Games and Economic Behavior 98 (2016) 68–77

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Monotone strategyproofness ☆

Guillaume Haeringer [a,*], Hanna Hałaburda [b,c]

[a] Department of Economics and Finance, Baruch College, New York, United States
[b] Bank of Canada, Ottawa, Canada
[c] New York University Stern School of Business, New York, United States

ARTICLE INFO

Article history:
Received 19 September 2014
Available online 8 June 2016

JEL classification:
C72
D41

Keywords:
Strategyproofness
Kemeny sets
Misrepresentations
Dominant strategy

ABSTRACT

We propose a way to compare the extent of preference misrepresentation between two strategies. We define a preference revelation mechanism to be *monotone strategyproof* if declaring a "more truthful" preference ordering dominates (with respect to the true preferences) declaring a "less truthful" preference ordering. Our main result states that a mechanism is strategyproof if, and only if, it is monotone strategyproof. This result holds for any deterministic social choice function on any domain; for probabilistic social choice functions it holds under a mild assumption on the domain.

1. Introduction

Truthful revelation is a primary goal in mechanism design. Ideally, it is a dominant strategy to truthfully reveal one's preferences, and a mechanism that induces such a dominant strategy for all agents and all preference profiles is said to be **strategyproof**. Non-trivial strategyproof mechanisms do not always exist if other desired properties are also imposed (Gibbard, 1973; Satterthwaite, 1975), but a number of environments have been identified for which non-trivial strategyproof mechanisms exist, e.g. voting, two-sided matching, house allocation, or auctions.[1]

Strategyproof mechanisms induce a radical division between strategies, for they distinguish the truthful strategy from all other strategies. All non-truthful strategies are deemed undesirable regardless of their other characteristics; a lie is a lie, whether big or small. This gave the prior literature little reason to scrutinize misrepresentations in strategyproof

☆ We are grateful to Miguel Ángel Ballester, Eric Budish, Eric Danan, Lars Ehlers, Matt Jackson, Łukasz Pomorski, Arunava Sen, John Weymark, and especially Vikram Manjunath for their comments and suggestions. Comments from participants to the 2011 GDRI workshop in Marseille and the 2011 Economic Design conference are gratefully acknowledged. We are grateful to two anonymous referees for their detailed comments and suggestions. Guillaume Haeringer acknowledges the support of Barcelona GSE Research Network and of the Government of Catalonia, Spanish Ministry of Science and Innovation through the grant "Consolidated Group-C" ECO2008-04756, the Spanish Ministry of Economy and Competitiveness through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2011-0075), and the *Generalitat de Catalunya* (SGR2009-419). Part of this research was conducted when he was visiting the Stanford Economics department, whose hospitality is gratefully acknowledged. Hanna Hałaburda acknowledges support of HBS Research Division. The views expressed in this paper are those of the authors. No responsibility for them should be attributed to the Bank of Canada.

* Corresponding author.
E-mail addresses: guillaume.haeringer@baruch.cuny.edu (G. Haeringer), hhalaburda@gmail.com (H. Hałaburda).

[1] See for instance Moulin (1980) for voting with single-peaked preferences, Dubins and Freedman (1981) and Roth (1982) for two-sided matching. See also Barberà (2011) for a recent survey.

http://dx.doi.org/10.1016/j.geb.2016.05.007
0899-8256/

mechanisms, for instance by measuring how much they deviate from the truth. We argue that this is an important omission and we focus in this paper specifically on non-truthful strategies in strategyproof mechanisms.

We believe there is a need for a general tool to analyze misrepresentations. There is indeed now growing evidence that strategyproof mechanisms perform poorly in the laboratory (see Chen (2008) for a survey).[2] Actually, experimental data from games with a dominant strategy also exhibit seemingly irrational behavior.[3] Overall, most experimental analysis of strategyproof mechanisms cannot go further than acknowledging the percentage of subjects not being truthful, and analyzing how this percentage varies when changing some environment parameters or the mechanism itself. However, the existing studies have not been able to rank non-truthful strategies on how close they are to the true preferences, save for some specific cases.[4] This is a serious limitation because what makes strategyproof mechanisms appealing is, among other things, their ability to generate quality data about individuals' preferences. Such information is crucial if one wishes to run counterfactuals and test potential new policies. Policy makers (and econometricians) may prefer a mechanism with a large percentage of individuals not being truthful but "close" to the truth over a mechanism with a smaller percentage of misrepresentations but consisting of large deviations from the truth.

