

Low Information Manipulations of Stable School Choice Mechanisms

by Benjamin Reveler

(6462847)

Major Paper presented to the
Department of Economics of the University of Ottawa
in partial fulfillment of the requirements of the M.A. Degree
Supervisor: Professor Vikram Manjunath

ECO 6999

Ottawa, Ontario
April 2019

1 Introduction

I consider the problem of assigning a set of agents to a set of schools (Abdulkadiroğlu and Sönmez, 2003). Agents have strict, transitive and complete preferences over schools and an outside option. The main normative criterion for assignments is stability. A stable outcome is such that (i) no agent desires entry at a school occupied by a second agent over whom she has priority, (ii) no agent prefers a school with remaining availability to that which she attends and (iii) no agent attends a school which she considers unacceptable.

The appeal of a mechanism certain to yield a stable outcome is evident, as such an outcome is fair towards all agents. Regrettably, with the exception of the agent-proposing deferred acceptance mechanism (Dubins and Freedman, 1981; Roth, 1982), all such mechanisms are vulnerable to strategic manipulations (Alcade and Barberà, 1994).

A mechanism in which truthful preference revelation is a dominant strategy is strategy-proof. Strategy-proofness entails that, in all possible states, an omniscient agent is unable to successfully manipulate the mechanism. Although strategy-proofness is a desirable property, it is also stringent, particularly when one considers that agents seldom possess detailed information regarding the preferences of other agents.

In a recent paper, Troyan and Morrill (2018) define a notion of non-manipulability that corresponds to a setting in which agents are completely unaware of all other agent’s preferences. Within this setting, they prove that, under a minor regularity assumption, all stable-dominating¹ and, by implication, all stable mechanisms are “not obviously manipulable.” Their result provides further credence to all stable mechanisms, as they demonstrate that agents are often unable to identify opportunities to manipulate stable mechanisms when they do not know other agents’ preferences (Troyan and Morrill, 2018).

The aim of this paper is to examine the ability of agents to manipulate stable mechanisms when endowed with varying amounts of information. Consider first the situation where each agent knows only the number of other agents who rank each school first. More generally, I suppose that each agent knows the number of agents who rank each list of k distinct possible schools as their preference over their k most preferred schools. Such knowledge allows agents to only consider a subset of possible outcomes associated with any given announcement of preferences. Under a

¹ A stable-dominating mechanism is any mechanism which is either stable or pareto-dominates a stable mechanism. See Alva and Manjunath (2019).

minor regularity assumption, I show that, with the exception of the agent-proposing deferred acceptance mechanism, all stable mechanisms are vulnerable to manipulation even when only statistics regarding other agents' single most preferred assignment are available.

To compare the possible outcomes under the different preferences that an agent could announce, I extend her preference over assignments to an incomplete preference over sets of assignments. I assume that she considers one set to be at least as desirable as another if (i) the worst assignment within the former set is at least as desirable as the worst assignment within the latter set and (ii) the best assignment within the former set is at least as desirable as the best assignment within the latter set. Previous authors have argued that more detailed set comparisons may be too complex if agents must consider the implications of actions in a large number of possible states (Bossert, et al. 2000; Troyan and Morrill, 2018) and that the best and worst possible assignment are also often the most influential (Troyan and Morrill, 2018). Most importantly, Bossert (1989) demonstrates that the extension described above is the only extension that satisfies the Gärdenfors principle,² monotonicity³ and neutrality.⁴

The contribution of this paper is a method to evaluate the manipulability of mechanisms based on the amount of information required for their manipulation.

2 The Model

Let N denote a finite set of n **agents** and S denote a finite set of m of **schools**. For each $s \in S$, $q_s \in \mathbb{N}$ represents the **capacity** of s . Let $Q = \sum_{s \in S} q_s$ denote the **total school capacity**. Let $q = (q_s)_{s \in S}$ define a **profile of capacity**.

