Relaxing Strategyproofness in One-sided Matching

TIMO MENNLE and SVEN SEUKEN
University of Zurich

We consider the one-sided matching problem, where agents must be matched to indivisible objects. Our first result is a novel characterization of strategyproof mechanisms by three intuitive axioms. Furthermore, we introduce *partial strategyproofness*, a new relaxation of strategyproofness that bridges the gap between full and weak strategyproofness. Partial strategyproofness finds application in the incentive analysis of non-strategyproof mechanisms, such as Probabilistic Serial, different variants of the Boston mechanism, and hybrid mechanisms. In this letter, we summarize the main results from [Mennle and Seuken 2014a] and demonstrate how they can be used for the design and analysis of matching mechanisms.[1]

Categories and Subject Descriptors: J.4.a [**Social and Behavioral Sciences**]: Economics

General Terms: Economics; Theory

Additional Key Words and Phrases: Matching, Mechanism Design, Strategyproofness

1. INTRODUCTION

The *(probabilistic) one-sided matching problem* is concerned with the allocation of indivisible goods to self-interested agents with privately known preferences in domains where monetary transfers are not permitted. Such problems often arise in situations that are of great importance to peoples' lives. For example, students must be matched to schools, teachers to training programs, or tenants to houses.

Strategyproof mechanisms, such as Random Serial Dictatorship (RSD), are appealing because they make truthful reporting a dominant strategy for all agents. Participation in the mechanism becomes an easy task as there is no need for deliberation about the best response, thus reducing cognitive costs for the agents and (likely) endowing the mechanism with correct information about agents' preferences.

While strategyproofness is certainly a desirable property, it also imposes severe restrictions. In particular, it is incompatible with ordinal efficiency and symmetry [Bogomolnaia and Moulin 2001], and it is also incompatible with rank efficiency [Featherstone 2011]. Therefore, recent research has sought to develop new methods for describing the incentive properties of non-strategyproof mechanisms. Azevedo and Budish [2013] introduced *strategyproofness in the large*, which formalizes the intuition that as markets get large, the incentives for agents to misreport their preferences vanish in the limit. While this is an interesting new concept, it does not provide any guarantees for finite settings. Pathak and Sönmez [2013] presented a concept to compare mechanisms by their *vulnerability to manipulation*. While this comparison is generally quite appealing, it yields inconclusive results in some important applications, e.g., when comparing different variants of the Boston

[1] [Mennle and Seuken 2014a] is available at: `http://www.ifi.uzh.ch/ce/publications/PSP.pdf`.

Authors' addresses: Department of Informatics, University of Zurich, {mennle,seuken}@ifi.uzh.ch

mechanism [Mennle and Seuken 2014c]. This highlights that the "toolbox" for the incentive analysis of non-strategyproof mechanisms is still insufficiently developed, and new concepts are needed.

In Section 2, we explain our first main result from [Mennle and Seuken 2014a], a new characterization of strategyproof mechanisms. In Section 3 we introduce *partial strategyproofness*, a novel concept for relaxing strategyproofness, which provides a single-parameter measure for "how strategyproofness" a non-strategyproof mechanism is. We apply the partial strategyproofness concept to various known and new mechanisms in Section 4, and conclude in Section 5.

2. AN AXIOMATIC CHARACTERIZATION OF STRATEGYPROOF MECHANISMS

Our first main result is that strategyproof mechanisms are characterized by three intuitive axioms. To explain the axioms, suppose an agent *swaps* the position of two adjacent objects in its reported preference order, say from $a \succ b$ to $b \succ a$. Our axioms limit the way in which a mechanism can change the allocation under this basic kind of manipulation.

The first axiom, *swap monotonicity*, requires the mechanism to be *direct* and *responsive*: if the allocation changes at all, then the allocation of a and b must change (*directness*), and in this case, the allocation for a must strictly decrease, and the allocation for b must strictly increase (*responsiveness*). To illustrate the axiom's significance, consider a mechanism that gives you chocolate ice cream if you ask for vanilla, and gives you vanilla if you ask for chocolate. This mechanism is obviously manipulable, and swap monotonicity prevents this kind of "defect."

