FACULTAD ECONOMÍA Y NEGOCIOS
UNIVERSIDAD DE CHILE
DEPARTAMENTO DE ECONOMÍA

SERIE DE DOCUMENTOS DE TRABAJO

SDT 553

STRATEGIC BEHAVIOR WITHOUT OUTSIDE OPTIONS

Autores:
Camilo J. Sirguiado
Juan Pablo Torres-Martínez

Santiago, Enero de 2024

sdt@econ.uchile.cl
econ.uchile.cl/publicaciones

STRATEGIC BEHAVIOR WITHOUT OUTSIDE OPTIONS

CAMILO J. SIRGUIADO AND JUAN PABLO TORRES-MARTÍNEZ

Abstract. In two-sided one-to-one matching markets, when agents have the ability to declare potential partners as unacceptable, no stable mechanism is strategy-proof (Roth, 1982), no Pareto efficient and individually rational mechanism is strategy proof, and each side of the market has a single stable mechanism that is strategy-proof for its members (Alcalde and Barberà, 1994).

These seminal results may no longer be valid when agents have no outside options and the mechanism designer knows it. Under these assumptions, Roth's impossibility theorem holds if and only if there are at least three agents on each side of the market, Alcalde and Barberà's impossibility theorem never holds, and Alcalde and Barberà's uniqueness result only holds for the short side, if there is one. Additionally, among the many stable mechanisms that are strategy-proof for the long side of the market, there is one that is Pareto-superior for the short side. Interestingly, this optimal mechanism does not coincide with any version of deferred acceptance unless the short side has only two agents. All these results can be extended to scenarios in which a part of the population has outside options. Furthermore, when all agents on one side of the market have outside options, Alcalde and Barberà's uniqueness result holds if and only if at most one agent on the other side may not declare any potential partner unacceptable.

Keywords: Matching markets - Outside options - Strategy-proofness - Stability
JEL Classification: D47, C78.

Date: January, 2024.

We are grateful to Gabriela Denis, Jorge Arenas, Marta Faias, Rodrigo Yáñez, and Sebastían Cea-Echenique for useful comments and suggestions.

Camilo J. Sirguiado (cjaramillos@uandes.cl)
School of Business and Economics, Universidad de los Andes, Chile

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

1. Introduction

Matching theory has established itself as an active field of research, thanks to its successful applicability to the design of school choice mechanisms, college admission systems, object allocation rules, and kidney exchange platforms.[1] One area of interest within this field has been the study and characterization of *strategy-proof mechanisms*, that is, centralized protocols that induce agents to truthfully report their preferences independently of the actions of others. Strategy-proofness promotes equity by not giving an advantage to more sophisticated agents and simplifies the interpretation of reported information.[2] Under a strategy-proof mechanism, policymakers should expect social goals to be achieved even when preferences are not observable.

Unfortunately, designing individually rational mechanisms with good stability or efficiency properties without incentivizing someone to distort preferences may be impossible. For instance, in two-sided one-to-one matching markets, Roth (1982, Theorem 3) shows that no stable mechanism is strategy-proof, and Alcalde and Barberà (1994, Proposition 1) established that no Pareto efficient and individually rational mechanism is strategy-proof. These impossibility results are demonstrated by appealing to agents' ability to declare potential partners inadmissible. Essentially, in scenarios in which preferences are heterogeneous, it is shown that some agents may strategically misreport a potential partner as unacceptable in order to be matched with someone better.

Despite these negative results, stability and strategy-proofness become compatible when only one side of the market has strategic agents: as established by Gale and Shapley (1982), Dubins and Freedman (1981), and Roth (1982), the *deferred acceptance mechanism* is stable and strategy-proof for those who make proposals. Moreover, Alcalde and Barberá (1994, Theorem 3) show that it is the only mechanism with these properties. The key to proving this uniqueness result is the fact that, when a stable mechanism other than deferred acceptance is implemented, each agent on the side of the market that makes proposals can get the best stable partner by simply declaring the mate under deferred acceptance as the only acceptable one.[3]

Evidently, when no one may declare others unacceptable, the manipulation strategies underlying the traditional proofs of Roth's impossibility theorem, Alcalde and Barberà's impossibility theorem, and Alcalde and Barberà's uniqueness result become infeasible. Therefore, it is natural to ask whether these classical properties are still valid.

Our objective is to study the effects on strategic behavior of the impossibility of declaring potential partners inadmissible. We assume that agents have no outside options and that the mechanism designer knows it. Under this information requirement, the lack of outside options—a characteristic of preferences—can be associated with the inability to misrepresent preferences by declaring potential partners unacceptable.

[1]We refer to Echenique, Inmorlica, and Vazirani (2023) for surveys on the applications of matching theory.

[2]Evidently, factors such as mistrust, altruism, or cognitive inability can induce agents to misrepresent preferences even when a strategy-proof mechanism is implemented (see Hassidim, Marciano, Romm, and Shorrer, 2017).

[3]The Rural Hospital Theorem (Gale and Sotomayor, 1985) ensures the success of this manipulation strategy, since it implies that those who form a pair when deferred acceptance is implemented will remain paired with someone under any other stable outcome. Therefore, regardless of the stable mechanism that is implemented, agents who declare their mate under deferred acceptance as the only acceptable one are not left alone.

Focusing on two-sided one-to-one matching markets, we show that:

- A stable and strategy-proof mechanism exists if and only if at least one side of the market has only two agents (see Theorem 1).
- A Pareto efficient, individually rational, and strategy-proof mechanism always exists.
- In balanced matching markets, for each side of the market there are many stable mechanisms that are strategy-proof for its members (see Theorem 2).[4]
- In unbalanced matching markets, only the long side of the market has many stable mechanisms that are strategy-proof for its members (see Theorem 2).

These properties fully characterize the scenarios in which Roth's impossibility theorem and Alcalde and Barberà's uniqueness result hold when no one may declare others unacceptable. Moreover, Alcalde and Barberà's impossibility theorem never holds in this context.

Assuming that each side of the market has at least three agents, we show that Roth's impossibility theorem still holds by finding a preference profile in which every stable mechanism is manipulable. Since potential partners cannot be declared unacceptable, the manipulation strategies will maintain the ranking of the best alternatives and reorder others (see the proof of Theorem 1). Alternatively, when a side of the market has only two agents, we prove that the deferred acceptance mechanism that is optimal for this side is strategy-proof for the entire population.[5]

Since individual rationality is trivially satisfied when agents have no outside options, Alcalde and Barberà's impossibility theorem is no longer valid. Indeed, it follows from Svensson (1999) that any version of the *serial dictatorship mechanism* is Pareto efficient, individually rational, and strategy-proof (see Remark 1).

In unbalanced markets, to prove that there are many stable mechanisms that are strategy-proof for the long side of the market, we describe a family of centralized protocols that combine the two versions of deferred acceptance (those obtained by choosing one side of the market to make proposals). Essentially, we find a subdomain of preferences in which the impossibility of declaring others unacceptable prevents agents on the long side of the market from manipulating the outcome of deferred acceptance when the short side makes the proposals. Thus, we construct a mechanism that is strategy-proof for the long side of the market by associating the *short-side optimal* stable matching in that subdomain and the *long-side optimal* stable matching otherwise (see the proof of Theorem 2). For balanced matching markets, a similar strategy allow us to prove that each side of the market has many stable mechanisms that are strategy-proof for its members.

To ensure that Alcalde and Barberà's uniqueness result is valid for the short side of an unbalanced market, we appeal to the Rural Hospital Theorem (Gale and Sotomayor, 1985). This seminal result guarantees that, for each preference profile, there is an agent on the long side of the market who is left alone in all stable matchings. Therefore, when a stable mechanism other than the *short-side optimal* deferred acceptance mechanism is implemented, there are preference profiles at which some

[4]A matching market is *balanced* when there are an equal number of agents on each side.

[5]Without restricting the population size, Alcalde and Barberà (1994, Theorem 2) show that deferred acceptance is strategy-proof when the preference profiles of those who receive proposals satisfy *top dominance*, a restriction that implies that the top choice of each agent determines the rest of her preferences.

agents on the short side can benefit from declaring as first-rated alternatives their best stable partner and someone who is left alone in any stable matchings, in this order (see the proof of Theorem 2).

It is important to note that—regardless of the existence of outside options—our results remain valid when the mechanism designer only knows that everyone considers all potential partners admissible. Indeed, this ensures that an agent may not misrepresent preferences by declaring a potential partner unacceptable, which is the main assumption underlying our findings.[6]

Our analysis can be extended to allow a number of agents to have outside options that they consider better than some potential partners (see Theorems 3 and 4). Among other properties, in this more general framework we show that:

- Roth's impossibly theorem holds when either no side of the market has only two agents or there is at most one agent on each side that may not declare potential partners unacceptable.
- Alcalde and Barberà's impossibly theorem holds as long as agents on one side of the market have no outside options and may not declare potential partners unacceptable.
- In balanced markets, Alcalde and Barberà's uniqueness result holds for one side of the market if and only if at most one agent on that side may not declare inadmissibilities.
- Alcalde and Barberà's uniqueness result holds for the long side of an unbalanced market if and only if at most one agent on that side may not declare potential partners unacceptable.
- If everyone on one side of the market has an outside option, then there are many stable mechanisms that are strategy-proof for the other side if and only if at least two agents on that side may not declare their potential partners unacceptable.