Let \emptyset denote the consumption of an **outside option**. For each $i \in N$, $S \cup \{\emptyset\}$ is the set of agent i 's potential assignments. Thus, each agent $i \in N$ either consumes her outside option or a seat at some school $s \in S$. Let \mathcal{A} denote the set of all **allocations**. An allocation, $\mu \in \mathcal{A}$, assigns each agent to a school or the outside option and respects that no school may exceed its capacity. For each $i \in N$, let $\mu_i \in S \cup \{\emptyset\}$ represent **agent i 's assignment** under μ . Similarly, for each $s \in S$, let $\mu_s \subseteq N$ denote the **set of agents assigned to school s** under μ . Formally, for each $\mu \in \mathcal{A}$,

² For linear orders, this says that (i) adding something superior to every element of a set improves it and (ii) adding something inferior to every element of a set worsens it. See Gärdenfors (1976).

³ For linear orders, this says that adding the same element to two sets does not reverse which set is superior.

⁴ This says that preferences are unaffected by the renaming of elements.

$i \in N$ and $s, s' \in S$, μ satisfies each of the following conditions: (i) $\mu_i = s$ if and only if $i \in \mu_s$, (ii) $|\mu_s| \leq q_s$ and (iii) $i \in \mu_s$ implies that for all $s \neq s'$, $i \notin \mu_{s'}$. For every $\mu \in \mathcal{A}$ and $i \in N$, agent i is a **participant** in μ if and only if $\mu_i \in S$.

Let \mathcal{P} denote the set of all transitive, complete and anti-symmetric binary relations over $S \cup \{\emptyset\}$. Each agent, $i \in N$, ranks each school and her outside option according to her **preference ordering**, $P_i \in \mathcal{P}$. For all $s, s' \in S \cup \{\emptyset\}$, let $s P_i s'$ indicate that agent i prefers assignment s to assignment s' when her preference is P_i . Likewise, for all $s, s' \in S \cup \{\emptyset\}$, let $s R_i s'$ indicate that either $s = s'$ or $s P_i s'$. Let $\mathbf{P} = (P_i)_{i \in N}$ represent a **preference profile**.

For all $i \in N$, $s \in S \cup \{\emptyset\}$ and $P_i \in \mathcal{P}$, let $\mathbf{B}(P_i, s) = \{s' \in S \cup \{\emptyset\} : s' P_i s\}$ represent the strict upper contour set of s at P_i . That is, $B(P_i, s)$ is the set of assignments which agent i prefers to assignment s .

Each school, $s \in S$, has a transitive, complete and anti-symmetric priority ordering over N represented by \succ_s . For all $s \in S$ and $i, j \in N$, let $i \succ_s j$ indicate that school s prioritizes agent i over agent j . Likewise, for all $s \in S$ and $i, j \in N$, let $i \succcurlyeq_s j$ indicate that either $i = j$ or $i \succ_s j$. Let $\succ = (\succ_s)_{s \in S}$ represent the **school priority profile**.

Fix N, S, q and \succ . For each preference profile, $P \in \mathcal{P}^N$, a **mechanism**, $\varphi: \mathcal{P}^N \rightarrow \mathcal{A}$, produces a allocation. For each $P \in \mathcal{P}^N$, let $\varphi(P) \in \mathcal{A}$ denote the **allocation selected by φ at P** . For each $i \in N$, let $\varphi_i(P) \in S \cup \{\emptyset\}$ denote **agent i 's assignment by the mechanism φ at P** .

3 Properties of Allocations and Mechanisms

In this section I describe several relevant axioms.

Individual Rationality For each $P \in \mathcal{P}^N$, an allocation, $\mu \in \mathcal{A}$, is individually rational at P if no agent prefers her outside option to her assignment under μ . That is, μ is individually rational at P if there does not exist $i \in N$ such that $\emptyset P_i \mu_i$. A mechanism, φ , is individually rational if for every $P \in \mathcal{P}^N$, $\varphi(P)$ is individually rational at P .

No Justified Envy For each $P \in \mathcal{P}^N$, an agent $i \in N$ and school $s \in S$ form a **blocking pair** at $\mu \in \mathcal{A}$ if both agent i and school s prefer the other to their assignment under μ . That is, for each $P \in \mathcal{P}^N$, agent i and school s form a blocking pair at μ if $s P_i \mu_i$ and for some agent $j \in \mu_s$, $i \succ_s j$.

For each $P \in \mathcal{P}^N$, an allocation, $\mu \in \mathcal{A}$, satisfies no justified envy if there does not exist a blocking pair. A mechanism, φ , satisfies no justified envy if for every $P \in \mathcal{P}^N$, $\varphi(P)$ satisfies no justified envy at P .

