPUEs of $G[F, M]$. For each random matching in $\Delta^{sd-SPUE}(G[F, M])$, the matchings which it assigns positive probabilities, its support, weakly Pareto dominate the outcome of the DA algorithm.

Theorem 5. *Given a device F and a market $M = (P, Ch)$, let η be a random matching in $\Delta^{sd-SPUE}(G[F, M])$. Then, every matching in the support of η weakly Pareto dominates the outcome of the DA algorithm at P.*

Appendix H contains the proof of Theorem 5.

5 Related literature and conclusion

The theoretical appeal of stability for matching markets has led most papers on decentralized matching markets to search for game forms that implement stable matchings. They are game forms where students simultaneously apply and colleges sequentially admit by Romero-Medina and Triossi (2012); game forms with simultaneous moves by Alcalde and Romero-Medina (2000) and Triossi (2012) and game forms with sequential moves by Haeringer and Wooders (2008), Pais (2005), Diamantoundi et al.(2015), Wu(1015) and Suh and Wen (2008). In some of these contributions, unstable matchings were starting to emerge in equilibrium (Suh and Wen, 2008; Diamantoundi et al., 2015; Wu, 2015; Triossi, 2012) and further conditions were imposed.

There is a prior reference to our contribution in Theorem 4. Indeed, Suh and Wen (2008) already identified a subset of the markets that we characterize in that theorem. They consider (1) a one-to-one matching market: $q_c = 1$ for each $c \in C$; (2) a balanced market: $|S| = |C|$; (3) a mutual acceptability: for each $c \in C$ and each $s \in S$, $cR_s\emptyset$ and $s \succ_c \emptyset$ and (4) an additional condition (α^M condition) that is equivalent to the claiming consistency we formulated. In an equivalent game, they proved that the game has a unique SPE such that its outcome is order independent and produces the outcome of the DA algorithm. However, their result does not hold if we drop market balancedness or mutual acceptability. Thus, Theorem 4 strengthens their result.

Next, it is well-known that no strategy-proof mechanism can Pareto dominates the DA mechanism. Recently, Bando (2014) unified two independent solutions— by Ehlers (2007), Wako (2010) and Kesten (2010)— which weakly Pareto dominate the DA mechanism; and, showed that these solutions coincide and can be supported by a strictly strong Nash equilibrium.

Finally, the introduction of first-order stochastic dominance criterion to evaluate probability distributions has a connection with prior results. The notion is introduced by D'Aspremont and Peleg (1988) and employed in matching theory by Ehlers and Masso

(2007, 2015) in incomplete information environment and Pais (2008) in decentralized random matching environment.

Unraveling has led many matching markets to adopt a centralized matching procedure. However, the result in this paper supports the continued use of early decision, at least when students can coordinate their applications. As part of the market design literature, this piece provides an understanding of an important aspect of the market that can improve students welfare.

References

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

[2] Alva, S. and Manjunath, V. (2016). "Strategy-proof Pareto-improvement". Working Paper. University of Texas.

[3] Avery, C. and Levin, J. (2010). "Early Admissions at Selective Colleges". *American Economic Review.* 100:2125-2156.

[4] Aygün, O. and Sönmez, T. (2012). "Matching with Contracts: The Critical Role of Irrelevance of Rejected Contracts". *Working Paper.* Boston University.

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

[6] Coles, P.; Kushnir, A. and Niederle, M. (2013). "Preference Signaling in Matching Markets". *American Economic Journal: Microeconomics.* 5:99-134.

[7] D'Aspremont, C. and Peleg, B. (1988). "Ordinal Bayesian incentive compatible representation of committees". *Social Choice Welfare.* 5:261-280.

[8] Diamantoudi, E.; Miyagawa, E. and Xue, L. (2015). "Decentralized matching: The role of commitment". *Games and Economic Behavior.* 92:1-17.

[9] Ehlers, L. (2007). "Von Neumann-Morgenstern stable sets in matching problems". *Journal of Economic Theory.* 134:537-547.

[10] Ehlers, L. and Masso, J. (2007). “Incomplete information and singleton cores in matching markets”. *Journal of Economic Theory.* 136: 587-600.

[11] Ehlers, L. and Masso, J. (2007). “Matching markets under (in)complete information”. *Journal of Economic Theory.* 157:295-314.

[12] Ehlers, L. and Klaus, B. (2014). “Object Allocation via Deferred-Acceptance: Strategy-Proofness and Comparative Statics”. *Games and Economic Behavior.* 97:128-146.

[13] Ergin, H. (2002). “Efficient Resource Allocation on the Basis of Priorities”. *Econometrica.* 70:2489-2497.

[14] Gale, D. and Shapley, L. (1962). “College Admissions and the stability of Marriage”. *American Mathematical Monthly.* 69:9-15.

[15] Kesten, O. (2010). “School choice with consent”. *Quarterly Journal of Economics.* 125: 1297-1348.

[16] Klijn, F. (2011) “On the Consistency of Deferred Acceptance when Priorities are Acceptant Substitutable”. *Mathematical Social Sciences.* 62:101-103.

[17] Hatfield, W. and Milgrom, P. (2005). “Matching with Contracts”. *American Economic Review.* 95:913-935.

[18] Kojima, F. and Manea, M. (2010). “Axioms For Deferred Acceptance”. *Econometrica.* 78:633-653.

[19] Kuhn, H. W. (1953). “Extensive games and the problem of information”. In H. W. Kuhn et A. W. Tucker (Eds.). Contributions to the theory of games II. Princeton. NJ: Princeton University Press. pp. 193-216.

[20] Moldovanu, B. and Winter, E. (1994). “Order independent equilibria”. *Games and Economic Behavior.* 9:21-34.

[21] Osborne, M. and Rubinstein A. (1994). “A course in game theory”. MIT Press. Cambridge.

[22] Pais, J. (2008). “Incentives in decentralized random matching markets”. *Games and Economic Behavior.* 64: 632-649.

[23] Romero-Medina, A. and Triossi, M. (2014). "Non-revelation mechanisms in many-to-one markets". *Games and Economic Behavior*. 87:624-630.

[24] Roth, A. (1982). "The economics of matching: Stability and incentives". *Mathematics of Operations Research*. 7:617-628.

[25] Roth, A. (1984). "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory". *Journal of Political Economy*. 92:991-1016.

