Games and Economic Behavior ••• (••••) •••–•••

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Note

Incentive properties for ordinal mechanisms ☆

Wonki Jo Cho

School of Economics, Sogang University, 35 Baekbeom-ro, Mapo-gu, Seoul, 04107, South Korea

ARTICLE INFO

Article history:
Received 24 March 2014
Available online xxxx

JEL classification:
C70
C78
D71

Keywords:
Ordinal mechanism
Strategy-proofness
Adjacent strategy-proofness
Mistake monotonicity

ABSTRACT

We study three incentive properties for ordinal mechanisms: (i) strategy-proofness, which requires that no agent gain by misrepresenting his preferences; (ii) adjacent strategy-proofness, which requires that no agent gain by switching the rankings of two adjacent alternatives; and (iii) mistake monotonicity, which requires that the welfare of each agent weakly decrease as he reports increasingly bigger mistakes. Each of these properties has three versions, depending on whether preferences over sure alternatives are extended to preferences over lotteries by the stochastic dominance, downward lexicographic, or upward lexicographic extension. We identify conditions on the preference domain that guarantee the equivalence of these properties. The universal domain and the domains of single-dipped and single-peaked preferences satisfy our conditions.

1. Introduction

Most real-life mechanisms that use lotteries to allocate resources are ordinal. Agents submit their ordinal preferences, or rankings, over sure alternatives and lotteries over those alternatives are chosen based only on this information. Examples include public housing allocation (Abdulkadiroğlu and Sönmez, 1999), probabilistic assignment of objects (Bogomolnaia and Moulin, 2001), school choice (Abdulkadiroğlu and Sönmez, 2003), and probabilistic voting (Gibbard, 1977). We consider three incentive properties for ordinal mechanisms, study their logical relations, and provide sufficient conditions on the preference domain that ensure their equivalence.

The most common incentive property is strategy-proofness, the requirement that no agent gain by misrepresenting his preferences. Some recent papers study if a weaker incentive property is sufficient for strategy-proofness (Carroll, 2012; Sato, 2013). As an extreme weakening of strategy-proofness, adjacent strategy-proofness requires that no agent benefit from reporting "small lies", namely those lies obtained by switching two alternatives that are adjacent in the true preference rankings. There are several reasons why adjacent strategy-proofness is an interesting notion. First, agents are sometimes constrained to choose preferences that are somewhat close to the truth. Big lies raise consistency and credibility issues. An agent's misrepresentation has to be consistent with the part of his private information that has already been disclosed. This limits the set of credible lies an agent can choose from. Second, weaker incentive properties may affect existing

☆ This paper is based on a chapter in my Ph.D. dissertation at the University of Rochester. I am grateful to William Thomson for his comments and encouragement. I thank an associate editor and two anonymous referees who gave extensive comments on the earlier version of this paper. I also thank Paulo Barelli, Battal Doğan, Srihari Govindan, Guillaume Haeringer, Biung-Ghi Ju, Asen Kochov, Hervé Moulin, Romans Pancs, and seminar participants at Korea University, Seoul National University, UC Riverside, Hitotsubashi University, Sogang University, 2011 Society for Economic Design Conference, and the 23rd International Conference on Game Theory for their comments.

E-mail address: chowonki@sogang.ac.kr.

http://dx.doi.org/10.1016/j.geb.2015.12.003
0899-8256/

results involving strategy-proofness. When combined with other desiderata, strategy-proofness often turns out very restrictive, characterizing a small family of rules or leading to an impossibility (e.g., Gibbard, 1973, 1977; Satterthwaite, 1975; Bogomolnaia and Moulin, 2001). When a weaker incentive property is imposed instead, we may obtain a larger family or restore a possibility. Third, if a weaker incentive property is equivalent to strategy-proofness (as we show), the task of designing a strategy-proof mechanism can be much simplified.