Non-wastefulness For each $P \in \mathcal{P}^N$, an allocation, $\mu \in \mathcal{A}$, is non-wasteful if no agent prefers a school with remaining capacity to her assignment under μ . That is, μ is non-wasteful if there does not exist $i \in N$ and $s \in S$ such that $s P_i \mu_i$ and $|\mu_s| < q_s$. A mechanism, φ , is non-wasteful if for every $P \in \mathcal{P}^N$, $\varphi(P)$ is non-wasteful at P .

Stability For all $P \in \mathcal{P}^N$, an allocation, $\mu \in \mathcal{A}$ is stable at P if μ is individually rational, non-wasteful and satisfies no justified envy at P . For all $P \in \mathcal{P}^N$, let $\mathcal{F}(P) \subseteq \mathcal{A}$ be the **set of all stable allocations** at P . A mechanism, φ , is stable if for every $P \in \mathcal{P}^N$, $\varphi(P) \in \mathcal{F}(P)$.

Define φ^{AO} as the **agent-optimal stable mechanism**. For every $P \in \mathcal{P}^N$, φ^{AO} is such that $\varphi^{AO}(P) \in \mathcal{F}(P)$. Furthermore, for every $P \in \mathcal{P}^N$ and $\mu \in \mathcal{F}(P)$, there does not exist $i \in N$ such that $\mu_i P_i \varphi_i^{AO}(P)$ (Gale and Shapley, 1962). The **agent-proposing deferred acceptance algorithm** computes $\varphi^{AO}(P)$ for each $P \in \mathcal{P}^N$ (Gale and Shapley, 1962).

Define φ^{AP} as the **agent-pessimal stable mechanism**. For every $P \in \mathcal{P}^N$, φ^{AP} is such that $\varphi^{AP}(P) \in \mathcal{F}(P)$. Furthermore, for every $P \in \mathcal{P}^N$ and $\mu \in \mathcal{F}(P)$, there does not exist $i \in N$ such that $\varphi_i^{AP}(P) P_i \mu_i$. The **school-proposing deferred acceptance algorithm** computes $\varphi^{AP}(P)$ for each $P \in \mathcal{P}^N$.

Strategy-Proofness A mechanism, φ , is strategy-proof if for every $P \in \mathcal{P}^N$, no agent can benefit by strategically misreporting her true preferences. That is, φ is strategy-proof if there does not exist $P \in \mathcal{P}^N$, $i \in N$ and $P'_i \in \mathcal{P}$, such that $\varphi_i(P'_i, P_{-i}) P_i \varphi_i(P)$.

A mechanism is strategy-proof only if, for all possible preferences of others, it is immune to manipulation even by agents fully informed of all other agents' preferences. However, if agents are unaware of other agents' preferences, they are likely ignorant of the existing possibilities to

beneficially misreport their preferences. As a result, a less restrictive notion of non-manipulability may be appropriate.

Consider an agent $i \in N$ with preferences $\bar{P}_i \in \mathcal{P}$. For all agents, $i \in N$, and all possible preference relations, $P'_i \in \mathcal{P}$, agent i considers the set of all possible assignments that could result from reporting P'_i as a function of other agents' preferences, $\{s \in S: s = \varphi_i(P'_i, P_{-i}), P_{-i} \in \mathcal{P}^{N \setminus \{i\}}\}$. For all $i \in N$ and $P'_i \in \mathcal{P}$, agent i 's true preferences, $\bar{P}_i \in \mathcal{P}$, cannot compare the set of possible assignments associated with \bar{P}_i and P'_i . An extension of \bar{P}_i to preferences over sets of assignments allows for such a comparison. For all $i \in N$ and $P'_i \in \mathcal{P}$, agent i compares the worst assignment associated with P'_i to the worst assignment associated with \bar{P}_i and the best possible assignments associated with P'_i to the best assignment associated with \bar{P}_i . For each $P'_i \in \mathcal{P}$, agent i evaluates the both the worst and the best assignment in $\{s \in S: s = \varphi_i(P'_i, P_{-i}), P_{-i} \in \mathcal{P}^{N \setminus \{i\}}\}$ with respect to her true preferences, \bar{P}_i .