From a theoretical perspective we argue that studying misrepresentations can help understanding further the anatomy of strategyproof mechanisms. By its definition, strategyproofness imposes the existence of a dominant strategy in the mechanism. But does it also impose any structure on misrepresentations? To address this question we classify misrepresentations so as to be able to rank strategies on how much they misrepresent the true preferences. Our contention is that such a classification must be linked to the cost of misrepresenting preferences. Drawing on the intuition for strategyproofness, small misrepresentations should have a lower impact on agents' welfare than large ones, or, put differently, small deviations should dominate large ones.[5] We call a mechanism satisfying this property **monotone strategyproof**. One might conjecture that imposing monotonicity between payoffs and distance from the truth would be more restrictive than the usual incentive compatibility, i.e., that some strategyproof mechanisms may not be monotone strategyproof. Our main contribution here is to show that monotone strategyproofness is actually equivalent to strategyproofness. This seemingly counterintuitive result turns out to be straightforward to show and holds for a very general class of environments.

Our result is derived within a typical environment where each individual has a strict preference relation over a finite set of alternatives and participates in a strategyproof mechanism.[6] We first devise a measure to compare the degree of preference misrepresentation. Given two preference orderings P_i and P_i', we define the **Kemeny set** of P_i and P_i' as the pairs of alternatives that are not ordered in the same way under these two preferences.[7] We compare the degree of misrepresentation by comparing Kemeny sets: Given a true preference ordering P_i, an ordering P_i' is defined to be more truthful than P_i'' when the Kemeny set of P_i' and P_i is a subset of that of P_i'' and P_i. That is, P_i' is more truthful than P_i'' when P_i'' has relatively more elements whose order disagrees with P_i. In this context, a mechanism is said to be monotone strategyproof if a more truthful strategy always dominates a less truthful one.[8] It is straightforward to see that monotone strategyproofness implies strategyproofness. Our main result (Theorem 2) states that the reverse also holds under a mild assumption on the domain of the mechanism. For deterministic social choice functions this equivalence actually holds for *any* environment (Theorem 1).

We compare strategies by comparing their Kemeny sets. A natural question is whether a non-truthful strategy P_i' that dominates another non-truthful strategy P_i'' is necessary closer to the true preferences in the way we define it. In other words, is Kemeny set inclusion equivalent to the dominance relation? It turns out that this equivalence is true for deterministic mechanisms, but not for the general case. For non-deterministic mechanisms we show how one preference ordering may dominate another without Kemeny set inclusion. This observation illustrates the complication added by non-deterministic mechanisms.

Two closely related papers are Carroll (2012) and Sato (2013). Like us, they also compare "large" and "small" misrepresentations, but they address a different question than we do. Both Carroll and Sato characterize conditions under which "local" strategyproofness implies "global" strategyproofness, that is, conditions under which restricting misrepresentations that only switch the ranking of two consecutive alternatives in one's preferences is enough to characterize strategyproofness. So their concern is more about the transitivity of strategyproofness.[9] Another related paper is Cho (2014). While considering closely related issues to ours, the analysis in Cho (2014) is constrained by a more restrictive environment. Cho studies probabilistic assignment mechanisms (Carroll, Sato and us consider any social choice mechanism). Cho's main contribution

[2] See for instance Cason et al. (2008) for the pivotal and the Groves–Clarke mechanisms, Chen and Sönmez (2006) or Calsamiglia et al. (2010) in a matching context.

[3] See Palacios-Huerta and Volij (2009) for the centipede game, Kagel and Levin (1986) for auction games or Andreoni (1995) for public good games.

[4] Chen and Sönmez (2006) and Calsamiglia et al. (2010) for instance analyze which type of alternative is likelier to be displaced in the preference orderings.

[5] See Jackson (1992) for a similar argument in the case of in an exchange economy.

[6] As we discuss at the end of this paper (in the Discussion section), our strategyproof equivalence result does not necessarily hold for weak preferences. It is surprising, as it holds (for a very general class of environments) under stochastic mechanisms.

[7] The cardinality of this set is the well-known *Kemeny distance* (Kemeny, 1959).

[8] The equivalent definition for stochastic mechanisms simply replaces dominance with stochastic dominance.

[9] Another, somewhat less related paper, is Pathak and Sönmez (2013), who also focus on misrepresentation of preferences. However, Pathak and Sönmez are interested in comparing mechanisms—and therefore consider mechanisms that are not strategyproof—while we are interested in comparing misrepresentations under strategyproof mechanisms.