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

[27] Roth, A. and Sotomayor, O.S. (1990). "Two-sided matching: A study in game theoretic modeling and analysis". Econometric society monograph. Cambridge University Press.

[28] Roth A. and Xing, X. (1997). "Turnaround time and bottlenecks in market clearing: decentralized matching in the market for clinical psychologists". *Journal of Political Economy*. 105:284-29.

[29] Suh, S. and Wen, Q. (2008). "Subgame Perfect Implementation of Stable Matchings in Marriage Problems". *Social Choice Welfare*. 31:163-174.

[30] Thomson, W. (2011). "Consistency and its Converse: An introduction". *Review of Economic Design*. 15:257-291.

[31] Triossi, M. (2009). "Hiring mechanisms, application costs and stability". *Games and Economic Behavior*. 66:566-575.

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

[33] Wu, Q. (2015). "A finite decentralized marriage market with bilateral search". *Journal of Economic Theory*. 160:216-242.

Appendices

Appendix A: Proof of Theorem 1

The proof uses known results that we collect first.

The DA mechanisms possess some form of monotonicity established by Kojima and Manea (2010). Given $s \in S$ and $P_s, \widehat{P}_s \in \widetilde{\mathcal{P}}$, we say that $\widehat{P}_s$ is an *individually rational (IR-) monotonic transformation*[14] of P_s at $v \in C \cup \{\emptyset_{na}\}$, in notation $\widehat{P}_s$ *i.r.m.t* P_s at v, if any college that is ranked above both v and $\emptyset_{na}$ under $\widehat{P}_s$ is also ranked above v under P_s, that is

$$\text{for each } c \in C,\ c\,\widehat{P}_s\,v \text{ and } c\,\widehat{P}_s\emptyset_{na} \;\Rightarrow\; c\,P_s\,v.$$

Of course, P_s *i.r.m.t* P_s at $v \in C \cup \{\emptyset_{na}\}$. The following case deserves to be illustrated separately as it will be the main form of IR-monotonic transformation we will be using. Given $c \in C$ and $s \in S$, let $P_s^c \in \widetilde{\mathcal{P}}$ be s's preference relation where c is his unique acceptable college and $P_s^{\emptyset_{na}}$ his preference relation where no college is acceptable.

Remark. *Given $s \in S$, $P_s \in \widetilde{\mathcal{P}}$ and $v \in C \cup \{\emptyset_{na}\}$, let $\mu \in \mathcal{M}$ be such that $v\,R_s\,\mu_s\,R_s\emptyset_{na}$. Then, P_s^v i.r.m.t P_s at μ_s.*

We say that $\widehat{P}$ is an *IR-monotonic transformation* of P at matching μ, in notation $\widehat{P}$ *i.r.m.t* P at μ, if for each $s \in S$, $\widehat{P}_s$ *i.r.m.t* P_s at μ_s.

IR-monotonicity. *A matching mechanism φ is IR-monotonic if for each $P, \widehat{P} \in \mathcal{P}^S$, $\widehat{P}$ i.r.m.t P at $\varphi(P) \Rightarrow \varphi(\widehat{P})\,\widehat{R}_S\,\varphi(P)$.*

Next, we will use the following incentive compatibility property of mechanisms.

Strategy-proofness. *A matching mechanism φ is strategy-proof if for each $P \in \mathcal{P}^S$, each $s \in S$ and each $\widehat{P}_s \in \mathcal{P}$, $\varphi_s(P)\,R_s\,\varphi_s(\widehat{P}_s, P_{-s})$.*

The DA mechanism φ^{Ch} induced by any profile Ch of choice functions is strategy-proof (Lemma 1). Beside this result (and others), the choice functions themselves have others properties. We collect the relevant ones.

Path-independence. *College c's choice function Ch_c is path-independent if for each $\widehat{S} \subseteq S$ and each $s \in S, Ch_c(Ch_c(\widehat{S}) \cup \{s\}) = Ch_c(\widehat{S} \cup \{s\})$.*

[14] See Kojima and Manea (2010) for further discussions.

Irrelevance of rejected students. *College c's choice function Ch_c satisfies the irrelevance of rejected students if for each $\widehat{S} \subsetneq S$ and each $s \notin \widehat{S}$, $s \notin Ch_c(\widehat{S} \cup \{s\}) \Rightarrow Ch_c(\widehat{S} \cup \{s\}) = Ch_c(\widehat{S})$.*

We now gather in a lemma these results.

Lemma 1. *Let Ch be a profile of choice functions and φ^{Ch} the DA mechanism induced by Ch. Then*

(1) φ^{Ch} is IR-monotonic (Kojima and Manea, 2010, Theorem 1).

(2) φ^{Ch} is strategy-proof (Hatfield and Milgrom, 2005).

(3) For each $c \in C$, Ch_c is path-independent (Ehlers and Klaus, 2016, Lemma 1).

(4) For each $c \in C$, Ch_c satisfies the irrelevance of rejected students (Ehlers and Klaus, 2016; Aygün and Sönmez, 2012).

Point 2 of Lemma 1 is established by Hatfield and Milgrom (2005) when each college's preference is substitutable and satisfies the *law of aggregate demand*, that is, for each $c \in C$ with choice function Ch_c, for each $\widehat{S} \subseteq S$ and each $S' \subseteq S$, $S' \subseteq \widehat{S}$ implies $|Ch_c(S')| \leq |Ch_c(\widehat{S})|$. However, an acceptant preference satisfies this law (Ehlers and Klaus, 2016).

Our proofs also use a result by Abdulkadiroğlu et al.(2009) when each college has a *responsive preference*,[15] and recently obtained by Alva and Munjunath (2016) in a more general model encompassing acceptant substitutable preferences.