Non-Obvious Manipulability (NOM) A mechanism, φ , is not obviously manipulable (NOM) if for every agent $i \in N$ and every true preference ranking, $\bar{P}_i \in \mathcal{P}$, there does not exist a strategic preference manipulation, $P'_i \in \mathcal{P}$, such that either:⁵

- (i) $\min_{P \in \mathcal{P}^N} \varphi_i(P'_i, P_{-i}) \bar{P}_i \min_{P \in \mathcal{P}^N} \varphi_i(\bar{P}_i, P_{-i})$
- (ii) $\max_{P \in \mathcal{P}^N} \varphi_i(P'_i, P_{-i}) \bar{P}_i \max_{P \in \mathcal{P}^N} \varphi_i(\bar{P}_i, P_{-i})$

If φ is not NOM then φ is **Obviously Manipulable (OM)**.

Strategy-proofness and non-obvious manipulability correspond to extreme situations in which agents either know everything or nothing about the preferences of other agents. However, in reality, agents may also have incomplete information regarding other agent's preferences. Below I formalize a concept of such knowledge.

For all $k = 0, \dots, m$, let $\Pi(k)$ be the **set of all lists of k distinct elements** of $S \cup \{\emptyset\}$. For all $i \in N$ and $P_i \in \mathcal{P}_i$, let $\tau^k(P_i) \in \Pi(k)$ represent agent i 's preferences over her k most preferred assignments under P_i . For all $P \in \mathcal{P}^N$, $k = 0, \dots, m$ and $\pi \in \Pi(k)$, x_π represents the number of agents, $j \in N$, who rank π as their preference over their k most preferred assignments under P . For all $P \in \mathcal{P}^N$, let $\mathbf{x} = (x_\pi)_{\pi \in \Pi(k)}$ represent each agent's **knowledge** regarding preferences. For all

⁵ Troyan and Morrill's definition of NOM includes an additional minor requirement.

$k = 0, \dots, m$, $i \in N$, $\bar{P}_i \in \mathcal{P}$ and $x \in \mathbb{N}^{\Pi(k)}$ such that $\sum_{\pi \in \Pi(k)} x_\pi = n$, let $\mathcal{K}(\bar{P}_i, x) = \{P \in \mathcal{P}^N : P_i = \bar{P}_i \text{ \& \forall } \pi \in \Pi(k) | j \in N : \tau^k(P_j) = \pi \} = x_\pi\}$ represent the set of preference profiles in which agent i infers the true preference profile exists.

k -Non-Obvious Manipulability (k -NOM) A mechanism is k -NOM if for every agent $i \in N$ and every true preference ranking, $\bar{P}_i \in \mathcal{P}$, there does not exist a set of knowledge, $x \in \mathbb{N}^{\Pi(k)}$, and strategic preference manipulation, $P'_i \in \mathcal{P}$, such that either:

- (i) $\min_{P \in \mathcal{K}(\bar{P}_i, x)} \varphi_i(P'_i, P_{-i}) \bar{P}_i \min_{P \in \mathcal{K}(\bar{P}_i, x)} \varphi_i(\bar{P}_i, P_{-i})$
- (ii) $\max_{P \in \mathcal{K}(\bar{P}_i, x)} \varphi_i(P'_i, P_{-i}) \bar{P}_i \max_{P \in \mathcal{K}(\bar{P}_i, x)} \varphi_i(\bar{P}_i, P_{-i})$

If φ is not k -NOM then φ is **k -Obviously Manipulable (k -OM)**.

Remark 1: 0-NOM is equivalent to NOM.

Remark 2: For all $k, k' = 0 \dots m$ such that $k < k'$, k -NOM is a weaker condition than k' -NOM. If φ is k' -NOM, then φ is also k -NOM. The reverse may not be true.

Remark 3: m -NOM is a weaker condition than strategy-proofness. If φ is strategy-proof, then φ is also m -NOM. The reverse may not be true, as for each $i \in N$, $s, s' \in S \cup \{\emptyset\}$, agent i only knows the number of other agents, $j \in N \setminus \{i\}$, who prefer assignment s to assignment s' .

In the remainder of this section, I define a regularity condition with regards to the following preference transformation.

Let $P \in \mathcal{P}^N$ and $\mu^{AO} = \varphi^{AO}(P)$. Let $P' \in \mathcal{P}^N$ be such that for all $i \in N$ and $s, s' \in S \cup \{\emptyset\} \setminus \mu_i^{AO}$, $s P'_i s'$ if and only if $s P_i s'$, and for all $s \in S \cup \{\emptyset\}$, $\mu_i^{AO} R'_i s$. Such P' is the **promotion of the best attainable assignment (PBAA)** at P .