Furthermore, among the stable mechanisms that are strategy-proof for one side of the market, there is always one that is Pareto-superior for the other side (see Theorem 5). In many situations, this optimal mechanism does not coincide with a version of deferred acceptance.

Our results have implications for *school choice* and *college admission* problems in which student priorities are determined by strict rankings (cf., Balinski and Sönmez, 1999; Abdulkadiroğlu and Sönmez, 2003). Indeed, it is well-known that by identifying each institution with a set of representatives, this kind of many-to-one matching problem can be viewed as a one-to-one matching market (cf., Gale and Sotomayor, 1985). Therefore, when students can be declared inadmissible or there are as many students as vacancies, our results imply that the following property holds: if at least two students consider all institutions acceptable, then a stable mechanism exists that improves the welfare of schools/colleges relative to the *student-optimal stable mechanism* and remains strategy-proof for students.

Related literature. To the best of our knowledge, there are no studies that examine the validity of the aforementioned results of Roth (1982) and Alcalde and Barberá (1994) in contexts where agents may not declare potential partners as unacceptable.

[6]It might be thought that it is sufficient to force agents (implicitly or explicitly) to report all potential partners as admissible. This happens in school choice systems in which, to prevent anyone from being excluded, it is assumed that all students find neighborhood schools acceptable (cf., Teo, Sethuraman, and Tan, 2001). However, studying strategy-proofness does not make sense in that type of context, because agents cannot always report their true preferences.

Teo, Sethuraman, and Tan (2001) highlight how discarding the strategic option of remaining single may significantly affects the ability of some agents to manipulate a stable mechanism. For balanced marriage markets, they show that the inability to declare potential partners unacceptable prevents a woman from reaching her best stable partner by manipulating the men-optimal deferred acceptance mechanism. Notice that, this never happens when women are the short side of the market, because any man who is left alone in a stable matching can be used to reach the best stable partner (see the proof of our Theorem 2). Additionally, Tadenuma and Toda (1998) show that a Nash implementable stable mechanism exists in a balanced market if and only if there are only two agents on each side of the market. Our Theorem 1 implies that an analogous property holds when Nash implementability is replaced by strategy-proofness, even when only one side has two agents.

For school choice problems in which institutions and students may not declare potential partners unacceptable, the results of Kesten (2010, Proposition 1) and Kesten and Kurino (2019, Corollary 3) characterize the existence of mechanisms that, from the perspective of students, are strategy-proof and Pareto dominate the student-optimal deferred acceptance mechanism.[7] They show that this type of mechanism exists if and only if the school system has more students than vacancies. We complement this property, as our Theorems 2 and 5 imply that there is a (stable) mechanism that Pareto dominates the student-optimal deferred acceptance mechanism *from the perspective of schools*, and remains strategy-proof for students, if and only if there are as many students as vacancies in the schools system.

The rest of the paper is organized as follows. Section 2 describes our framework and Section 3 motivates our main results by studying the two-agent case. Sections 4 and 5 analyze the validity of Roth (1982, Theorem 3) and Alcalde and Barberá (1994, Proposition 1 and Theorem 3) when (some) agents may not report inadmissibilities. Section 6 shows that each side of the market has an optimal stable mechanism among those that are strategy-proof for the other side. Remarks on topics for future research are included in Section 7. Some proofs are left to the Appendix.

2. Model

We study matching markets in which the mechanism designer knows that agents have no outside options. Hence, no one may misrepresent preferences by declaring potential partners as inadmissible.

Let $[M, W, (\succ_i)_{i\in M\cup W}]$ be a two-sided one-to-one matching markets in which the population is divided into two finite sets, M and W, with at least two agents each. Given $H \in \{M, W\}$, each agent $h \in H$ has a complete, transitive, and strict preference $\succ_h$ defined on H^c, where $M^c \equiv W$ and $W^c \equiv M$. Let $\mathcal{P}$ be the set of preference profiles $\succ = (\succ_i)_{i\in M\cup W}$ satisfying the conditions above.

A *matching* is a function $\mu : M \cup W \to M \cup W$ determining a partner for each agent in $M \cup W$. That is, $\mu(h) \in H^c \cup \{h\}$ and $\mu(\mu(h)) = h$ for each $H \in \{M, W\}$ and $h \in H$. Let $\mathcal{M}$ be the set of matchings between M and W. A matching μ is *stable* when no pair of agents can block it, in the

[7] From the perspective of students, when schools may be declared unacceptable, no strategy-proof mechanism Pareto dominates the student-optimal deferred acceptance mechanism (see Abdulkadiroglu, Pathak, and Roth, 2009; Erdil, 2014).

sense that there is no $(m,w) \in M \times W$ such that $w \succ_m \mu(m)$ and $m \succ_w \mu(w)$. Since there are no outside options, no one is interested in blocking a matching to be alone.

A mechanism is a centralized protocol that associates a matching to each preference profile. Given a mechanism $\Omega : \mathcal{P} \to \mathcal{M}$, consider the following properties:

- Ω is *stable* when for each $\succ \in \mathcal{P}$ the matching $\Omega[\succ]$ is stable in $[M, W, \succ]$.
- Given a side of the market, $H \in \{M, W\}$, Ω is *strategy-proof for* H when for any pair of preference profiles $\succ, \succ' \in \mathcal{P}$ there is no agent $h \in H$ such that

$$\Omega[\succ'_h, \succ_{-h}](h) \succ_h \Omega[\succ](h),$$

  where $\succ_{-h} = (\succ_i)_{i \neq h}$. Ω is *strategy-proof* when it is strategy-proof for M and W.

Therefore, Ω is strategy-proof for H when truth-telling is a dominant strategy for agents in H in the game in which every $i \in M \cup W$ reports preferences $\succ_i$ and $\Omega[(\succ_i)_{i \in M \cup W}]$ is implemented.

Let $\mathrm{DA}_H : \mathcal{P} \to \mathcal{M}$ be the stable mechanism that associates to each preference profile the outcome of the *deferred acceptance algorithm* when agents in $H \in \{M, W\}$ make proposals.

Some classical properties of DA_H will be key to prove our main results:

- $\mathrm{DA}_H[\succ]$ is the best stable matching of $[M, W, \succ]$ for agents in H (Gale and Shapley, 1962).
- $\mathrm{DA}_H : \mathcal{P} \to \mathcal{M}$ is strategy-proof for H (Dubins and Freedman, 1981; Roth, 1982).

We will also appeal to the *Rural Hospital Theorem* (Gale and Sotomayor, 1985): those who are single in a stable matching of $[M, W, \succ]$ remain single in any other stable outcome.

When agents may declare others as unacceptable, Roth (1982) shows that no stable mechanism is strategy-proof, while Alcalde and Barberà (1994) guarantee that DA_H is the only stable mechanism that is strategy-proof for the members of $H \in \{M, W\}$. We will prove that these results may no longer hold when agents have no outside options and the mechanism designer knows it.

Remark 1. If potential partners can be declared unacceptable, then no strategy-proof mechanism is Pareto efficient and individually rational (Alcalde and Barberà, 1994).[8] This result is no longer valid when agents have no outside options and the mechanism designer knows it.

Actually, since individual rationality trivially holds in the absence of outside options, the *serial dictatorship* mechanism $\mathrm{SD}^f_H : \mathcal{P} \to \mathcal{M}$ induced by an order f of the agents in $H \in \{M, W\}$ is Pareto efficient, individually rational, and strategy-proof (cf., Svensson, 1999).[9] □

[8] A mechanism $\Omega : \mathcal{P} \to \mathcal{M}$ is *Pareto efficient* if there is no preference profile $\succ \in \mathcal{P}$ and matching $\mu \in \mathcal{M}$ such that $\mu(i)$ is at least as preferred as $\Omega[\succ](i)$ for each $i \in M \cup W$, and $\mu(h) \succ_h \Omega[\succ](h)$ for some $h \in M \cup W$. The mechanism Ω is *individually rational* when, for each $\succ \in \mathcal{P}$ and $i \in M \cup W$, we have that $\Omega[\succ](i) \succ_i i$ when $\Omega[\succ](i) \neq i$.

