Journal of Mathematical Economics 72 (2017) 16–24

Contents lists available at ScienceDirect

Journal of Mathematical Economics

journal homepage: www.elsevier.com/locate/jmateco

When preference misreporting is Harm[less]ful?

Mustafa Oğuz Afacan [a,1], Umut Mert Dur [b,*]

[a] Faculty of Arts and Social Sciences, Sabancı University, Orhanli, Tuzla, 34956, Istanbul, Turkey
[b] Poole School of Management, North Carolina State University, 27695, Raleigh, NC, USA

ARTICLE INFO

Article history:
Received 30 August 2016
Received in revised form 17 March 2017
Accepted 10 April 2017
Available online 10 June 2017

Keywords:
Harmless
Harmful
Matching
Mechanism
Non-bossiness
Characterization

ABSTRACT

In a school choice problem, we say that a mechanism is harmless if no student can ever misreport his preferences so that he is not hurt but someone else is. We consider two large classes of mechanisms, which include the Boston, the agent-proposing deferred acceptance, and the school-proposing deferred acceptance (*sDA*) mechanisms. Among all the rules in these two classes, the *sDA* is the unique harmless mechanism. We next provide two axiomatic characterizations of the *sDA*. First, the *sDA* is the unique stable, non-bossy, and "independent of an irrelevant student mechanism". The last axiom requires that the outcome does not depend on the presence of a student who prefers being unassigned to any school. As harmlessness implies non-bossiness, the *sDA* is also the unique stable, harmless, and independent of an irrelevant student mechanism. To our knowledge, these axiomatizations as well as the well-known Gale and Shapley's (1962), which reveals that the *sDA* is the student-pessimal stable mechanism, are the only characterizations of the *sDA*.

1. Introduction

Initiated by Gale and Shapley (1962), matching theory has been fruitful both in theory and practice. Theoretical findings have been put into practice, and some important real-life markets have been successfully redesigned by researchers under the guidance of theory. Examples include the National Residency Matching Program in the USA, placing doctors at hospitals, and the New York City and Boston student placement systems.[2]

In the redesign of matching markets, one of the main concerns has been the agents' incentives in reporting their preferences to relevant matching institutions. More specifically, both researchers and market practitioners put high premium on market designs giving right incentives to agents to submit their true preferences in a detail-free way. This property is known as strategy-proofness in the literature. In terms of positive economics, in compliance with the Wilson doctrine, strategy-proofness is highly desirable because it provides participants a very simple optimal strategy. Moreover, from normative point of view, it sustains fairness among sophisticated (strategic) and unsophisticated (sincere) participants by leveling the playing field. The unfairness of the non-strategy-proof Boston mechanism (henceforth, *BM*) has been well-documented by Abdulkadiroğlu et al. (2006). By using the 2001–2002 Boston Public Schools (BPS) data, they show the presence of both sophisticated and sincere students and observe that the former group have indeed hurt the latter.[3] In fact, the following memo from BPS Superintended Payzant to the School Committee on May 25, 2005 reveals the importance that BPS staff gives to strategy-proofness due to its fairness aspect:

> "*A strategy-proof algorithm levels the playing field by diminishing the harm done to parents who do not strategize or do not strategize well.*"

The above empirical finding of Abdulkadiroğlu et al. (2006) is confirmed in both theory and lab by Pathak and Sönmez (2008) and Chen and Sönmez (2006), respectively. Especially, Pathak and Sönmez (2008) show that the Pareto-dominant Nash equilibrium under the *BM* favors sophisticated students at the expense of sincere ones, providing a theoretical support for the unfairness of the *BM* stemming from its lack of strategy-proofness. On the other hand, Kumano (2013) reveals that the *BM* becomes strategy-proof in a very special school-priority domain; thereby it is hard to make the *BM* strategy-proof through playing with priorities as well.[4]

* Corresponding author.
E-mail addresses: mafacan@sabanciuniv.edu (M.O. Afacan), udur@ncsu.edu (U.M. Dur).
1 Fax: +90 216 483 9250.