The following proposition demonstrates that above transformation is innocuous when one only considers stable allocations.

Proposition 1: If P' is the PBAA at P then $\mathcal{F}(P) = \mathcal{F}(P')$.

Proof:

Consider a fixed allocation $\mu \in \mathcal{A}$ and let $\mu^{AO} = \varphi^{AO}(P)$. Below I illustrate that (i) μ is individually rational at P if and only if μ is individually rational at P' , (ii) μ is non-wasteful at P'

if μ is stable at P , (iii) μ is wasteful at P' if μ is wasteful at P , (iv) μ satisfies no justified envy at P' if μ is stable at P' and (v) μ violates no justified envy at P' if μ violates no justified envy at P .

Let $\bar{N} = \{i \in N: \mu_i = \mu_i^{AO}\}$.

By definition of P' , for all $i \in \bar{N}$, since $\mu_i = \mu_i^{AO}$, $B(P'_i, \mu_i)$ is empty.

By definition of P' , for all $i \in N \setminus \bar{N}$, $B(P'_i, \mu_i) \supseteq B(P_i, \mu_i)$ and $\mu_i^{AO} P'_i \mu_i$.

- (i) Since μ^{AO} is individually rational at P , there does not exist $i \in N$ such that $\emptyset P_i \mu_i^{AO}$. Thus, by definition of P' , for all $i \in N$, $\emptyset P_i \mu_i$ if and only if $\emptyset P'_i \mu_i$. As a result, μ is individually rational at P if and only if μ is individually rational at P' .
- (ii) Suppose $\mu \in \mathcal{F}(P)$. Since μ^{AO} is the agent-optimal stable allocation at P , there does not exist $i \in N$ such that $\mu_i P_i \mu_i^{AO}$. Thus, by definition of P' , for all $i \in N$, $B(P'_i, \mu_i) \subseteq B(P_i, \mu_i)$. Since μ is non-wasteful at P , there does not exist $i \in N$ and $s \in S$ such that $s P_i \mu_i$ and $|\mu_s| < q_s$. Suppose there exists $i \in N$ and $s \in S$ such that $\emptyset P'_i \mu_i$ and $|\mu_s| < q_s$. This contradicts that for all $i \in N$, $B(P'_i, \mu_i) \subseteq B(P_i, \mu_i)$. As a result, μ is also non-wasteful under P' .
- (iii) Suppose μ is wasteful at P . Therefore, there exists $i \in N$ and $s \in S$ such that $s P_i \mu_i$ and $|\mu_s| < q_s$. It must either be the case that $i \in \bar{N}$ or $i \in N \setminus \bar{N}$.
 - (a) Consider a case in which $i \in \bar{N}$. Since μ^{AO} is non-wasteful at P and $s P_i \mu_i$, there exists $l \in N$ such that $s = \mu_l^{AO}$. Furthermore, because $|\mu_s| < q_s$, $l \in N \setminus \bar{N}$ such that $s = \mu_l^{AO}$. Moreover, because $l \in N \setminus \bar{N}$, $\mu_l^A P'_l \mu_l$. As a result, μ is wasteful at P' .
 - (b) Consider a case in which there exists $i \in N \setminus \bar{N}$. Suppose there does not exist $i \in N \setminus \bar{N}$ and $s \in S$ such that $s P'_i \mu_i$ and $|\mu_s| < q_s$. This contradicts that for all $i \in N \setminus \bar{N}$, $B(P'_i, \mu_i) \supseteq B(P_i, \mu_i)$. As a result, μ is wasteful at P' .
- (iv) Suppose $\mu \in \mathcal{F}(P)$. Since μ^{AO} is the agent-optimal stable allocation at P , there does not exist $i \in N$ such that $\mu_i P_i \mu_i^{AO}$. Thus, by definition of P' , for all $i \in N$, $B(P'_i, \mu_i) \subseteq B(P_i, \mu_i)$. Since μ satisfies no justified envy at P , there does not exist $i, j \in N$ and $s \in$

S such that $j \in \mu_s$, $i \succ_s j$ and $s P_i \mu_i$. Suppose there exists $i, j \in N$ and $s \in S$ such that $j \in \mu_s$, $i \succ_s j$ and $s P'_i \mu_i$. This contradicts that for all $i \in N$, $B(P'_i, \mu_i) \subseteq B(P_i, \mu_i)$. As a result, μ also satisfies no justified envy at P' .