[9] Let $f : \{1, \ldots, |H|\} \to H$ be a bijective function. For each $\succ \in \mathcal{P}$, the matching $\mathrm{SD}^f_H[\succ]$ is obtained by the sequential process in which agents in H choose partners in H^c according to the order determined by f. That is, at each stage $t \in \{1, \ldots, |H|\}$, the agent $f(t)$ is matched with the best available partner in H^c, if any.

3. The two-agent case

We will appeal to the simplest two-sided one-to-one matching market to illustrate the effects that the non-existence of outside options has on incentives. When $M = \{m_1, m_2\}$, $W = \{w_1, w_2\}$, and all potential partners are acceptable, the only matchings that could be stable are

$$\mu = \{(m_1, w_1), (m_2, w_2)\}, \qquad \eta = \{(m_1, w_2), (m_2, w_1)\}.$$

In this context, there are four stable mechanism defined in $\mathcal{P}$:

Scenario	Preference Profile				Stable Mechanisms			
	Agents in W		Agents in M		DA_M	Ω_1	Ω_2	DA_W
	$m_1 \succ m_2$	$m_2 \succ m_1$	$w_1 \succ w_2$	$w_2 \succ w_1$				
[1]	w_2	w_1	m_1	m_2	μ		η	
[2]	w_1	w_2	m_2	m_1	η	μ	η	μ
[3]	w_1, w_2	—	m_1, m_2	—	μ			
[4]	—	w_1, w_2	m_1, m_2	—	η			
[5]	w_1	w_2	m_1, m_2	—	μ			
[6]	w_2	w_1	m_1, m_2	—	η			
[7]	w_1, w_2	—	—	m_1, m_2	η			
[8]	—	w_1, w_2	—	m_1, m_2	μ			
[9]	w_1	w_2	—	m_1, m_2	μ			
[10]	w_2	w_1	—	m_1, m_2	η			
[11]	w_1, w_2	—	m_1	m_2	μ			
[12]	—	w_1, w_2	m_1	m_2	μ			
[13]	w_1	w_2	m_1	m_2	μ			
[14]	w_1, w_2	—	m_2	m_1	η			
[15]	—	w_1, w_2	m_2	m_1	η			
[16]	w_2	w_1	m_2	m_1	η			

We claim that Roth's impossibility theorem does not hold in $\mathcal{P}$, because DA_M, Ω_1, Ω_2, and DA_W are strategy-proof. Notice that, since DA_H is stable and strategy-proof for $H \in \{M, W\}$, these four mechanisms are strategy-proof in the subdomain determined by the scenarios [3]-[16].

As a consequence, the only preference profiles in which an agent could misrepresent the preferences are those described in [1] and [2]—because DA_M and DA_W do not coincide—and those in which they can reach the scenarios [1] or [2] through a unilateral deviation. However, since an agent matched with their best alternative never has incentives to misrepresent preferences, it follows that:

- The only cases in which an agent $m \in M$ could misrepresent his preferences are:
  - When Ω_2 or DA_W are implemented and preferences are those of scenario [1].
  - When Ω_1 or DA_W are implemented and preferences are those of scenario [2].

  It is not difficult to verify that a deviation of m would not change his partner in these situations (see scenarios [5], [6], [9], and [10]).
- The only cases in which an agent $w \in W$ could misrepresent her preferences are:
  - When DA_M or Ω_1 are implemented and preferences are those of scenario [1].
  - When DA_M or Ω_2 are implemented and preferences are those of scenario [2].

  However, a deviation of w would not change her partner (see [11], [12], [14], and [15]).

Therefore, DA_M, Ω_1, Ω_2, and DA_W are stable and strategy-proof mechanisms, which in turn implies that they are strategy-proof for both M and W. Hence, Alcalde and Barbera's uniqueness result is not valid in this context. Underlying these results is the fact that $|M| = |W| = 2$ implies that $\min\{|M|, |W|\} = 2$ and $|M| \geq |W|$ (see Theorems 1 and 2).

4. Strategy-proofness and stability without outside options

In this section, we prove that in a two-sided one-to-one matching market in which the mechanism designer knows that agents have no outside options, Roth's impossibility theorem is valid if and only if there are at least three agents on each side of the market. Furthermore, Alcalde and Barberà's uniqueness result only holds for the short side of the market, if there is one.

To understand in which scenarios Roth's impossibility theorem still holds, it is good to describe the main arguments of its proof. Let $\succ \in \mathcal{P}$ be a preference profile such that $\text{DA}_M[\succ]$ and $\text{DA}_W[\succ]$ are the only stable matchings of $[M, W, \succ]$. Given a stable mechanism $\Omega : \mathcal{P} \to \mathcal{M}$, we have that $\Omega[\succ] \in \{\text{DA}_M[\succ], \text{DA}_W[\succ]\}$. As a consequence, if $\Omega[\succ] \neq \text{DA}_H[\succ]$ and everyone may report potential partners as unacceptable, then at least one agent $h \in H$ can manage to improve through an all-or-nothing manipulation strategy in which $\text{DA}_H[\succ](h)$ is reported as the only acceptable partner.[10] Therefore, no stable mechanism is strategy-proof.

Although these manipulation strategies are not feasible within our framework, we will demonstrate that in scenarios analogous to those described above, an agent could benefit by ranking in a different order those who consider them the best alternative. However, in order to succeed in misrepresenting preferences in this way, it will be necessary for M and W to have at least three agents. The following result formalizes these ideas.

Theorem 1. *In a two-sided one-to-one matching market between M and W, there exists a stable and strategy-proof mechanism defined in $\mathcal{P}$ if and only if* $\min\{|M|, |W|\} = 2$.

Proof. Without loss of generality, assume that $\min\{|M|, |W|\} = |W|$.

($\Longrightarrow$) Let $M = \{m_1, \dots, m_r\}$ and $W = \{w_1, \dots, w_s\}$, where $r \geq s$. Suppose that $|W| \geq 3$ and consider a preference profile $\succ = (\succ_i)_{i \in M \cup W} \in \mathcal{P}$ such that:

$\succ_{m_1}$	$\succ_{m_2}$	$\succ_{m_3}$	$\succ_{m_4}$	$\cdots$	$\succ_{m_s}$	$\cdots$	$\succ_{m_r}$	$\succ_{w_1}$	$\succ_{w_2}$	$\succ_{w_3}$	$\succ_{w_4}$	$\cdots$	$\succ_{w_s}$
w_2	w_1	w_1	w_4	$\cdots$	w_s	$\cdots$	w_s	m_1	m_2	m_1	m_4	$\cdots$	m_s
w_1	w_2	w_3	$\vdots$		$\vdots$		$\vdots$	m_2	m_1	m_3	$\vdots$		$\vdots$
w_3	w_3	$\vdots$	$\vdots$		$\vdots$		$\vdots$	m_3	m_3	$\vdots$	$\vdots$		$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$

[10] If $\Omega[\succ](h) \neq \text{DA}_H[\succ](h)$, the matching $\text{DA}_H[\succ]$ continues to be stable in the market in which preferences are given by $\succ' = (\succ'_h, \succ_{-h})$, where $\text{DA}_H[\succ](h) \succ'_h h \succ'_h \cdots$. Therefore, the Rural Hospital Theorem guarantees that h forms a pair with $\text{DA}_H[\succ](h)$ in *any* stable matching of $[M, W, \succ']$.

It follows that $[M, W, (\succ_i)_{i\in M\cup W}]$ has only two stable matchings:

$$\begin{aligned} \mathrm{DA}_M[\succ] &= \{(m_1,w_2),(m_2,w_1),(m_3,w_3),\ldots,(m_s,w_s),m_{s+1},\ldots,m_r\},\\ \mathrm{DA}_W[\succ] &= \{(m_1,w_1),(m_2,w_2),(m_3,w_3),\ldots,(m_s,w_s),m_{s+1},\ldots,m_r\}. \end{aligned}$$

Given any stable mechanism $\Omega : \mathcal{P} \to \mathcal{M}$, we have two alternatives:

- If $\Omega[\succ] = \mathrm{DA}_M[\succ]$, then w_1 can improve her situation by reporting any preference relation satisfying $m_1 \succ'_{w_1} m_3 \succ'_{w_1} m_2 \succ'_{w_1} \cdots$. Indeed, $\mathrm{DA}_W[\succ]$ is the only stable matching of the market $[M, W, (\succ'_{w_1}, \succ_{-w_1})]$ and $\mathrm{DA}_W[\succ](w_1) = m_1 \succ_{w_1} m_2 = \mathrm{DA}_M[\succ](w_1)$.
- If $\Omega[\succ] = \mathrm{DA}_W[\succ]$, then m_1 can improve his situation by reporting any preference relation satisfying $w_2 \succ'_{m_1} w_3 \succ'_{m_1} w_1 \succ'_{m_1} \cdots$. Indeed, $\mathrm{DA}_M[\succ]$ is the only stable matching of the market $[M, W, (\succ'_{m_1}, \succ_{-m_1})]$ and $\mathrm{DA}_M[\succ](m_1) = w_2 \succ_{m_1} w_1 = \mathrm{DA}_W[\succ](m_1)$.