Lemma 2 (Abdulkadiroğlu et al. 2009, claim in Theorem 1)**.** *Let $M = (P, Ch)$ be a market and μ^* the outcome of the DA algorithm at M. If a matching μ Pareto dominates μ^* at P, then, the same set of students are matched in both μ and μ^*, in notation, $\mu_C = \mu^*_C$.*

We are now ready to prove the theorem. Fix an order π and a market $M = (P, Ch)$ and let $G[\pi, M]$ be the game induced by π and M. By convention, whenever we consider a history $h^t = (h^0, a_1 ..., a_t)$, $a_{\widehat{t}}$ is the application that $s_{\widehat{t}}$ makes. To simplify the notation, we suppress the reference to the choice profile Ch in the DA mechanism φ^{Ch} induced

[15] Let $\succ_c$ be c's preference relation over $S \cup \{\emptyset\}$. A preference relation $\bar{\succ}_c$ *over* 2^S *is responsive to* $\succ_c$, if for any subset $\widehat{S}$ of students with $|\widehat{S}| < q_c$, $\widehat{S} \cup \{s\} \bar{\succ}_c \widehat{S} \cup \{s'\}$ if and only if $s \succ_c s'$ and $\widehat{S} \cup \{s\} \bar{\succ}_c \widehat{S}$ if and only if $s \succ_c \emptyset$.

by Ch and just write $\varphi := \varphi^{Ch}$. We continue by establishing two important results as lemmas. For each $t = 1, ..., n$, each history $h^t = (h^0, a_1, ..., a_t)$ and each $v \in C \cup \{\emptyset_{na}\}$, let $S_v(h^t) := \{s_{\widehat{t}} | \widehat{t} \leq t, a_{\widehat{t}} = v\}$.

Lemma 3. *Fix a history $h^{n-1} \in \mathcal{H}_\pi^{n-1}$ and let $\mu^{n-1} := \varphi(\boldsymbol{P}(h^{n-1}))$.*

(i) assume that s_t applies according to solidarity strategy, and let $\widehat{v} = \sigma^*_{s_n}(h^{n-1})$. Let $h^n = (h^t, \widehat{v})$ be the terminal history following such application. Then, for each $c \in C$, $\mu_c^{n-1} = Ch_c(S_c(h^n))$.

(ii) assume that $\mu_{s_n}^{n-1} = \emptyset$ and s_n applies to an acceptable college $\widehat{c}$ and let $\widehat{h}^n = (h^{n-1}, \widehat{c})$ be the terminal history following such application. Then, for each $c \in C$, $\mu_c^{n-1} = Ch_c(S_c(\widehat{h}^n))$.

Proof of Lemma 3. (i). Since we only have two histories to consider, we further simplify the notations; $S_c := S_c(h^{n-1})$ and $\widehat{S}_c := S_c(h^n)$.

Now consider the DA algorithm that produces μ^{n-1} at $(\boldsymbol{P}(h^{n-1}), Ch)$. Recall that $\boldsymbol{P}_{s_n}(h^{n-1}) = P_{s_n}$ and for each $t \neq n$, we have $\boldsymbol{P}_{s_t}(h^{n-1}) = P_{s_t}^{a_t}$. Therefore, each student in $S \setminus \{s_n\}$ makes a proposal (if any) no further than the first step of the algorithm. Furthermore, by $\sigma^*_{s_n}(h^{n-1}) = \mu_{s_n}^{n-1} = \widehat{v}$, we have

$$s_n \in \widehat{S}_{\widehat{v}}. \tag{1}$$

Fix $c \in C$ and consider the set of students proposing to c in the DA algorithm. There are two cases regarding s_n's proposals.

Case 1. s_n did not propose to c.

Then, c received (in the first step) only proposals from students in S_c.[16] Therefore, $\mu_c^{n-1} = Ch_c(S_c)$ and $c \neq \widehat{v}$ imply that $s_n \notin \mu_c^{n-1}$. Now by (1) $s_n \notin \widehat{S}_c$ and we draw that $\widehat{S}_c = S_c$. By substitution, $\mu_c^{n-1} = Ch_c(\widehat{S}_c) = Ch_c(S_c(h^n))$.

Case 2. s_n proposed to c.

If s_n proposed to c in Step 1 of the algorithm, then $\mu_c^{n-1} = Ch_c(S_c \cup \{s_n\})$. If s_n proposed to c in a step later than Step 1, then $\mu_c^{n-1} = Ch_c(Ch_c(S_c) \cup \{s_n\})$. Since Ch_c is path-independent, $Ch_c(Ch_c(S_c) \cup \{s_n\}) = Ch_c(S_c \cup \{s_n\})$. In any case,

$$\mu_c^{n-1} = Ch_c(S_c \cup \{s_n\}). \tag{2}$$

First, assume that $s_n \in \mu_c^{n-1}$. Then $c = \widehat{v}$ and by (1), $s_n \in \widehat{S}_c$. Hence $\widehat{S}_c = S_c \cup \{s_n\}$.

[16] S_c is the set of students from whom c receives applications before s_n's turn in the game $G[\pi, M]$. Thus, $\boldsymbol{P}_s(h^{t-1}) = P_s^c$ for all $s \in S_c$.

Then combining this and (2) obtains $\mu_c^{n-1} = Ch_c(\widehat{S}_c)$. Second, assume that $s_n \notin \mu_c^{n-1}$. Then from (2), $s_n \notin Ch_c(S_c \cup \{s_n\})$. Since Ch_c satisfies the irrelevance of rejected students, $Ch_c(S_c \cup \{s_n\}) = Ch_c(S_c)$. Again $c \neq \widehat{v}$ and by (1), $s_n \notin \widehat{S}_c$ and thus, $\widehat{S}_c = S_c$. Consequently, $Ch_c(S_c \cup \{s_n\}) = Ch_c(S_c) = Ch_c(\widehat{S}_c)$. Finally combining the later result with (2) obtains $\mu_c^{n-1} = Ch_c(\widehat{S}_c) = Ch_c(S_c(h^n))$.

(ii). For this case, we further simplify the notation and let $S_c^* := S_c(\widehat{h}^n)$ for each $c \in C$. With regard to the proof of (i), we only need to consider college $\widehat{c}$. Since s_n applied to $\widehat{c}$, $S_{\widehat{c}}^* = S_{\widehat{c}} \cup \{s_n\}$. Consider the DA algorithm that produces μ^{n-1}. Since $\mu_{s_n}^{n-1} = \emptyset_{na}$ and $\widehat{c}$ is acceptable to him, s_n has proposed to that college in some step of the algorithm. Then using (2) obtain $\mu_{\widehat{c}}^{n-1} = Ch_{\widehat{c}}(S_{\widehat{c}} \cup \{s_n\})$. Combining this with $S_{\widehat{c}}^* := S_{\widehat{c}} \cup \{s_n\}$ obtains $\mu_{\widehat{c}}^{n-1} = Ch_{\widehat{c}}(S_{\widehat{c}}^*) = Ch_{\widehat{c}}(S_{\widehat{c}}(\widehat{h}^n))$. □