2 See Abdulkadiroğlu et al. (2005, 2006), Balinski and Sönmez (1999), Abdulkadiroğlu et al. (2005), and Roth and Peranson (1999).

3 They show that many unsophisticated students were unassigned, which would not have been the case if they had chosen a wise preference submission strategy. Similarly, by using data from Wake County School District, Dur et al. (2015) show that sophisticated students benefit under the *BM* at the expense of sincere students.

4 For instance, in order to be out of that special domain, it is enough to have two schools whose capacities are less than the total number of students (see Kumano, 2013).

http://dx.doi.org/10.1016/j.jmateco.2017.04.005

In spite of the advantages of strategy-proofness, there is still a type of unfair situation that strategy-proofness fails to rule out. It is when a student changes someone else's assignment while himself being unaffected. This has already been studied in the literature, and the property eliminating such situations is called "non-bossiness" (see Satterthwaite and Sonnenschein, 1981). It is immediate to observe that if a mechanism is bossy, then there exists a problem where a student harms someone else by misreporting his preferences without himself being unaffected. This is arguably unfair to the harmed student. Moreover, bossy mechanisms call for group manipulations where an agent can form a coalition with someone else to make him to report a false preference list that yields a more favorable assignment for the former agent while the latter is unaffected (the latter can be convinced to do that in return of money).[5]

Motivated by the unfairness aspect of non-strategy-proofness and bossiness, we introduce a notion guaranteeing that no student can ever harm someone else without himself being harmed. More specifically, we say that a mechanism is harmless if whenever a student misreports his preferences without being worse off, then no other student becomes worse off as well. While it is immediate to realize that harmlessness implies non-bossiness, there is no logical relation between strategy-proofness and harmlessness. Moreover, for a strategy-proof mechanism, harmlessness is equivalent to non-bossiness. Hence, if a strategy-proof mechanism lacks harmlessness, then it would be vulnerable to group manipulations.

We first show that the school-proposing deferred acceptance (*sDA*) and the Top Trading Cycles (*TTC*) mechanisms are harmless, whereas the commonly used agent-proposing deferred acceptance (*aDA*) mechanism and the *BM* are not.

Next, we conduct an axiomatic analysis to deepen our understanding of the set of stable and harmless mechanisms. A mechanism is independent of an irrelevant student if the outcome never depends on the presence of a student who prefers being unassigned to every school. We show that a mechanism is stable, non-bossy, and independent of an irrelevant student if and only if it is the *sDA*. As harmlessness implies non-bossiness, as a corollary of this result, we also obtain that the *sDA* is the unique stable, harmless, and independent of an irrelevant student mechanism. Although the *sDA* is widely used in real life matching practices,[6] to the best of our knowledge, there has been no axiomatization of the *sDA* other than being the student-pessimal stable mechanism (Gale and Shapley, 1962); thereby these characterizations shed further light on our *sDA* understanding.

Lastly, in order to investigate whether other harmless mechanisms exist, we consider two large classes of mechanisms, including the *BM*, the *aDA*, and the *sDA*. A subclass of them, which is in use in China for college admissions, was first introduced by Chen and Kesten (2017). Within these classes, the *sDA* happens to be the unique harmless mechanism as well.

2. Related literature

This paper is broadly related to the extensive manipulations in matching markets literature. In two-sided matching markets where both sides are active agents, Roth (1982) shows that no stable mechanism is strategy-proof. On the other hand, whenever either side has commonly known preferences in a one-to-one matching problem, a stable and strategy-proof rule exists (Dubins and Freedman (1981) and Roth (1982)). Sönmez (1997, 1999) demonstrates that stable mechanisms are vulnerable to capacity and pre-arrangement manipulations, respectively. In contrast to stable mechanisms, the *BM* and the *TTC* are immune to capacity manipulations (Kesten, 2012). Some of these negative results are overturned in large markets. Kojima and Pathak (2009) prove that under some regularity conditions, the scope of profitable preference and capacity manipulations diminishes under the *aDA* as the market becomes large. Pápai (2000) shows that strategy-proofness and non-bossiness together are equivalent to group strategy-proofness.[7] Kojima (2010) finds that no stable mechanism is non-bossy in a two-sided matching context.