Therefore, no stable mechanism $\Omega : \mathcal{P} \to \mathcal{M}$ is strategy-proof when $\min\{|M|, |W|\} > 2$.

($\Longleftarrow$) We claim that no agent can manipulate the mechanism $\mathrm{DA}_W : \mathcal{P} \to \mathcal{M}$ when $|W| = 2$. Indeed, Dubins and Freedman (1981, Theorem 9) and Roth (1982, Theorem 5) ensure that DA_W is strategy-proof for W. Moreover, given $m \in M$ and $\succ \in \mathcal{P}$ such that $w_i \succ_m w_j$:

- If $\mathrm{DA}_W[\succ](m) = w_i$, then m has no incentives to manipulate DA_W at $\succ$.
- If $\mathrm{DA}_W[\succ](m) = w_j$, then m does not receive a proposal from w_i when DA_W is implemented. Moreover, $\mathrm{DA}_W[\succ'_m, \succ_{-m}](m) = \mathrm{DA}_W[\succ](m)$ when $\succ'_m$ is characterized by $w_j \succ'_m w_i$. Since the mechanism designer knows that m does not have an outside option, he cannot improve his situation by misrepresenting preferences.[11]
- If $\mathrm{DA}_W[\succ](m) = m$, then m does not receive any proposal when DA_W is implemented. This will remain the case regardless of the preferences that he decides to report. Therefore, m cannot improve his situation by misrepresenting preferences.

Hence, DA_W is stable and strategy-proof in the preference domain $\mathcal{P}$ when $|W| = 2$. □

Some remarks are in order:

- The arguments made in the proof of Theorem 1 imply that DA_H is strategy-proof in the preference domain $\mathcal{P}$ provided that $|H| = 2$.
- When $\min\{|M|, |W|\} = 2$, there may exist stable and strategy-proof mechanisms different from DA_M and DA_W. For instance, this happens in the two-agent case (see Section 3).

If an agent $m \in M$ may not declare potential partners as inadmissible, he loses the ability to manipulate a stable mechanism other than DA_M by simply declaring $\mathrm{DA}_M[\succ](m)$ as the *only* acceptable partner. This manipulation strategy is a key ingredient in the classical proof of Alcalde and Barberà's uniqueness result. Thus, the assumption that the mechanism designer knows that there are no outside options may compromise the validity of this result.

[11]This is the only step in the proof of Theorem 1 where we use that agents on the large side of the market may not declare potential partners as unacceptable. Furthermore, this proof works even when agents in the short side have outside options and may declare some potential partners as unacceptable (see Corollary 1).

However, when M is the short side of the market, it is possible to show that DA_M remains the only stable mechanism that is strategy-proof for M. More precisely, when $|M| < |W|$, for each preference profile $\succ \in \mathcal{P}$ there exists an agent $w^* \in W$ that stays alone in all stable matchings of $[M, W, \succ]$. Therefore, when a stable mechanism other than DA_M is implemented, each $m \in M$ can report $\text{DA}_M[\succ](m)$ and w^* as the best alternatives, in this order. This manipulation strategy ensures that $\text{DA}_M[\succ]$ remains stable under the new preferences and, as a consequence of the Rural Hospital Theorem, w^* stands alone. In addition, m never forms a stable pair with someone less preferred to w^*, because w^* always considers him acceptable. Hence, by misrepresenting preferences in this way, m manage to pair with $\text{DA}_M[\succ](m)$—the best partner he would get in a stable matching. Therefore, no stable mechanism other than DA_M is strategy-proof for M.

Nevertheless, when M is the long side of the market, the scarcity of potential partners in W will allow the existence of many stable mechanisms defined on $\mathcal{P}$ that are strategy-proof for M. Essentially, we will manage to obtain a family of mechanisms with these properties by properly combining DA_M and DA_W. To gain some intuition about this claim, consider a preference profile in which agents in W have heterogeneous best partners. In this scenario, agents in W do not compete when the DA_W is implemented, and the inability of agents in M to declare others unacceptable makes it impossible for them to improve through misrepresenting preferences (cf., Teo, Sethuraman, and Tan, 2001). The next result formalizes the ideas described above.

Theorem 2. *In a two-sided one-to-one matching market between M and W, there are many stable mechanisms defined in $\mathcal{P}$ that are strategy-proof for M if and only if $|M| \geq |W|$.*

Proof. ($\Longleftarrow$) Let $M = \{m_1, \ldots, m_r\}$ and $W = \{w_1, \ldots, w_s\}$. Suppose that $r \geq s$ and consider a preference profile $\succ^* \in \mathcal{P}$ characterized by:

- Given $i \in \{1, 2\}$, m_i considers w_i the best potential partner.
- Given $i, j \in \{1, 2\}$ with $i \neq j$, w_i considers m_j her best potential partner.
- Given $i \in \{3, \ldots, s\}$, agents m_i y w_i consider each other the best alternative.
- The agents $m_1, \ldots, m_s$ are the top alternatives for each $w \in W$.

It follows that the two-sided market $[M, W. \succ^*]$ has only two stable matchings:

$$\begin{aligned}\text{DA}_M[\succ^*] &= \{(m_1, w_1), (m_2, w_2), (m_3, w_3), \ldots, (m_s, w_s), m_{s+1}, \ldots, m_r\},\\ \text{DA}_W[\succ^*] &= \{(m_1, w_2), (m_2, w_1), (m_3, w_3), \ldots, (m_s, w_s), m_{s+1}, \ldots, m_r\}.\end{aligned}$$

Let $\overline{M} = \{m_3, \ldots, m_s\}$ and $K = \left\{\succ \in \mathcal{P} : (\succ_i)_{i \in \overline{M} \cup W} = (\succ^*_i)_{i \in \overline{M} \cup W}\right\}$.

Consider the mechanism $\Omega : \mathcal{P} \to \mathcal{M}$ defined by

$$\Omega[\succ] = \begin{cases} \text{DA}_W[\succ], & \text{when } \succ \in K, \\ \text{DA}_M[\succ], & \text{when } \succ \notin K. \end{cases}$$

Since $\Omega[\succ^*] = \text{DA}_W[\succ^*] \neq \text{DA}_M[\succ^*]$, the stable mechanisms Ω and DA_M are different.

We claim that Ω is strategy-proof for M. By contradiction, suppose that there exists $m \in M$ and $\succ, \succ' \in \mathcal{P}$ such that $\Omega[\succ'_m, \succ_{-m}](m) \succ_m \Omega[\succ](m)$.

There are two relevant cases to analyze:

- Suppose that $\succ \in K$. Since each agent in $\overline{M}$ is matched with his best alternative in $\text{DA}_W[\succ]$, it follows that $m \notin \overline{M}$. Hence, $(\succ'_m, \succ_{-m}) \in K$. Moreover, as no one has an outside option and $(\succ_w)_{w\in W} = (\succ^*_w)_{w\in W}$, the deferred acceptance algorithm DA_W finishes with the pairs formed at the first step when it is applied to either $(\succ'_m, \succ_{-m})$ or $\succ$.[12] Therefore, $\Omega[\succ'_m, \succ_{-m}] = \text{DA}_W[\succ'_m, \succ_{-m}] = \text{DA}_W[\succ] = \Omega[\succ]$, a contradiction.
- Suppose that $\succ \notin K$. Since DA_M is strategy-proof for M, we have that $(\succ'_m, \succ_{-m}) \in K$. Thus, $(\succ_w)_{w\in W} = (\succ^*_w)_{w\in W}$ and $m = m_i$ for some $i \in \{3, \dots, s\}$, which imply that $w_i = \text{DA}_W[(\succ'_m, \succ_{-m})](m) = \Omega[(\succ'_m, \succ_{-m})](m) \succ_m \Omega[\succ](m) = \text{DA}_M[(\succ)](m)$. Therefore, although w_i considers m the best alternative under $\succ_{w_i}$, she rejects his proposal when deferred acceptance is applied to $\succ$. A contradiction.

($\Longrightarrow$) Assuming that $|M| < |W|$, we want to prove that there is a single stable mechanism defined in $\mathcal{P}$ that is strategy-proof for M. By contradiction, assume that there exists $\Omega : \mathcal{P} \to \mathcal{M}$, stable and strategy-proof for M, satisfying $\Omega[\succ] \neq \text{DA}_M[\succ]$ for some $\succ \in \mathcal{P}$.

Since DA_M is the M-optimal stable mechanism (Gale and Shapley, 1962; Theorem 2), it follows that $\text{DA}_M[\succ](m) \succ_m \Omega[\succ](m)$ for some $m \in M$. In particular, $\text{DA}_M[\succ](m) \in W$. Furthermore, $|M| < |W|$ implies that there exists $w^* \in W$ such that $\text{DA}_M[\succ](w^*) = w^*$.