Lemma 4. *Let $t = 1, ..., n-1$ and fix a history $h^{t-1} = (h^0, a_1, ..., a_{t-1})$. Assume that student s_t applies accordine to $\sigma_{s_t}^*(h^{t-1}) = v$ and let $h^t = (h^0, a_1, ..., a_{t-1}, v)$ be the history following that decision. Let $\mu^{t-1} := \varphi(\boldsymbol{P}(h^{t-1}))$ and $\mu^t := \varphi(\boldsymbol{P}(h^t))$. Then, the same set of students are matched in both μ^{t-1} and μ^t, that is, $\mu_C^{t-1} = \mu_C^t$.*

Proof of Lemma 4. By definition, $\sigma_{s_t}^*(h^{t-1}) = \varphi(\boldsymbol{P}(h^{t-1})$. Thus, $v = \mu_{s_t}^{t-1}$. Since $\boldsymbol{P}_{s_t}(h^{t-1}) = P_{s_t}$, we have $\boldsymbol{P}(h^{t-1}) = (P_{s_t}, \boldsymbol{P}_{-s_t}(h^{t-1}))$ and $\boldsymbol{P}(h^t) = (P_{s_t}^v, \boldsymbol{P}_{-s_t}(h^{t-1}))$. Next, because $P_{s_t}^v$ *i.r.m.t* P_{s_t} at v, we have $\boldsymbol{P}(h^t)$ *i.r.m.t* $\boldsymbol{P}(h^{t-1})$ at $\varphi(\boldsymbol{P}(h^{t-1}))$. Since φ is IR-monotonic, $\varphi(\boldsymbol{P}(h^t))\, \boldsymbol{R}(h^t)\, \varphi(\boldsymbol{P}(h^{t-1}))$ or equivalently,

$$\mu^t\, \boldsymbol{R}(h^t)\, \mu^{t-1}.$$

For s_t, $\mu_{s_t}^t = v$ because $\boldsymbol{P}_{s_t}(h^t) = P_{s_t}^v$ and $\mu_{s_t}^{t-1} = v$. For each student $s \neq s_t$, $\boldsymbol{P}_s(h^t) = \boldsymbol{P}_s(h^{t-1})$. Student s_t is matched to v in both μ^{t-1} and μ^t and the preference relation of each $s \neq s_t$ is the same in both $\boldsymbol{P}(h^{t-1})$ and $\boldsymbol{P}(h^t)$. Therefore, we have

$$\mu^t\, \boldsymbol{R}(h^{t-1})\, \mu^{t-1}. \tag{3}$$

Now (3) and Lemma 2 give $\mu_C^{t-1} = \mu_C^t$. □

We now turn to the proof of Theorem 1. We show that the solidarity strategy σ^* is the unique backwards-induction strategy of $G[\pi, M]$, by induction on $t = 1, ..., n$, starting from n.

Induction base ($t = n$): We verify that at any history h^{n-1}, applying according to $\sigma_{s_n}^*$ is s_n's unique best application at that history. Assume first that $\sigma_{s_n}^*(h^{n-1}) = \widehat{c}$ for some $\widehat{c} \in C$. We show that if s_n applies to $\widehat{c}$, then $\widehat{c}$ will admit him. This is in fact the conclusion of Lemma 3. By this result, if $\widehat{c}$ is s_n's first choice according to P_{s_t}, then we are done. Otherwise, let $c \in C$ be such that

$$c \ P_{s_n} \widehat{c} \tag{4}$$

and assume that s_n applies to c at h^{n-1}. We show that c will not admit him. Let $\mu^{n-1} = \varphi(\boldsymbol{P}(h^{n-1}))$ and for the purpose of this proof let $S_c^* = S_c(h^{n-1})$. Next, pick $s_t \in \mu_c^{n-1}$ with $t \neq n$. Then s_t applied to c and therefore $s_t \in S_c^*$. Therefore, $\mu_c^{n-1} \subseteq S_c^*$ and $\mu_c^{n-1} \cup \{s_n\} \subseteq S_c^* \cup \{s_n\}$. Since c's preference is substitutable, if $s_n \in Ch_c(S_c^* \cup \{s_n\})$, then

$$s_n \in Ch_c(S_c^* \cup \{s_n\}) \cap (\mu_c^{n-1} \cup \{s_n\}) \subseteq Ch_c(\mu_c^{n-1} \cup \{s_n\}). \tag{5}$$

From (4), (5) and the fact that $\mu_{s_n}^{n-1} = \widehat{c}$, the pair (s_n, c) blocks μ^{n-1} at $(\boldsymbol{P}(h^{n-1}), Ch)$, contradicting the stability of μ^{n-1} at $(\boldsymbol{P}(h^{n-1}), Ch)$. In conclusion, $s_n \notin Ch_c(S_c^* \cup \{s_n\})$ and c will not admit s_n. Assume now that $\sigma^*(h^{n-1}) = \emptyset_{na}$. Then, the last conclusion says that if s_n applies to an acceptable college he will not be admitted. Clearly, by costly application choosing to remain unmatched ($\emptyset_{na}$) is better than applying to an unacceptable college and either not being admitted ($\emptyset_a$ is the resulted outcome) or being admitted.

Induction hypothesis: Let t be such that $t < n$ and assume that for each $\widehat{t}$ with $t < \widehat{t} \leq n$, student $s_{\widehat{t}}$ uses solidarity strategy, that is, for each $h^{\widehat{t}-1} \in \mathcal{H}_\pi^{\widehat{t}-1}$, $\sigma_{s_{\widehat{t}}}^*(h^{\widehat{t}-1}) = \varphi_{s_{\widehat{t}}}(\boldsymbol{P}(h^{\widehat{t}-1}))$.

Induction step: Fix an arbitrary history $h^{t-1} \in \mathcal{H}_\pi^{t-1}$. We show that the action $\sigma_{s_t}^*(h^{t-1})$ is the unique best response for s_t to $\sigma_{-s_t}^*|_{h^{t-1}}$ in $G[\pi, M|_{h^{t-1}}]$. We distinguish two cases:

Case 1: $\sigma_{s_t}^*(h^{t-1}) = \widehat{c}$ for some $\widehat{c} \in C$.