Pathak and Sönmez (2008) conduct an equilibrium analysis with sophisticated and sincere students under the *BM*. Among other results, they report that sophisticated students weakly prefer the Pareto-dominant Nash equilibrium outcome of the *BM* to the dominant-strategy outcome of the *aDA*. Other papers on manipulations in matching markets include Afacan (2013b), Ergin (2002), Kojima (2006, 2011), Konishi and Ünver (2006), and Sönmez and Pathak (2013).

This paper is also related to the axiomatic characterization literature. Gale and Shapley (1962) show that the *sDA* is the worst stable mechanism for students, in other words, it is the student-pessimal stable mechanism. To our knowledge this was the only *sDA* axiomatization in the literature, and the current paper now provides two new axiomatizations of it in terms of non-bossiness and harmlessness. As opposed to the *sDA*, there are various axiomatizations of the *aDA*, including Alcalde and Barbera (1994), Balinski and Sönmez (1999), Ehlers and Klaus (2014), Kojima and Manea (2010), and Morrill (2013a). The other well-known and in use school choice mechanisms *BM* and the *TTC* have been characterized by Abdulkadiroglu and Che (2010), Afacan (2013a), Dur (2015), Kojima and Ünver (2014), and Morrill (2013b).

3. Model

A **school choice problem** (Abdulkadiroğlu and Sönmez, 2003), or simply a problem, consists of:

- a finite set of students $I = \{i_1, i_2, \ldots, i_n\}$,
- a finite set of schools $S = \{s_1, s_2, \ldots, s_m\}$,
- a quota vector $q = (q_s)_{s\in S}$ where q_s is the number of available seats in school s,
- a list of preference relations $P = (P_i)_{i\in I}$ where P_i is the strict preference of student i over S and the null-school, denoted by $\emptyset$, representing the being unassigned option,
- a list of priority orders $\succ = (\succ_s)_{s\in S}$ where $\succ_s$ is the strict priority relation of school s over I.

Let $q_\emptyset = |I|$. In the rest of the paper, unless otherwise written, we suppress I, S, q, and $\succ$ from the problem notation, and just write P to denote the problem. For any student i and pair of schools s, s', we write sR_is' only if $s = s'$ or sP_is'. Let $\mathcal{P}$ be the set of all possible strict preferences of a student. We say that school s is **acceptable** to student i if $sP_i\emptyset$. Otherwise, it is called unacceptable.

A **matching** $\mu : I \to S \cup \{\emptyset\}$ is a function such that $|\mu(i)| = 1$ and $|\mu^{-1}(s)| \le q_s$ for all $i \in I$ and $s \in S$. Let $\mathcal{M}$ be the set of all matchings.

A matching $\mu \in \mathcal{M}$ **Pareto dominates** another matching $\nu \in \mathcal{M}$ if $\mu(i)R_i\nu(i)$ for each student $i \in I$, with holding strictly for some student. A matching μ is **Pareto efficient** if it is not Pareto dominated.

A matching μ is **non-wasteful** if there exists no student-school pair (i, s) such that $|\mu^{-1}(s)| < q_s$ and $sP_i\mu(i)$.[8] A matching μ is

[5] For a detailed analysis on non-bossiness, refer to Thomson (2016).

[6] Centralized college admissions in Turkey and Ukraine, secondary school assignments in France and Finland, high school assignments in Ireland, and teacher assignment in France are done through variants of the *sDA*.

[7] A mechanism is group strategy-proof if no group of agents can ever collectively misreport their preferences so that none of them is worse off, with at least one of them being better off.

[8] As the null-school is not scarce, non-wastefulness implies individual rationality, which requires from a matching μ that $\mu(i)R_i\emptyset$ for any $i \in I$.