Let $\succ'_m$ be a preference profile in which $\text{DA}_M[\succ](m)$ and w^* are the best alternatives for m, in this order. Since the matching $\text{DA}_M[\succ]$ is stable under $(\succ'_m, \succ_{-m})$, it follows from the Rural Hospital Theorem (Gale and Sotomayor, 1985, Theorem 1) that w^* is alone in every stable matching of the market $[M, W, (\succ'_m, \succ_{-m})]$. This implies that $\Omega[\succ'_m, \succ_{-m}](m) \succ'_m w^*$, because otherwise (m, w^*) would block the stable matching $\Omega[\succ'_m, \succ_{-m}]$ (remember that m is always acceptable for w^*).[13] As a consequence, $\Omega[\succ'_m, \succ_{-m}](m) = \text{DA}_M[\succ](m)$, which implies that $\Omega[\succ'_m, \succ_{-m}](m) \succ_m \Omega[\succ](m)$. This contradicts the strategy-proofness of Ω. $\square$

Assuming that $|M| \geq |W|$, the arguments made in the proof of Theorem 2 ensure that the stable mechanism $\Omega : \mathcal{P} \to \mathcal{M}$ defined by

$$\Omega[\succ] = \begin{cases} \text{DA}_W[\succ], & \text{when } \succ \in K, \\ \text{DA}_M[\succ], & \text{when } \succ \notin K, \end{cases}$$

is strategy-proof for M when the subdomain of preferences K is characterized by the existence of a set $\overline{M} \subseteq M$ that satisfies the following conditions:

- For each $\succ \in K$, $\text{DA}_W[\succ](h)$ is the best alternative for $h \in \overline{M} \cup W$.
- Agents in $M \setminus \overline{M}$ have unrestricted preferences in K.

Notice that, to ensure that $\Omega \neq \text{DA}_M$ it is necessary that $\text{DA}_W[\succ] \neq \text{DA}_M[\succ]$ for some $\succ \in K$, which requires that $|M \setminus \overline{M}| \geq 2$.[14]

[12] The arguments made up to this point remain valid when there are at least two agents in M who may not declare inadmissibilities, even when everyone in W may consider some potential partners unacceptable (see Theorem 4).

[13] This is the only step in the proof in which we use that agents in W may not declare others unacceptable.

[14] To construct the set K it is enough to choose two agents from M to leave out of $\overline{M}$. Therefore, when $|M| \geq |W|$, we can construct at least $|M|(|M| - 1)/2$ stable mechanisms defined in $\mathcal{P}$ which are strategy-proof for M.

5. Extensions

In this section, we discuss the validity of the classical results of Roth (1982, Theorem 3) and Alcalde and Barberà (1994, Proposition 1 and Theorem 3) in scenarios in which some agents may declare potential partners as unacceptable.

In a two-sided one-to-one matching market between M and W, denote by $H_{\otimes}$ the subset of agents in $H \in \{M, W\}$ that always consider all potential partners acceptable. We assume that the identity of the agents in $H_{\otimes}$ is known by the mechanism designer and, therefore, they never misrepresent preferences by declaring some potential partners inadmissible. Let $\mathcal{Q}(M_{\otimes}, W_{\otimes})$ be the set of preference profiles $(\succ_i)_{i \in M \cup W}$ such that, given $H \in \{M, W\}$:

- For each $h \in H_{\otimes}$, $\succ_h$ is a complete, transitive, and strict preference defined on H^c.
- For each $h \in H \setminus H_{\otimes}$, $\succ_h$ is a complete, transitive, and strict preference defined on $H^c \cup \{h\}$.

Notice that, $\mathcal{P} = \mathcal{Q}(M, W) \subseteq \mathcal{Q}(M_{\otimes}, W_{\otimes})$ for all $M_{\otimes} \subseteq M$ and $W_{\otimes} \subseteq W$.

Given $\succ \in \mathcal{Q}(M_{\otimes}, W_{\otimes})$, a matching $\mu \in \mathcal{M}$ is *stable* under $\succ$ when the following properties hold:

- *Individual rationality*: there is no $i \in (M \setminus M_{\otimes}) \cup (W \setminus W_{\otimes})$ such that $i \succ_i \mu(i)$.
- There is no $(m, w) \in M \times W$ such that $w \succ_m \mu(m)$ and $m \succ_w \mu(w)$.

If $\mathcal{Q} \equiv \mathcal{Q}(M_{\otimes}, W_{\otimes})$, denote by $\mathbb{S}_H(\mathcal{Q})$ the non-empty set of stable mechanisms $\Omega : \mathcal{Q} \to \mathcal{M}$ that are strategy-proof for agents in $H \in \{M, W\}$.

Remark 2. Given $M_{\otimes} \subseteq M$ and $W_{\otimes} \subseteq W$, it is well-known that the serial dictatorship mechanism $\mathrm{SD}_M^f : \mathcal{Q}(M_{\otimes}, W_{\otimes}) \to \mathcal{M}$ induced by an order f of agents in M is Pareto efficient, strategy-proof, and individually rational for the agents in M (cf., Svensson, 1999). Moreover, SD_M^f is individually rational for the agents in W as long as $W_{\otimes} = W$. Therefore, if the mechanism designer knows that all agents on one side of the market have no outside options, then the impossibility result of Alcalde and Barberà (1994, Proposition 1) does not hold. □

When all agents may declare some potential partners as unacceptable (i.e. $M_{\otimes} \cup W_{\otimes} = \emptyset$), the impossibility result of Roth (1982, Theorem 3) states that $\mathbb{S}_M(\mathcal{Q}) \cap \mathbb{S}_W(\mathcal{Q}) = \emptyset$.

Theorem 3. *In a two-sided one-to-one matching market between M and W, we have that:*

(i) *If $M_{\otimes} = M$ and $|W| = 2$, then $\mathrm{DA}_W \in \mathbb{S}_M(\mathcal{Q}) \cap \mathbb{S}_W(\mathcal{Q})$.*
(ii) *If $M_{\otimes} = M$, $|W| = 2$, and $|W \setminus W_{\otimes}| \geq 1$, then $\mathbb{S}_M(\mathcal{Q}) \cap \mathbb{S}_W(\mathcal{Q}) = \{\mathrm{DA}_W\}$.*
(iii) *If $\min\{|M|, |W|\} > 2$ or $\max\{|M_{\otimes}|, |W_{\otimes}|\} < 2$, then $\mathbb{S}_M(\mathcal{Q}) \cap \mathbb{S}_W(\mathcal{Q}) = \emptyset$.*

The proof is given in the Appendix.

When W has only two agents, the property (i) implies that the absence of outside options for agents in M has an important implication: DA_W becomes strategy-proof for M. In addition, if at least one agent of W may declare inadmissibilities, then DA_W is the only mechanism that is strategy-proof for the whole population (property (ii)).

Moreover, as a direct consequence of properties (i) and (iii) we can extend the result of Theorem 1 to ensure that, even when agents on the short side of the market may declare others as unacceptable, a stable and strategy-proof mechanism exists if and only if $\min\{|M|,|W|\}=2$.

Corollary 1. *When $|M|\geq|W|$ and $M_\otimes=M$, there exists a stable and strategy-proof mechanism defined in $\mathcal{Q}(M_\otimes,W_\otimes)$ if and only if W has only two agents.*

It also follows form Theorem 3(iii) that Roth's impossibly theorem holds when either no side of the market has only two agents or there is at most one agent on each side with no outside option.

In terms of our notation, when all agents may declare potential partners as unacceptable, the uniqueness result of Alcalde and Barberà (1994, Theorem 3) shows that $|\mathbb{S}_M(\mathcal{Q})|=|\mathbb{S}_W(\mathcal{Q})|=1$.

Theorem 4. *In a two-sided one-to-one matching market between M and W, we have that:*

(a) *If $|M_\otimes|<2$, then DA_M is the only mechanism in $\mathbb{S}_M(\mathcal{Q})$.*
(b) *If $|M|<|W|$, $|M_\otimes|\geq 2$, and $|W_\otimes|\leq|M|$, then $|\mathbb{S}_M(\mathcal{Q})|>1$.*
(c) *If $|M|\geq|W|$ and $|M_\otimes|\geq 2$, then $|\mathbb{S}_M(\mathcal{Q})|>1$.*
(d) *If $|M|<|W|$ and $W_\otimes=W$, then DA_M is the only mechanism in $\mathbb{S}_M(\mathcal{Q})$.*

The proof is given in the Appendix.