We consider the case where s_t applies to $\widehat{c}$ (Case 1.1) and a case where s_t applies to a college c with $c \, P_{s_t} \, \widehat{c}$ (Case 1.2), if any.

Case 1.1. s_t applies to $\widehat{c}$. Then $\widehat{c}$ will admit him.

Let $h^t, ..., h^n$ be the histories reached after s_t's application and each of the remaining students apply according to solidarity strategy. Let $\mu^{\widehat{t}} := \varphi(\boldsymbol{P}(h^{\widehat{t}}))$, $\widehat{t} = t-1, ..., n$. We know that $s_t \in \mu_{\widehat{c}}^{t-1}$ since $\varphi_{s_t}(\boldsymbol{P}(h^{t-1})) := \sigma_{s_t}^*(h^{t-1}) = \widehat{c}$ and $\mu^{t-1} := \varphi(\boldsymbol{P}(h^{t-1}))$. Now apply Lemma 4 to obtain $s_t \in \mu_C^t$,..., $s_t \in \mu_C^{n-1}$. Next because $\boldsymbol{P}_{s_t}(h^{n-1}) = P_{s_t}^{\widehat{c}}$ and

μ^{n-1} is IR at $(\boldsymbol{P}(h^{n-1}), Ch)$, $s_t \in \mu_C^{n-1}$ implies that $\mu_{s_t}^{n-1} = \widehat{c}$. Finally, apply Lemma 3 to obtain $s_t \in \mu_{\widehat{c}}^{n-1} = Ch_{\widehat{c}}(S_{\widehat{c}}(h^n))$. Therefore, $\widehat{c}$ will admit s_t.

By Case 1.1, if $\widehat{c}$ is s_t's first choice, then applying to $\widehat{c}$ is his unique best response at h^{t-1}. Otherwise, let $c \in C$ be such that

$$c\, P_{s_t}\, \widehat{c} = \sigma_{s_t}^*(h^{t-1}).$$

Case 1.2. s_t applies to c. Then c will not admit him.

Let $\widehat{h}^t, ..., \widehat{h}^n$ be the histories reached after this application and each of the remaining students apply according to solidarity strategy. Let $\widehat{\mu}^t := \varphi(\boldsymbol{P}(\widehat{h}^t))$, $\widehat{t} = t, ..., n-1$. By definition, $\boldsymbol{P}(h^{t-1}) = (P_{s_t}, \boldsymbol{P}_{-s_t}(h^{t-1}))$ and $\boldsymbol{P}(\widehat{h}^t) = (P_{s_t}^c, \boldsymbol{P}_{-s_t}(h^{t-1}))$. Since φ is strategy-proof, we have $\varphi_{s_t}(P_{s_t}, \boldsymbol{P}_{-s_t}(h^{t-1}))\; R_{s_t}\; \varphi_{s_t}(P_{s_t}^c, \boldsymbol{P}_{-s_t}(\widehat{h}^t))$ or equivalently

$$\mu_{s_t}^{t-1}\; R_{s_t}\; \widehat{\mu}_{s_t}^t. \tag{6}$$

Since $c P_{s_t} \widehat{c} = \mu_{s_t}^{t-1}$ by assumption and R_{s_t} is strict, (6) implies $c P_{s_t} \widehat{\mu}_{s_t}^t$ and thus $c \neq \widehat{\mu}_{s_t}^t$. Now because $\boldsymbol{P}_{s_t}(\widehat{h}^t) = P_{s_t}^c$ and $\widehat{\mu}_{s_t}^t$ is IR at $(\boldsymbol{P}(\widehat{h}^t), Ch)$, we have $\widehat{\mu}_{s_t}^t = \emptyset_{na}$. Next, apply Lemma 4 to obtain that $s_t \notin \widehat{\mu}_C^t, ..., s_t \notin \widehat{\mu}_C^{n-1}$. Thus, $s_t \notin \widehat{\mu}_c^{n-1}$. Finally apply Lemma 3 to obtain that $s_t \notin Ch_c(S_c(\widehat{h}^n)) = \widehat{\mu}_c^{n-1}$. Hence, c will not admit s_t.

Case 2: $\sigma_{s_t}^*(h^{t-1}) = \emptyset_{na}$.

By an argument similar to Case 1.2, s_t cannot be admitted by an acceptable college. Suppose that at history h^{t-1}, he applies to an unacceptable college, say c. Then, either c admits him and he would have been better off not applying or c does not admit him and since applications are costly choosing to remain unmatched would have been a best response.

In conclusion, σ^* is the unique backwards-induction strategy of $G[\pi, M]$.

Appendix B: Proof of Theorem 2

Let π be an order and $M = (P, Ch)$ a market with $P \in \mathcal{P}^S$. We need to show that the set of SPUEs of $G[\pi, M]$ corresponds to the profiles $\sigma = (\sigma_s)_{s \in S}$ of strategies such that for each $s \in S$, σ_s is a solidarity strategy or a bossy strategy. An SPUE of $G[\pi, M]$ is a strategy profile that induces an undominated Nash equilibria in every subgame of $G[\pi, M]$. It is itself an SPE of $G[\pi, M]$. We continue by establishing two results as lemmas.

Lemma 5. *Let $t = 1, ..., n-1$ and $h^{t-1} \in \mathcal{H}_\pi^{t-1}$ a history such that $\varphi_{s_t}(\boldsymbol{P}(h^{t-1})) = \emptyset$.*

Assume that student s_t applies to an acceptable college, say c and let $h^t = (h^{t-1}, c)$ be the history following such application. Let $\mu^{t-1} := \varphi(\boldsymbol{P}(h^{t-1}))$ and $\mu^t := \varphi(\boldsymbol{P}(h^t))$. Then, the same set of students are matched in both μ^{t-1} and μ^t, that is, $\mu_C^{t-1} = \mu_C^t$.