The second axiom, *upper invariance*, requires that an agent cannot influence the allocation of one of its better choices by swapping two less preferred objects. Hashimoto et al. [2014] introduced this axiom as one of the central axioms to characterize the Probabilistic Serial mechanism.[2] For individually rational mechanisms and with a null object present, upper invariance is equivalent to *truncation robustness*, i.e., agents cannot benefit by declaring some objects as unacceptable.

The third axiom, *lower invariance*, requires that an agent cannot influence the allocation for *less* preferred objects by swapping two *more* preferred objects. This is a natural complement to upper invariance.

In combination, the three axioms characterize strategyproof mechanisms.

Theorem 1 [Mennle and Seuken 2014a]. *A mechanism is strategyproof if and only if it is swap monotonic, upper invariant, and lower invariant.*

This result yields a new way of establishing (or falsifying) strategyproofness. In particular, it highlights the severity of the restriction that strategyproofness imposes on the mechanism design space: if an agent swaps two objects in its reported preference order (from $a \succ b$ to $b \succ a$), all that a strategyproof mechanism can do is to (weakly) increase the probability for b and to reduce the probability for a by the same amount, while the allocation for all other objects must remain unchanged.

[2]Hashimoto et al. [2014] call this axiom *weak invariance*, but in the context of our characterization, *upper invariance* is the more appropriate name as it complements *lower invariance*.

3. PARTIAL STRATEGYPROOFNESS

Arguably, lower invariance is the least intuitive axiom. By dropping this axiom, a new class of mechanisms emerges, which we call *partially strategyproof*. These mechanisms remain strategyproof, but only on a particular domain restriction, where the agent's underlying utility functions are bounded away from indifference. Formally, a utility function u satisfies *uniformly relatively bounded indifference with respect to* $r \in [0,1]$ (URBI(r)) if for all objects a, b with $u(a) \geq u(b)$ we have

$$r \cdot (u(a) - \min u) \geq u(b) - \min u. \quad (1)$$

We say that a mechanism is *r-partially strategyproof* if it is strategyproof in the domain constrained by URBI(r). If a mechanism is r-partially strategyproof for some non-trivial $r > 0$, then we sometimes simply say that it is *partially strategyproof*. Our second main result is a characterization of partially strategyproof mechanisms for any fixed setting (i.e., number of objects, number of agents, vector of capacities).

Theorem 2 [Mennle and Seuken 2014a]. *In any setting, a mechanism is partially strategyproof if and only if it is swap monotonic and upper invariant.*

Theorems 1 and 2 show that requiring full strategyproofness beyond partial strategyproofness "buys" lower invariance. Interestingly, this insight also leads to a better understanding of what lower invariance actually contributes in terms of incentives. Recall that an r-partially strategyproof mechanism is strategyproof for agents with utilities *inside* URBI(r). However, an agent who is close to indifferent between two objects a and b may benefit from sacrificing some probability for a to gain more probability for the less preferred object b. Under partially strategyproof mechanisms, any beneficial manipulation must involve such a trade-off decision. Thus, lower invariance adds that this trade-off is *never* beneficial, independent of the relative utility differences, even for agents with utilities *outside* URBI(r).

We have also shown that the URBI(r) domain restriction is maximal, i.e., there is no larger domain on which *all* r-partially strategyproof mechanisms are also strategyproof (Theorem 3 in [Mennle and Seuken 2014a]). Thus, for any partially strategyproof mechanism and in any fixed setting, we can consider the largest admissible indifference bound as a single-parameter measure for "how strategyproof" the mechanism is. We call this value the *degree of strategyproofness* and show that it is computable and consistent with the *vulnerability to manipulation* concept.