The property (a) ensures that, independently of the characteristics of the agents in W, the existence of many stable mechanisms that are strategy-proof for M requires that at least two agents in M have no outside options. The property (b) determines sufficient conditions for the existence of many stable mechanisms that are strategy-proof for the *short side* of the market, a situation that never arises when no one has outside options (see Theorem 2). Essentially, assuming that $|M|<|W|$, if there are at most $|M|$ agents in W without outside options, the arguments in the proof of Theorem 4 ensure that the stable mechanism $\Omega:\mathcal{P}\to\mathcal{M}$ defined by

$$\Omega[\succ]=\begin{cases}\mathrm{DA}_W[\succ], & \text{when } \succ\in K,\\ \mathrm{DA}_M[\succ], & \text{when } \succ\notin K,\end{cases}$$

is strategy-proof for M provided that the subdomain of preferences K satisfies the following properties for some sets $\overline{M}\subseteq M$ and $\overline{W}\subseteq W$:

- For each $\succ\in K$, $\mathrm{DA}_W[\succ](h)$ is the best alternative for $h\in\overline{M}\cup\overline{W}$.
- Agents in $M\setminus\overline{M}$ have unrestricted preferences in K and belong to $M_\otimes$.
- For each preference profile $\succ\in K$, all agents in $W\setminus\overline{W}$ consider those in $M\setminus\overline{M}$ unacceptable and $\mathrm{DA}_W[\succ](M)=\overline{W}$.[15]

To guarantee that Ω is different from DA_M, it is key that $|M\setminus\overline{M}|\geq 2$. This is the only reason why the statement of property (b) requires that at least two agents in M have no outside options.

[15] Since these properties imply that $|W\setminus\overline{W}|\leq|W\setminus W_\otimes|$ and $|M|=|\overline{W}|$, they require that $|W_\otimes|\leq|M|$.

Remark 3. Within the family of preference domains $\{\mathcal{Q}(M_\otimes, W_\otimes) : M_\otimes \subseteq M,\ W_\otimes \subseteq W\}$, a set $\mathcal{Q}(\overline{M}_\otimes, \overline{W}_\otimes)$ is a *maximal domain* in which Alcalde and Barberá's uniqueness result does not hold if and only if $|\overline{M}_\otimes| = 2$ and $\overline{W}_\otimes = \emptyset$ (see properties (a)-(c) of Theorem 4). In other words, to ensure that many stable mechanisms are strategy-proof for M, at least two members of M must be unable to declare their potential partners as unacceptable. □

Notice that, properties (c)-(d) of Theorem 4 imply that the result of Theorem 2 holds when everyone in W and at least two agents in M have no outside options.

Corollary 2. *When $W_\otimes = W$ and $|M_\otimes| \geq 2$, there are many stable mechanisms defined in $\mathcal{Q}(M_\otimes, W_\otimes)$ that are strategy-proof for M if and only if $|M| \geq |W|$.*

The next corollary is a direct consequence of properties (a), (b), and (c) of Theorem 4.

Corollary 3. *When either $W_\otimes = \emptyset$ or $|M| \geq |W|$, there are many stable mechanisms defined in $\mathcal{Q}(M_\otimes, W_\otimes)$ that are strategy-proof for M if and only if $|M_\otimes| \geq 2$.*

This result has implications for *college admission problems*, namely many-to-one matching problems in which institutions may declare some students inadmissible (cf., Balinski and Sönmez, 1999; Abdulkadiroğlu and Sönmez, 2003). Indeed, when student priorities are determined by strict rankings, a college admission problem can be viewed as a one-to-one matching market by identifying each institution with a group of representatives, as many as the number of vacancies (cf., Gale and Sotomayor, 1985). Thus, if at least two students never declare a college unacceptable, it follows from Corollary 3 that there exists a stable mechanism that all colleges prefer to DA_M and that is strategy-proof for students.

6. On W-optimal mechanisms in $\mathbb{S}_M(\mathcal{Q})$

In this section, we prove the existence of a stable mechanism that is optimal for all agents in W among those in which agents in M do not have incentives to misrepresent preferences.

Given a side of the market $H \in \{M, W\}$, consider the partial order $\geqslant_H$ defined over the family of mechanisms $\Omega : \mathcal{Q} \to \mathcal{M}$, where $\mathcal{Q} \equiv \mathcal{Q}(M_\otimes, W_\otimes)$, and characterized by

$$\Omega_1 \geqslant_H \Omega_2 \qquad \Longleftrightarrow \qquad \Omega_1[\succ](h) \succeq_h \Omega_2[\succ](h), \quad \forall h \in H, \forall \succ \in \mathcal{Q}.$$

Hence, $\Omega_1 \geqslant_H \Omega_2$ if and only if, regardless of preferences in $\mathcal{Q}$, all agents in H weakly prefer the outcome of Ω_1 to that of Ω_2. As usual, $\Omega_1 >_H \Omega_2$ indicates that $\Omega_1 \geqslant_H \Omega_2$ and $\Omega_1 \neq \Omega_2$. Notice that, $\Omega_1 >_H \Omega_2$ if and only if Ω_1 is Pareto-superior to Ω_2 for agents in H.

A mechanism $\Omega_W : \mathcal{Q} \to \mathcal{M}$ is *W-optimal in* $\mathbb{S}_M(\mathcal{Q})$ when it is the greatest element of $\mathbb{S}_M(\mathcal{Q})$ under the partial order $\geqslant_W$. That is, from the point of view of agents in W, the mechanism Ω_W is Pareto-superior to any other mechanism in $\mathbb{S}_M(\mathcal{Q})$. Since agents in W compete with each other, the existence of such a mechanism seems non-trivial.

Theorem 5. *In a two-sided one-to-one matching market between M and W, there always exists a mechanism $\Omega_W : \mathcal{Q} \to \mathcal{M}$ that is W-optimal in $\mathbb{S}_M(\mathcal{Q})$. Moreover, the following properties hold:*

(i) *Each agent in $M \setminus M_\otimes$ has the same partner in all mechanisms in $\mathbb{S}_M(\mathcal{Q})$.*

(ii) *When agents have no outside options, we have that:*

$$\Omega_W \neq \mathrm{DA}_M \iff |M| \geq |W|, \qquad \Omega_W \neq \mathrm{DA}_W \iff |W| > 2.$$

(iii) *Ω_W differs from DA_M and DA_W as long as $|M_\otimes| \geq 2$ and $|M| \geq |W| > 2$.*

To prove this result, we will show that the set of stable and one-side strategy-proof *mechanisms* is a lattice under the same binary operations for which Knuth (1976) shows that the collection of stable *matchings* has this property (see the Appendix).

The property (i) of Theorem 5 formalizes the intuition that agents without outside options are the only ones that can be negatively affected when a mechanism $\Omega \in \mathbb{S}_M(\mathcal{Q})$ different from DA_M is implemented.[16] We conclude with a result that is obtained from property (ii) of Theorem 5 and Kesten and Kurino (2019, Corollary 3).

Corollary 4. *When $|M| > |W|$, for each $H \in \{M, W\}$ there exists a mechanism $\Omega : \mathcal{P} \to \mathcal{M}$ that is strategy-proof for M and satisfies $\Omega >_H \mathrm{DA}_M$.*

7. Concluding remarks

In two-sided one-to-one matching markets, assuming that some agents may not misrepresent preferences by declaring potential partners unacceptable, we have found necessary and sufficient conditions to ensure either the existence of a stable and strategy-proof mechanism or the multiplicity of stable mechanisms that are strategy-proof for one side of the market. We have also shown that, among the stable mechanisms that are strategy-proof for one side of the market, there is one that is optimal for the other side. In many contexts, which we describe in terms of the relative size of the sides of the market, this optimal mechanism does not coincide with any version of deferred acceptance. The problem of fully characterizing this optimal mechanism is open.

A natural extension of our results is to the context of *matching markets with contracts*, that is, many-to-one matching markets between *hospitals* and *doctors* in which side payments are allowed. Hatfield and Milgrom (2005) show that many of the mechanism design properties of one-to-one matching markets remain valid in this context as long as doctors are *substitutes* and the *law of aggregate demand* holds.[17] Moreover, Hatfield and Milgrom (2005), Sakai (2011), and Hirata and Kasuya (2017) show that Roth's impossibility theorem and Alcalde and Barberà's uniqueness result hold when agents may declare others unacceptable. Although this is a topic for future research, we believe that our results should be adaptable to this more general framework.

[16] The advantages of having an outside option can reach the point of inducing agents to always prefer a manipulable mechanism over a strategy-proof one (see Akbarpour, Kapor, Neilson, Van Dijk, Zimmerman, 2022).

[17] Doctors are substitutes when any contract that is selected from a set of alternatives continues to be chosen when some of those alternatives are no longer available. Hospital's preferences satisfy the law of aggregate demand whenever the number of contracts chosen does not decrease as the set of alternatives expands.

Appendix: omitted proofs

Proof of Theorem 3.
(i) Arguments identical to those presented in the second part of the proof of Theorem 1 can be applied to show this result (see footnote 11).