Proof of Lemma 5. First, by definition we know that $\boldsymbol{P}_{s_t}(h^{t-1}) = P_{s_t}$ and $\boldsymbol{P}_{s_t}(h^t) = (P^c_{s_t}, \boldsymbol{P}_{-s_t}(h^{t-1}))$. Since $\varphi_{s_t}(\boldsymbol{P}(h^{t-1})) = \emptyset$, φ is strategy-proof and $\varphi_{s_t}(\boldsymbol{P}(h^t))$ is IR at $\boldsymbol{P}(h^t)$ we have $\varphi_{s_t}(\boldsymbol{P}(h^t)) = \emptyset$. Because $\boldsymbol{P}(h^t)$ *i.r.m.t* $\boldsymbol{P}(h^{t-1})$ at $\varphi(\boldsymbol{P}(h^{t-1}))$ and φ is IR-monotonic, we have $\varphi(\boldsymbol{P}(h^t))\ \boldsymbol{R}(h^t)\ \varphi(\boldsymbol{P}(h^{t-1}))$ or equivalently

$$\mu^t\, \boldsymbol{R}(h^t)\, \mu^{t-1}.$$

Now with $\boldsymbol{P}_{-s_t}(h^t) = \boldsymbol{P}_{-s_t}(h^{t-1})$ and $\mu^{t-1}_{s_t} = \emptyset = \mu^t_{s_t}$, we conclude that

$$\mu^t\ \boldsymbol{R}(h^{t-1})\ \mu^{t-1}. \tag{7}$$

Finally, (7) and Lemma 2 obtain $\mu^t_C = \mu^{t-1}_C$. □

From Lemma 5 we establish Lemma 6 which will be used twice in the proof.

Lemma 6. *Let $t = 1, ..., n-1$ and consider student s_t. Assume that each student $s_{\widehat{t}}$ with $\widehat{t} > t$ uses a strategy $\sigma_{s_{\widehat{t}}}$ that is either a solidarity strategy or a bossy strategy. Fix a history $h^{t-1} \in \mathcal{H}^{t-1}_\pi$ and let $v = \varphi_{s_t}(\boldsymbol{P}(h^{t-1}))$. Let c be such that $c\, P_{s_t} v$ and assume that s_t applies to c. Then c will not admit s_t.*

Proof of Lemma 6. Let $h^t, ..., h^n$ be the histories following s_t's application and each student $s_{\widehat{t}}$ with $\widehat{t} > t$ takes the decision $\sigma_{s_{\widehat{t}}}(h^{\widehat{t}-1})$. Let $\mu^{\widehat{t}} := \varphi(\boldsymbol{P}(h^{\widehat{t}}), \widehat{t} = t-1, ..., n-1$. By definition, we have $\boldsymbol{P}(h^{t-1}) = (P_{s_t}, \boldsymbol{P}_{-s_t}(h^{t-1}))$ and $\boldsymbol{P}(h^t) = (P^c_{s_t}, \boldsymbol{P}_{-s_t}(h^{t-1}))$. Since φ is strategy-proof, we have $\varphi_{s_t}(\boldsymbol{P}(h^{t-1}))\, R_{s_t}\, \varphi_{s_t}(\boldsymbol{P}(h^t))$. As $c\, P_{s_t}\, v$ and $v = \varphi_{s_t}(\boldsymbol{P}(h^{t-1}))$, we have $\varphi_{s_t}(\boldsymbol{P}(h^t)) \neq c$. Thus, $\varphi_{s_t}(\boldsymbol{P}(h^t)) = \emptyset$ as $\boldsymbol{P}_{s_t}(h^t) = P^c_{s_t}$ and $\varphi(\boldsymbol{P}(h^t))$ is IR at $(\boldsymbol{P}(h^t), Ch)$. Next, as $\boldsymbol{P}(h^t)$ *i.r.m.t* $\boldsymbol{P}(h^{t-1})$ at $\varphi(\boldsymbol{P}(h^{t-1}))$ and φ is IR-monotonic, $\varphi(\boldsymbol{P}(h^t))\ \boldsymbol{R}(h^t)\ \varphi(\boldsymbol{P}(h^{t-1}))$ or equivalently

$$\mu^t\ \boldsymbol{R}(h^t)\ \mu^{t-1}.$$

Since $\mu^{t-1}_{s_t} = \emptyset = \mu^t_{s_t}$ and $\boldsymbol{P}_{-s_t}(h^t) = \boldsymbol{P}_{-s_t}(h^{t-1})$, we can conclude that $\mu^t\, \boldsymbol{R}(h^{t-1})\, \mu^{t-1}$. By Lemma 2, the same set of students is matched in both μ^t and μ^{t-1}. Finally, by Lemma 5, the same set of students is matched in μ^t through μ^{n-1}. Now apply Lemma

3 (ii) to obtain $\mu_c^{n-1} = Ch_c(S(h^n))$. Thus, $s_t \notin \mu_c^t$ implies $s_t \notin Ch_c(S(h^n))$ and c will not admit s_t. □

We now proceed to the proof of the theorem. Let π be an order and $M = (P, Ch)$ a market. The proof is divided into two steps. We first show that any strategy profile in which each student uses either a solidarity strategy or a bossy strategy is an SPUE of $G[\pi, M]$ (Step 1) and that any SPUE of $G[\pi, M]$ is of this form (Step 2). Let $\sigma = (\sigma_s)_{s\in S}$ be a strategy profile such that for each $s \in S$, σ_s is a solidarity strategy or a bossy strategy.

Step 1: σ is an SPUE of $G[\pi, M]$.

We show this by induction on $t = 1, ..., n$, starting from n.

Induction base ($t = n$): It is a matter of verification fairly similar to the induction base case of the proof of Theorem 1, that for each $h^{n-1} \in \mathcal{H}_\pi^{n-1}$ the decision $\sigma_{s_n}(h^{n-1})$ is a best response for s_n. We leave the details aside and focus on the induction step.

Induction hypothesis: Fix t with $t < n-1$ and assume that for each $\widehat{t}$ such that $t < \widehat{t} \leq n$, student $s_{\widehat{t}}$ uses $\sigma_{s_{\widehat{t}}}$.

Induction step: Fix an arbitrary history $h^{t-1} \in \mathcal{H}_\pi^{t-1}$ and let $\varphi_{s_t}(\boldsymbol{P}(h^{t-1})) = v$. First assume that $v = c$ for some $c \in C$. If s_t applies to c, it will admit him. This is the conclusion of Case 1.1 when its proof is based on Lemma 3 (ii) and Lemma 5. If s_t applies to a college $\widehat{c}$ with $\widehat{c} P_{s_t} c$, then $\widehat{c}$ will not admit him. This is the conclusion of Lemma 5. Therefore, applying to c is the best response of s_t to $(\sigma_{s_{\widehat{t}}})_{\widehat{t}>t}$ at h^{t-1}, that is, applying according to σ_{s_t} at h^{t-1}.[17] Second, assume that $v = \emptyset$. Then by Lemma 5, if he applies to an acceptable college, it will not admit him. Since applying to an unacceptable college is weakly dominated in the subgame $G[\pi, M|_{h^{t-1}}]$, applying according to σ_{s_t} is a best response of s_t to $(\sigma_{s_{\widehat{t}}})_{\widehat{t}>t}$ at h^{t-1}. Indeed, choosing to remain unmatched or applying to any acceptable college at h^{t-1} both lead to the same outcome for s_t of remaining unmatched.