- (v) Suppose μ violates no justified envy at P . Therefore, there exists $i, j \in N$ and $s \in S$ such that $j \in \mu_s$, $s P_i \mu_i$ and $i \succ_s j$. It must either be the case that $i \in \bar{N}$ or $i \in N \setminus \bar{N}$.
- (a) Consider a case in which $i \in \bar{N}$. Since μ^{AO} is non-wasteful and satisfies no justified envy at P and $s P_i \mu_i$, there exists $l \in N$ such that $s = \mu_l^{AO}$ and $l \succ_s i$. Furthermore, because $j \in \mu_s$, $l \in N \setminus \bar{N}$ such that $s = \mu_l^{AO}$ and $l \succ_s i$. By transitivity, it must be the case that $l \succ_s j$. Moreover, because $l \in N \setminus \bar{N}$, $\mu_l^{AO} P'_l \mu_l$. As a result, μ violates no justified envy at P' .
- (b) Consider a case in which $i \in N \setminus \bar{N}(\mu)$. Suppose there does not exist $i \in N \setminus \bar{N}(\mu)$, $j \in N$ and $s \in S$ such that $j \in \mu_s$, $s P'_i \mu_i$ and $i \succ_s j$. This contradicts that for all $i \in N \setminus \bar{N}$, $B_i(P'_i, \mu_i) \supseteq B_i(P_i, \mu_i)$. As a result, μ violates no justified envy at P' .

■

Proposition 1 suggests that, in regards to stable mechanisms, the following is a mild axiom, as the stable set is unaffected by the promotion of the best attainable assignment.

Invariance to Promotion of the Best Attainable Assignment (IPBAA) A mechanism, φ , is IPBAA if for every $P, P' \in \mathcal{P}^N$ such that P' is the PBAA at P , $\varphi(P) = \varphi(P')$.

Corollary 1: φ^{AO} and φ^{AP} are IPBAA.

5 Results

Theorem 1: If φ is IPBAA and stable but not strategy-proof, then φ is 1 – OM.

Proof:

Let φ be a stable, IPBAA and non-strategy-proof mechanism.

Note that φ^{AO} is the only strategy-proof stable mechanism (Alcade and Barberà, 1994) and pareto-dominates all other stable mechanisms (Gale and Shapley, 1962). As a result, there exists $\bar{P} \in \mathcal{P}^N$ and $i \in N$ such that $\varphi_i^{AO}(\bar{P}) \bar{P}_i \varphi_i(\bar{P})$. Moreover, because φ satisfies IPBAA, there exist $\bar{P} \in \mathcal{P}^N$ such that, for all $i \in N$, $B_i(\bar{P}_i, \varphi_i^{AO}(\bar{P}))$ is empty and such that for some $i \in N$, $\varphi_i^{AO}(\bar{P}) \bar{P}_i \varphi_i(\bar{P})$.

Consider a fixed preference profile $\bar{P} \in \mathcal{P}^N$ such that, for all $i \in N$, $B(\bar{P}_i, \varphi_i^{AO}(\bar{P}))$ is empty and such that for some $i \in N$, $\varphi_i^{AO}(\bar{P}) \bar{P}_i \varphi_i(\bar{P})$.

Let $i \in N$ be the agent such that $\varphi_i^{AO}(\bar{P}) \bar{P}_i \varphi_i(\bar{P})$. Let $s' = \varphi_i^{AO}(\bar{P})$.

Let $\mathcal{K}(\bar{P}_i, x) = \{P \in \mathcal{P}^N : P_i = \bar{P}_i \text{ \& \ } \forall \pi \in \Pi(k) | j \in N : \tau^k(P_j) = \pi\} = x_\pi\}$ represent the statistics available to agent i .

When $k = 1$, each $\pi \in \Pi(k)$ is a singleton representing an individual school. Thus, for all $s \in S$, x_s denotes the number of agents $j \in N$ that rank s as their most preferred school. At every preference profile $P \in \mathcal{P}^N$, $\varphi^{AO}(P)$ is feasible. As a result, for all $s \in S$, $x_s \leq q_s$.