(ii) Suppose that $M_\otimes = M$, $|W| = 2$, and $|W \setminus W_\otimes| \geq 1$. It follows from the property (i) that DA_W belongs to $\mathbb{S}_M(\mathcal{Q}) \cap \mathbb{S}_W(\mathcal{Q})$. Moreover, since $|W| = 2$ and $|W \setminus W_\otimes| \geq 1$ ensure that $|W_\otimes| < 1$, the property (a) of Theorem 4 implies that DA_W is the only mechanism in $\mathbb{S}_W(\mathcal{Q})$. Therefore, $\mathbb{S}_M(\mathcal{Q}) \cap \mathbb{S}_W(\mathcal{Q}) = \{\mathrm{DA}_W\}$.

(iii) Arguments identical to those presented in the first part of the proof of Theorem 1 can be applied to show that $\min\{|M|, |W|\} > 2$ implies that $\mathbb{S}_M(\mathcal{Q}) \cap \mathbb{S}_W(\mathcal{Q}) = \emptyset$. On the other hand, given $H \in \{M, W\}$, it follows from Theorem 4(a) that $|H_\otimes| < 2$ implies $\mathbb{S}_H(\mathcal{Q}) = \{\mathrm{DA}_H\}$. As a consequence, when $\max\{|M_\otimes|, |W_\otimes|\} < 2$, we have that $\mathbb{S}_M(\mathcal{Q}) \cap \mathbb{S}_W(\mathcal{Q}) = \emptyset$, because $\mathrm{DA}_M \neq \mathrm{DA}_W$. □

Proof of Theorem 4.
(a) Suppose that $M_\otimes = \{\overline{m}\}$.[18] By contradiction, assume that there is $\Omega : \mathcal{Q} \to \mathcal{M}$ different from DA_M that is stable and strategy-proof for M. Let $\succ \in \mathcal{Q}$ such that $\Omega[\succ] \neq \mathrm{DA}_M[\succ]$. Since agents in M consider the matching $\mathrm{DA}_M[\succ]$ the best stable outcome in $[M, W, \succ]$, for each $m \in M$ we have $\mathrm{DA}_M[\succ](m) \succ_m \Omega[\succ](m)$ or $\mathrm{DA}_M[\succ](m) = \Omega[\succ](m)$. Notice that, if there exists $m \neq \overline{m}$ such that $\mathrm{DA}_M[\succ](m) \succ_m \Omega[\succ](m)$, then $\mathrm{DA}_M[\succ](m) \in W$ and m can manipulate the mechanism Ω by reporting any preference $\succ'_m$ such that $\mathrm{DA}_M[\succ](m) \succ'_m m \succ'_m \cdots$. Indeed, since $\mathrm{DA}_M[\succ]$ is stable in $[M, W, (\succ'_m, \succ_{-m})]$, it follows from the Rural Hospital Theorem that $\Omega[\succ'_m, \succ_{-m}](m) = \mathrm{DA}_M[\succ](m)$, which implies that $\Omega[\succ'_m, \succ_{-m}] \succ_m \Omega[\succ](m)$. This contradicts the strategy-proofness of Ω. Therefore, $\overline{w} \equiv \mathrm{DA}_M[\succ](\overline{m}) \succ_{\overline{m}} \Omega[\succ](\overline{m})$. Furthermore, as $\Omega[\succ](m) = \mathrm{DA}_M[\succ](m)$ for all $m \neq \overline{m}$, we have that $\Omega[\succ](\overline{w}) = \overline{w}$.[19] Therefore, $(\overline{m}, \overline{w})$ blocks the matching $\Omega[\succ]$ under $\succ$. A contradiction.

(b) We want to prove that $|\mathbb{S}_M(\mathcal{Q})| > 1$ whenever $|M| < |W|$, $|M_\otimes| \geq 2$, and $|W_\otimes| \leq |M|$.

Suppose that $M = \{m_1, \ldots, m_r\}$ and $W = \{w_1, \ldots, w_s\}$, where $2 \leq r < s$. Also, $\{m_1, m_2\} \subseteq M_\otimes$ and $W_\otimes \subseteq \{w_1, \ldots, w_r\}$. Let $\widetilde{\succ} \in \mathcal{Q}$ be a preference profile that satisfies the following conditions:

1. Given $i \in \{1, 2\}$, m_i considers w_i the best potential partner
2. Given $i, j \in \{1, 2\}$ with $i \neq j$, w_i considers m_j her best potential partner.
3. For each $i \in \{3, \ldots, r\}$, agents m_i y w_i consider each other the best alternative.
4. Each agent $w \in \{w_{r+1}, \ldots, w_s\}$ considers m_1 and m_2 unacceptable.

It follows that the market $[M, W, \widetilde{\succ}]$ has only two stable matchings:

$$\begin{aligned}
\mathrm{DA}_M[\widetilde{\succ}] &= \{(m_1, w_1), (m_2, w_2), (m_3, w_3), \ldots, (m_r, w_r), w_{r+1}, \ldots, w_s\}, \\
\mathrm{DA}_W[\widetilde{\succ}] &= \{(m_1, w_2), (m_2, w_1), (m_3, w_3), \ldots, (m_r, w_r), w_{r+1}, \ldots, w_s\}.
\end{aligned}$$

Define $\overline{M} = M \setminus \{m_1, m_2\}$ and $K = \left\{\succ \in \mathcal{Q} : (\succ_i)_{i \in \overline{M} \cup W} = (\widetilde{\succ}_i)_{i \in \overline{M} \cup W}\right\}$. Let $\Omega : \mathcal{Q} \to \mathcal{M}$ be the mechanism such that

$$\Omega[\succ] = \begin{cases} \mathrm{DA}_W[\succ], & \text{when } \succ \in K, \\ \mathrm{DA}_M[\succ], & \text{when } \succ \notin K. \end{cases}$$

[18] When $M_\otimes = \emptyset$, Alcalde and Barberà (1994, Theorem 3) guarantee that DA_M is the only mechanism in $\mathbb{S}_M(\mathcal{Q})$.

[19] This argument also contradicts the Rural Hospital Theorem, because $\overline{w} = \mathrm{DA}_M[\succ](\overline{m})$ is left alone in $\Omega[\succ]$.

Since $\Omega[\widetilde{\succ}] = \text{DA}_W[\widetilde{\succ}] \neq \text{DA}_M[\widetilde{\succ}]$, Ω is different from DA_M. We claim that Ω is strategy-proof for M. By contradiction, suppose that there are $m \in M$ and $\succ, \succ' \in \mathcal{Q}$ such that $\Omega[\succ'_m, \succ_{-m}](m) \succ_m \Omega[\succ](m)$. In this context, there are two relevant cases:

- Suppose that $\succ \in K$. Since each agent in $\overline{M}$ is matched with his best alternative in $\text{DA}_W[\succ]$, it follows that $m \in \{m_1, m_2\}$. Hence, $(\succ'_m, \succ_{-m}) \in K$. This implies that $(\succ_w)_{w\in W} = (\widetilde{\succ}_w)_{w \in W}$ and condition (4) ensures that the deferred acceptance algorithm DA_W finishes with the pairs that are formed at the first step when it is applied to either $(\succ'_m, \succ_{-m})$ or $\succ$. We conclude that $\Omega[\succ'_m, \succ_{-m}] = \text{DA}_W[\succ'_m, \succ_{-m}] = \text{DA}_W[\succ] = \Omega[\succ]$. A contradiction.
- Suppose that $\succ \notin K$. Since DA_M is strategy-proof for M in the preference domain $\mathcal{Q}$, we have that $(\succ'_m, \succ_{-m}) \in K$. Therefore, $(\succ_w)_{w\in W} = (\widetilde{\succ}_w)_{w\in W}$ and $m = m_i$ for some $i \in \{3, \ldots, r\}$, which imply that $w_i = \text{DA}_W[(\succ'_m, \succ_{-m})](m) = \Omega[(\succ'_m, \succ_{-m})](m) \succ_m \Omega[\succ](m) = \text{DA}_M[(\succ)](m)$. Therefore, although w_i considers m the best alternative under $\succ_{w_i}$, she rejects his proposal when deferred acceptance is applied to $\succ$. A contradiction.

We conclude that there are many stable mechanisms defined in $\mathcal{Q}$ that are strategy-proof for M.

(c) + (d) The proof of Theorem 2 works to show these properties (see footnotes 12 and 13). □

Proof of Theorem 5.