Step 2: Every SPUE of $G[\pi, M]$ is a strategy profile σ such that for each $s \in S$, σ_s is either a solidarity strategy or a bossy strategy of $G[\pi, M]$.

Let $\widehat{\sigma} = (\widehat{\sigma}_s)_{s\in S}$ be an SPUE of $G[\pi, M]$. Assume by contradiction that there exists some t, such that $\widehat{\sigma}_{s_t}$ is neither a solidarity strategy nor a bossy strategy of $G[\pi, M]$. Since the restriction of $\widehat{\sigma}_{s_t}$ to every subgame is not weakly dominated, s_t does not apply to an

[17] Note that according to σ_{s_t}, if $\varphi_{s_t}(\boldsymbol{P}(h^{t-1})) = c$ for some $c \in C$, then s_t applies to c at that history.

unacceptable college under P_s at any history. Thus, there exists a history $h^{t-1} \in \mathcal{H}_\pi^{t-1}$ and a college c such that $\varphi_{s_t}(\boldsymbol{P}(h^{t-1})) = c$ and $\widehat{\sigma}_{s_t}(h^{t-1}) \neq c$. First, $t \neq n$. Indeed, assuming not, by Step 1 of this proof, if s_n applies to c, then it will admit him. Thus, $\widehat{\sigma}_{s_n}(h^{n-1}) P_{s_n} c$. Since $\varphi(\boldsymbol{P}(h^{n-1}))$ is IR at $(\boldsymbol{P}(h^{n-1}), Ch)$ and $\boldsymbol{P}_{s_n}(h^{n-1}) = P_{s_n}$, then $\widehat{\sigma}_{s_n}(h^{n-1}) = \widehat{c}$ for some $\widehat{c} \in C$. Now by Lemma 6, $\widehat{c}$ will not admit him, contradicting optimality of $\widehat{\sigma}_{s_n}(h^{n-1})$ as he could have been admitted by c. In conclusion $t \neq n$. Inductively, fix $\widehat{t} < n$ and assume that $t \neq t'$, for $t' = \widehat{t} + 1, ..., n$. We show that $t \neq \widehat{t}$. Assuming not, we use the previous argument to reach the contradiction that $\widehat{\sigma}_{s_{\widehat{t}}}(h^{\widehat{t}-1})$ is not optimal. Thus, t does not exist, the desire contradiction; which finishes the proof.

Appendix C: Proof of Theorem 3

Let $h^0, ..., h^n$ be the histories in an execution path of a solidarity strategy unique SPE of $\widetilde{G}[P, M]$ (the reasoning is similar with an SPUE). Fix $t = 1, ..., n$ and let $\mu^{\widehat{t}} := \varphi(\boldsymbol{P}(h^{\widehat{t}}))$, $\widehat{t} = 0, ..., n$. Now, recall that by definition, for each $\widehat{t} \leq t$, $\boldsymbol{P}_{s_t}(h^{\widehat{t}}) = P_{s_t}$. Therefore, by (3) or (7), we have $\mu_{s_t}^{t-1} \, R_{s_t} ... \, R_{s_t} \, \mu_{s_t}^0$. Since P_{s_t} is strict, the outcome of the SPE for s_t is $\mu_{s_t}^{t-1}$ and the and the outcome of the DA algorithm at M is μ^0, the following result concludes the proof.

$$\mu_{s_t}^{t-1} \, R_{s_t} \, \mu_{s_t}^0.$$

Appendix D: Proof of Proposition 1

We prove that $\bigcup_{\pi \in \mathcal{O}} \mathcal{M}^{SPUE}(G[\pi, M])$ satisfies the rural hospital theorem. Let $\mu, \widehat{\mu} \in \bigcup_{\pi \in \mathcal{O}} \mathcal{M}^{SPUE}(G[\pi, M])$. Let $\mu^* := \varphi^{Ch}(P)$. By Theorem 2, $\mu \, R \, \mu^*$ and $\widehat{\mu} \, R \, \mu^*$. It is sufficient if we prove that μ and μ^* satisfies part (i) and (ii) of the definition of the rural hospital theorem.

Part (i) follows from Lemma 2. For part (ii), assume that for some $c \in C$, $|\mu_c| < q_c$. By part (i), $|\mu^*| = |\mu_c| < q_c$. This part is complete if we show that $\mu_c \subseteq \mu_c^*$. Let $s \in \mu_c$. If $\mu_s^* \neq c$, then $c \, P_s \, \mu_s^*$, in contradiction with the stability of μ^* since $|\mu_c^*| < q_c$ and c's preference is acceptant. Thus, $\mu_c \subseteq \mu_c^*$; proving that $\mu_c = \mu_c^*$. Alva and Munjunath (2016) independently proved a similar result in a more general model.

Appendix E: Proof of Corollary 1

The result follows from order independence and Theorem 1 in which every student is matched to his outcome under the DA algorithm when he is ordered first.

Appendix F: Proof of Theorem 4

Proof. " $\Rightarrow$ ". Fix a market $M = (P, Ch)$ with $P \in \widetilde{\mathcal{P}}^S$ and assume that it induces an order independent $\mathcal{G}$-outcome and let μ be this unique outcome. Consider the DA mechanism φ^{Ch} induced by Ch. In the remainder we drop the reference to Ch in φ^{Ch}. By Corollary 1,

$$\text{for each } s \in S, \mu_s = \varphi_s(P). \tag{8}$$

We now show that φ is claims consistent at P. Pick an arbitrary subset $\widehat{S} \subseteq S$. We then show that $\varphi(P^{\varphi}_{\widehat{S}}, P_{-\widehat{S}}) = \varphi(P)$. Now pick $s \in S$. We distinguish two cases:

Case 1: $s \in \widehat{S}$. Since, $(P^{\varphi}_{\widehat{S}}, P_{-\widehat{S}})$ *i.r.m.t* P at $\varphi(P)$, and φ is IR-monotonic, it is easily established that $\varphi_s(P^{\varphi}_{\widehat{S}}, P_{-\widehat{S}}) = \varphi_s(P)$.