Let $P'_i \in \mathcal{P}$ truncate \bar{P}_i at s' .

P'_i
s'
\emptyset
.
.

For every $P \in \mathcal{K}(\bar{P}_i, x)$, $\varphi_i^{AO}(P'_i, P_{-i}) = s'$ as it is feasible to give each agent her most preferred assignment. Therefore, such an allocation must be stable at P .

For every $P \in \mathcal{K}(\bar{P}_i, x)$, $\varphi_i^{AO}(P'_i, P_{-i}) \in S$. By the Rural Hospital Theorem (Roth, 1986), for every $P \in \mathcal{K}(\bar{P}_i, x)$, $\varphi_i(P'_i, P_{-i}) \in S$. Furthermore, because φ is stable and for every $s \in S \setminus \{s'\}$, $\emptyset P'_i s$, it is the case that for every $P \in \mathcal{K}(\bar{P}_i, x)$, $\varphi_i(P'_i, P_{-i}) = s'$.

Finally, $\bar{P} \in \mathcal{K}(\bar{P}_i, x)$ and is such that $s' \bar{P}_i \varphi_i(\bar{P})$.

Thus, $\min_{P \in \mathcal{K}(\bar{P}_i, x)} \varphi_i(P'_i, P_{-i}) \bar{P}_i \min_{P \in \mathcal{K}(\bar{P}_i, x)} \varphi_i(\bar{P}_i, P_{-i})$ which indicates that φ is 1-OM.

■

Alcade and Barberà (1994) characterize φ^{AO} as the only stable and strategy-proof mechanism. The following corollary of Theorem 1 is a novel characterization.

Corollary 2: φ^{AO} is the only 1-NOM, stable and IPBAA mechanism.

In the proof of Theorem 1, I demonstrate that if a stable and IPBAA mechanism is not strategy-proof, then it is possible for an agent to ensure that she obtains her most preferred assignment. Moreover, to accomplish this, she only requires statistics regarding other agents' single most preferred assignments. My definition of 1-OM is based on a specific extension of each agent's preferences from assignments to sets of assignments. Still, one could weaken this definition to the point where, for a mechanism to be considered manipulable the worst possible assignment from acting strategically would have to be weakly better than the best possible assignment from acting truthfully and strictly better than the worst possible assignment from acting truthfully. The above observation regarding the proof of Theorem 1 indicates that the result would hold, even for a weaker definition of non-manipulability such as this one.

References

- Abdulkadiroğlu, Atila and Tayfun Sönmez (2003) "School Choice: A Mechanism Design Approach." *American Economic Review*, 93, 729-747.
- Alcade, José and Salvador Barberà (1994) "Top Dominance and the Possibility of Strategy-Proof Stable Solutions to Matching Problems." *Economic Theory*, 4, 417-435.
- Alva, Samson and Vikram Manjunath (2019) "Strategy-Proof Pareto-Improvement." *Journal of Economic Theory*, 181, 121-142.
- Bossert, Walter (1988) "On the Extension of Preferences over a Set to the Power Set: An Axiomatic Characterization of Quasi-ordering." *Journal of Economic Theory*, 49, 84-92.
- Bossert, Walter; Prasanta K. Pattanaik and Yongsheng Xu (2000) "Choice under Complete Uncertainty: Axiomatic Characterizations of Some Decision Rules." *Economic Theory*, 16, 295-312.
- Dubins, Lester E. and D. A. Freedman (1981) "Machiavelli and the Gale-Shapely Algorithm." *American Mathematical Monthly*, 88, 485-494.

- Gale, David and Lloyd Shapley (1962) "College Admissions and the Stability of Marriage." *American Mathematical Monthly*, 69, 9-15.
- Gärdenfors, Peter (1976) "Manipulation of Social Choice Functions." *Journal of Economic Theory*, 13, 217-228.
- Roth, Alvin E. (1982) "The Economics of Matching: Stability and Incentives." *Mathematics of Operating Research*, 7, 617-628.
- Roth, Alvin E. (1986) "On the Allocation of Residents to Rural Hospitals." *Econometrica*, 54, 425-427.
- Troyan, Peter and Thayer Morrill (2018) "Obvious Manipulations." Mimeo.