Given $\Omega_1, \Omega_2 \in \mathbb{S}_M(\mathcal{Q})$, let $\vee$ and $\wedge$ be the binary operations such that, for each agent $m \in M$ and preference profile $\succ \in \mathcal{Q}$, we have that

$$\Omega_1 \vee \Omega_2[\succ](m) = \max_{\succ_m}\{\Omega_1[\succ](m), \Omega_2[\succ](m)\}, \qquad \Omega_1 \wedge \Omega_2[\succ](m) = \min_{\succ_m}\{\Omega_1[\succ](m), \Omega_2[\succ](m)\},$$

where $\max_{\succ_m}\{w, w'\} = w$ if and only if either $w \succ_m w'$ or $w = w'$, and $\min_{\succ_m}\{w, w'\} = w$ if and only if either $w' \succ_m w$ or $w = w'$. Since the stable matchings of $[M, W, \succ]$ form a lattice under the binary operations $\vee$ and $\wedge$ (Knuth, 1976; Theorem 7, attributed to J. H. Conway), the mechanisms $\Omega_1 \vee \Omega_2$ and $\Omega_1 \wedge \Omega_2$ are well-defined and stable. We claim that $\Omega_1 \vee \Omega_2$ and $\Omega_1 \wedge \Omega_2$ belong to $\mathbb{S}_M(\mathcal{Q})$. Indeed, by contradiction, suppose that there exist $m \in M$ and $\succ, \succ' \in \mathcal{Q}$ such that $\Omega_1 \vee \Omega_2[\succ'_m, \succ_{-m}](m) \succ_m \Omega_1 \vee \Omega_2[\succ](m)$. It follows from the definition of $\vee$ that $\Omega_1 \vee \Omega_2[\succ'_m, \succ_{-m}](m) = \Omega_i[\succ'_m, \succ_{-m}](m)$ for some $i \in \{1, 2\}$, which implies that $\Omega_i[\succ'_m, \succ_{-m}](m) \succ_m \Omega_i[\succ](m)$. This contradicts the fact that $\Omega_i \in \mathbb{S}_M(\mathcal{Q})$. Analogously, if $\Omega_1 \wedge \Omega_2[\succ'_m, \succ_{-m}](m) \succ_m \Omega_1 \wedge \Omega_2[\succ](m)$, then $\Omega_1 \wedge \Omega_2[\succ](m) = \Omega_j[\succ](m)$ for some $j \in \{1, 2\}$. Hence, $\Omega_j[\succ'_m, \succ_{-m}](m) \succ_m \Omega_j[\succ](m)$, which contradicts the fact that $\Omega_j \in \mathbb{S}_M(\mathcal{Q})$.

Therefore, $(\mathbb{S}_M(\mathcal{Q}), \vee, \wedge)$ is a lattice. Notice that, $\Omega_1 \geqslant_M \Omega_2$ if and only if $\Omega_1 = \Omega_1 \vee \Omega_2$. Since agents from different sides of the market have opposed preferences for stable matchings (Knuth, 1976; Corollary 1), it follows that $\Omega_1 \geqslant_W \Omega_2$ if and only if $\Omega_2 \geqslant_M \Omega_1$. This relationship implies that a mechanism is W-optimal in $\mathbb{S}_M(\mathcal{Q})$ if and only if it is the least element of $\mathbb{S}_M(\mathcal{Q})$ under the partial order $\geqslant_M$. Since every finite lattice has a least element, we conclude that there is $\Omega_W : \mathcal{Q} \to \mathcal{M}$ that is W-optimal in $\mathbb{S}_M(\mathcal{Q})$.

(i) By contradiction, suppose that there exists $m \in M \setminus M_\otimes$ such that $\Omega[\succ](m) \neq \text{DA}_M[\succ](m)$ for some preference profile $\succ \in \mathcal{Q}$ and mechanism $\Omega \in \mathbb{S}_M(\mathcal{Q})$. Since $\text{DA}_M[\succ]$ is the best stable matching of $[M, W, \succ]$ for agents in M, $w \equiv \text{DA}_M[\succ](m) \succ_m \Omega[\succ](m)$. Let $\succ'_m$ be a preference relation defined on $W \cup \{m\}$ such that $w \succ'_m m \succ'_m \cdots$ (remember that m may declare others unacceptable). Since $\text{DA}_M[\succ]$ is stable under $(\succ'_m, \succ_{-m})$ and w is the only acceptable partner of the agent m under $\succ'_m$, the Rural Hospital Theorem implies that $\Omega(\succ'_m, \succ_{-m})(m) = w$. Hence, $\Omega(\succ'_m, \succ_{-m})(m) \succ_m \Omega[\succ](m)$, which contradicts the strategy-proofness of Ω.

(ii) Suppose that agents have no outside options (i.e., $M_\otimes = M$ and $W_\otimes = W$). In this case, the preference domain $\mathcal{Q} \equiv \mathcal{Q}(M_\otimes, W_\otimes)$ coincides with $\mathcal{P}$.

- For any $\succ \in \mathcal{P}$, $\mathrm{DA}_M[\succ]$ is the best stable matching for agents in M. Hence, $\Omega >_W \mathrm{DA}_M$ for all $\Omega \in \mathbb{S}_M(\mathcal{P}) \setminus \{\mathrm{DA}_M\}$, which implies that $\Omega_W \neq \mathrm{DA}_M$ if and only if $|\mathbb{S}_M(\mathcal{P})| > 1$. Therefore, it follows form Theorem 2 that $\Omega_W \neq \mathrm{DA}_M$ if and only if $|M| \geq |W|$.
- Since DA_W implements the best stable matching for agents in W, the arguments made in the proof of Theorem 1 ensure that $\Omega_W = \mathrm{DA}_W$ when $|W| = 2$. Thus, $\Omega_W \neq \mathrm{DA}_W$ implies that $|W| > 2$.

  We claim that DA_W is not strategy-proof for M when $|W| > 2$. Indeed, let $\succ \in \mathcal{P}$ be a preference profile such that, for some $m_1, m_2 \in M$ and $w_1, w_2, w_3 \in W$ we have that:

$$w_1 \succ_{m_1} w_2 \succ_{m_1} w_3 \succ_{m_1} \cdots, \qquad w_2 \succ_{m_2} w_1 \succ_{m_2} w_3 \succ_{m_2} \cdots,$$

$$m_2 \succ_{w_1} m_1 \succ_{w_1} \cdots, \qquad m_1 \succ_{w_2} m_2 \succ_{w_2} \cdots, \qquad m_1 \succ_{w_3} m_2 \succ_{w_3} \cdots.$$

  Then, $\mathrm{DA}_W[\succ](m_1) = w_2$. However, when m_1 reports preferences $w_1 \succ'_{m_1} w_3 \succ'_{m_1} w_2 \succ'_{m_1} \cdots$, we have that $\mathrm{DA}_W[\succ'_{m_1}, \succ_{-m_1}](m_1) = w_1$. Thus, DA_W is not strategy-proof for M.

(iii) This property is a direct consequence of Theorem 3(iii) and Theorem 4(c). □

References

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

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

[3] Akbarpour, M., A. Kapor, C. Neilson, W. Van Dijk, and S. Zimmerman (2022): "Centralized school choice with unequal outside options," *Journal of Public Economics*, 210, 104644.

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

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

[6] Dubins, L., and D.A. Freedman (1981): "Machiavelli and the Gale-Shapley algorithm," *The American Mathematical Monthly*, 88, 485-494.

[7] Echenique, F., N. Immorlica, and V. Vazirani (2023): "Online and matching-based market design," Cambridge University Press.

[8] Erdil, A. (2014): "Strategy-proof stochastic assignment," *Journal of Economic Theory*, 151, 146-162.

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

[10] Gale, D., and M. Sotomayor (1985): "Some remarks on the stable matching problem," *Discrete Applied Mathematics*, 11, 223-232.

[11] Hassidim, A., D. Marciano, A. Romm, and R. Shorrer (2017): "The mechanism Is truthful, why aren't you?," *American Economic Review: Papers & Proceedings*, 107, 220-224.

[12] Hatfield, J., and P. Milgrom (2005): "Matching with contracts," *American Economic Review*, 95, 913-935.

[13] Hirata, D., and Y. Kasuya (2017): "On stable and strategy-proof rules in matching markets with contracts," *Journal of Economic Theory*, 168, 27-43.

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

[15] Kesten, O., and M. Kurino (2019): "Strategy-proof improvements upon deferred acceptance: A maximal domain for possibility," *Games and Economic Behavior,* 117, 120-143.

[16] Knuth, D. (1976): "Mariages stables et leurs relations avec d'autres problèmes combinatoires," Les Presses de l'Université de Montréal, Montréal

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

[18] Sakai, T. (2011): "A note on strategy-proofness from the doctor side in matching with contracts," *Review of Economic Design*, 15, 337-342.

[19] Svensson, L-G. (1999): "Strategy-proof allocation of indivisible goods," *Social Choice and Welfare*, 16, 557-567.

[20] Tadenuma, K., and M. Toda (1998): "Implementable stable solutions to pure matching problems," *Mathematical Social Sciences*, 35, 121-132.

[21] Teo, C., J. Sethuraman, and W. Tan (2001): "Gale-Shapley stable marriage problem revisited: strategic issues and applications," *Management Science*, 47, 1252-1267.