4. APPLICATION OF PARTIAL STRATEGYPROOFNESS

Our partial strategyproofness concept finds numerous applications in the design and analysis of non-strategyproof mechanisms. First, we show that the *Probabilistic Serial mechanism* (PS) [Bogomolnaia and Moulin 2001] is swap monotonic and upper invariant, and hence partially strategyproof. Moreover, numerical evidence suggests the following conjecture: as the setting becomes large, i.e., the capacities of the objects increases, the degree of strategyproofness of PS converges to 1.

Second, we consider the traditional "naïve" *Boston mechanism* (NBM) [Abdulkadiroğlu and Sönmez 2003] in a setting with no priorities. This mechanism has often been criticized for its manipulability, but it is nevertheless in frequent use in school choice settings in practice. In [Mennle and Seuken 2014c] we have introduced an adaptive variant of the Boston mechanism (ABM): instead of applying to their kth choice in the kth round, ABM lets agents apply to their *best available* choice in each round. This removes obvious opportunities for manipulation, yet a comparison between NBM and ABM by *vulnerability to manipulation* remains inconclusive, except in the most basic case. However, we have shown that ABM is in fact partially strategyproof, while NBM is not (Prop. 1 & Cor. 1 in [Mennle and Seuken 2014c]). Thus, partial strategyproofness makes the different incentive properties explicit.

Third, in [Mennle and Seuken 2014b] we have demonstrated how hybrid mechanisms can be used to trade off between strategyproofness and efficiency. The main idea is to consider convex combinations of two component mechanisms, one of which brings "good incentives" while the other brings "good efficiency." We have shown that, under certain technical assumptions, these *hybrid mechanisms* are partially strategyproof, while also being more efficient than the less efficient component. Furthermore, this trade-off is scalable in the sense that the mechanism designer can accept a lower degree of strategyproofness in exchange for more efficiency. Prior to the introduction of partial strategyproofness, no measure existed to evaluate the incentive properties of such non-strategyproof hybrid mechanisms.

5. CONCLUSION

Strategyproofness is the most attractive incentive property in mechanism design, but it also severely restricts the mechanism design space. Our new characterization by swap monotonicity, upper invariance, and lower invariance provides an easy way of establishing (or falsifying) strategyproofness. Furthermore, our partial strategyproofness concept enables a new kind of incentive analysis of non-strategyproof mechanisms and applies not only in the limit, but in settings of any size. It bridges the gap between weak and full strategyproofness and yields a parametric and computable measure for the degree of strategyproofness. We believe that our results will lead to new insights in the analysis of existing non-strategyproof matching mechanisms and facilitate the design of new ones.

REFERENCES

Abdulkadiroğlu, A. and Sönmez, T. 2003. School Choice: A Mechanism Design Approach. *American Economic Review 93(3)*, 729–747.

Azevedo, E. M. and Budish, E. 2013. Strategyproofness in the Large as a Desideratum for Market Design. Working Paper.

Bogomolnaia, A. and Moulin, H. 2001. A New Solution to the Random Assignment Problem. *Journal of Economic Theory 100,* 2 (Oct.), 295–328.

Featherstone, C. R. 2011. A Rank-based Refinement of Ordinal Efficiency and a new (but Familiar) Class of Ordinal Assignment Mechanisms. Working Paper.

Hashimoto, T., Hirata, D., Kesten, O., Kurino, M., and Ünver, U. 2014. Two Axiomatic Approaches to the Probabilistic Serial Mechanism. *Theoretical Economics 9,* 1.

Mennle, T. and Seuken, S. 2014a. An Axiomatic Approach to Characterizing and Relaxing Strategyproofness of One-sided Matching Mechanisms. Working Paper.

Mennle, T. and Seuken, S. 2014b. Hybrid Mechanisms: Trading off Strategyproofness and Efficiency in One-Sided Matching. Working Paper.

Mennle, T. and Seuken, S. 2014c. The Naïve versus the Adaptive Boston Mechanism. Working Paper.

Pathak, P. and Sönmez, T. 2013. School Admissions Reform in Chicago and England: Comparing Mechanisms by Their Vulnerability to Manipulation. *American Economic Review 103(1)*, 80–106.