Case 2: $s \notin \widehat{S}$. Let π be an order such that all students in $\widehat{S}$ are ordered first, and $\pi(|\widehat{S}|+1) = s$. Let $h^{|\widehat{S}|} = (h^o, a_1, ..., a_{|\widehat{S}|})$ be a history in the execution path histories of the solidarity strategy profile of $G[\pi, M]$. By (8) every student in $\widehat{S}$ is matched under μ to his mate under $\varphi(P)$. Therefore, we have $a_1 = \varphi_{s_1}(P)$,...,$a_{|\widehat{S}|} = \varphi_{s_{|\widehat{S}|}}(P)$ and thus $\boldsymbol{P}(h^{|\widehat{S}|}) = (P^{\varphi}_{\widehat{S}}, P_{-\widehat{S}})$. Since $s = s_{|\widehat{S}|+1}$, according to solidarity strategy,

$$\mu_s = \varphi_s(P^{\varphi}_{\widehat{S}}, P_{-\widehat{S}}). \tag{9}$$

Combining (8) and (9) obtains $\varphi_s(P^{\varphi}_{\widehat{S}}, P_{-\widehat{S}}) = \varphi_s(P)$.

Case 1 and Case 2 establish that φ is claims consistent.

" $\Leftarrow$ ". Fix a market $M = (Ch, P)$ and assume that φ is claims consistent at P. We prove that M induces an order independent $\mathcal{G}$-outcome. Given an order π, let $\mu^{\pi} \equiv \mathcal{M}^{SPE}(G[\pi, M])$. We now show by induction on $t = 1, ..., n$ that for each order π and each $s \in S$, $\mu^{\pi}_s = \varphi_s(P)$. By Theorem 1, for each order π and each $s \in S$, $s = s_1$ implies

$$\mu^{\pi}_{s_1} = \varphi_{s_1}(P). \tag{10}$$

Equation (10) is the induction base ($t = 1$). As an induction hypothesis, let $t > 1$ and assume that for each $\widehat{t} < t$, each order π and each $s \in S$, $s = s_{\widehat{t}}$ implies $\mu^{\pi}_{s_{\widehat{t}}} = \varphi_{s_{\widehat{t}}}(P)$.

We now prove the induction step. Fix an arbitrary order π and let $\widehat{S} \equiv \{s_{\widehat{t}} | \widehat{t} < t\}$. Let $h^{t-1} = (h^o, a_1, ..., a_{t-1})$ be a history of the execution path of σ^*. Then by the induction assumption $a_{\widehat{t}} = \varphi_{s_{\widehat{t}}}(P)$, $\widehat{t} < t$. Thus, $\boldsymbol{P}(h^{t-1}) = (P^{\varphi}_{\widehat{S}}, P_{-\widehat{S}})$ and by solidarity strategy,

$$\mu^{\pi}_{s_t} = \varphi_{s_t}(P^{\varphi}_{\widehat{S}}, P_{-\widehat{S}}). \tag{11}$$

Since φ is claims consistent at P, $\varphi_{s_t}(P) = \varphi_{s_t}(P^{\varphi}_{\widehat{S}}, P_{-\widehat{S}})$; and together with (11) obtain $\mu^{\pi}_{s_t} = \varphi_{s_t}(P)$. □

Appendix G: Proof of Proposition 2

Fix a market $M = (P, Ch)$ with $P \in \widetilde{\mathcal{P}}^S$, a student s and two orders π and $\widehat{\pi}$ for which the set $S \setminus \{s\}$ has the same relative ranking under π and $\widehat{\pi}$ and assume that $\widehat{\pi}^{-1}(s) < \pi^{-1}(s)$. Without loss of generality, assume that π and $\widehat{\pi}$ are adjacent as represented below; the difference between π and $\widehat{\pi}$ occurs only on the elements in boxes:

$$\widehat{\pi} : s_1 ... s_{t-1} \boxed{s \; \widehat{s}} ... s_n$$
$$\pi : s_1 ... s_{t-1} \boxed{\widehat{s} \; s} ... s_n.$$

Let σ^* and $\widehat{\sigma}^*$ be the solidarity strategy profiles of $G[\pi, M]$ and $G[\widehat{\pi}, M]$ respectively. By Theorem 1, they are the unique SPEs of $G[\pi, M]$ and $G[\widehat{\pi}, M]$. Consider the histories in the execution paths of these strategies. Since the ordering of the first $t-1$ students and the market are the same in both games, the first t histories in these paths are the same. We represent them as follows:

$$\text{Execution path of } \widehat{\sigma}^* : \quad h^0, ..., h^{t-1}, \widehat{h}^t, ..., \widehat{h}^n.$$
$$\text{Execution path of } \sigma^* : \quad h^0, ..., h^{t-1}, h^t, ..., h^n.$$

Let $\mu^{\pi} \equiv \mathcal{M}^{SPE}(G[\pi, M])$ and $\mu^{\widehat{\pi}} \equiv \mathcal{M}^{SPE}(G[\widehat{\pi}, M])$. We show that $\mu^{\pi}_s R_s \, \mu^{\widehat{\pi}}_s$. Since s decides after history h^{t-1} in $G[\widehat{\pi}, M]$ and after h^t in $G[\pi, M]$, we have

$$\mu^{\widehat{\pi}}_s = \varphi_s(\boldsymbol{P}(h^{t-1})) \tag{12}$$

and

$$\mu^{\pi}_s = \varphi_s(\boldsymbol{P}(h^t)). \tag{13}$$

Now because $\boldsymbol{P}(h^t)$ *i.r.m.t* $\boldsymbol{P}(h^{t-1})$ at $\varphi(\boldsymbol{P}(h^{t-1}))$, we have $\varphi(\boldsymbol{P}(h^t)) \; \boldsymbol{R}(h^t) \; \varphi(\boldsymbol{P}(h^{t-1}))$ as φ is IR-monotonic. Since $\boldsymbol{R}_s(h^t) = R_s$, (12) and (13) obtain $\mu^{\pi}_s R_s \mu^{\widehat{\pi}}_s$ as desired.