Another incentive property, which we propose, is mistake monotonicity. It requires that an agent's welfare weakly decrease as he reports increasingly bigger mistakes. Given two preference relations over alternatives, one is a bigger mistake than the other if the former is obtained from the latter by switching some pairs of adjacent alternatives, where each such switching is a mistake according to the true preference relation. Discovering one's own preferences is a complex process and agents may make a mistake in their reports to the mechanism. The designer typically seeks to maximize agents' welfare based on the reported preferences. If mistakes are involved in reports, strategy-proofness does not guarantee that the allocation chosen for the reported preferences delivers the maximum social welfare, or a welfare level close to it, for the true preferences. However, mistake monotonicity ensures that a mechanism be robust to mistakes by correlating the degree of mistake with social welfare. This robustness is independent of strategic issues and cannot be captured by strategy-proofness. Clearly, mistake monotonicity is a strengthening of strategy-proofness, but the two are equivalent on some domains.

Our objective is to investigate logical relations among strategy-proofness, adjacent strategy-proofness, and mistake monotonicity. But there is a fundamental issue we need to address to pursue this objective. Since agents only submit preferences over sure alternatives, probabilistic outcomes cannot be directly evaluated according to the elicited preferences. To circumvent this problem, Gibbard (1977) and Bogomolnaia and Moulin (2001) extend preferences over sure alternatives to preferences over lotteries using (first-order) stochastic dominance. We refer to this procedure as the *sd*-extension. Most papers taking the ordinal approach follow this practice (e.g., Che and Kojima, 2010; Hashimoto et al., 2014; Katta and Sethuraman, 2006; Liu and Pycia, 2012). Once we adopt the *sd*-extension, it is automatically embedded in properties of mechanisms and affects their content. For instance, the notion of strategy-proofness based on the *sd*-extension says that for each agent, the lottery he obtains by reporting his true preferences should stochastically dominate any lottery he obtains by lying. Although the *sd*-extension plays a key role in defining properties of ordinal mechanisms, its role has not been investigated so far.

We define an extension as a mapping from preferences over sure alternatives to preferences over lotteries. The *sd*-extension is an example. We consider two extensions related to lexicographic preferences (Hausner, 1954) in parallel with the *sd*-extension. Our first alternative is the "downward lexicographic" extension, or the *dl*-extension. The *dl*-extension gives the following preference relation over lotteries. Lotteries are compared in a lexicographic fashion, starting from the probabilities for the most preferred alternative. Given two lotteries, the lottery that assigns a higher probability to the most preferred alternative is preferred. If the two probabilities are equal, the lottery that assigns a higher probability to the second most preferred alternative is preferred, and so on. The other alternative is the "upward lexicographic" extension, or the *ul*-extension. The preference relation over lotteries obtained by the *ul*-extension performs lexicographic comparison in the opposite way, preferring a lottery that minimizes probabilities for less preferred alternatives. For each of the three incentive properties, the *sd*-, *dl*-, and *ul*-extensions give rise to different notions. Therefore, we prefix them by the corresponding extensions; e.g., for an arbitrary extension e, *e-strategy-proofness*.

We show that under some conditions on the (preference) domain, for each $e \in \{sd, dl, ul\}$, *e-adjacent strategy-proofness* is equivalent to *e-strategy-proofness* (Theorem 1). Our equivalence result generalizes Sato (2013). He restricts attention to deterministic mechanisms and considers two domain conditions: "connectedness" and "non-restoration". Given two preference relations over alternatives, we can always change one to the other by consecutively switching two adjacent alternatives. Two preference relations are connected in a domain if we can change one to the other by performing such "adjacent-pair-switch" operations without leaving the domain. A domain is connected if any two preference relations are connected in the domain. Non-restoration says that for each pair of connected preference relations, we can change one to the other by performing the adjacent-pair-switch operations, without leaving the domain and without reversing the rankings of any two alternatives twice. For deterministic mechanisms (for which the three notions of strategy-proofness coincide), if the domain satisfies connectedness and non-restoration, then adjacent strategy-proofness and strategy-proofness are equivalent (Sato, 2013). We find that non-restoration, together with connectedness, remains sufficient for the equivalence of *sd-adjacent strategy-proofness* and *sd-strategy-proofness*. This is not covered by Carroll (2012) who shows the same equivalence for the "polyhedral type space".

On the other hand, for the *dl*- and *ul*-extensions, non-restoration is not enough. Under the *dl*-extension, probabilities for preferred alternatives give extremely high utility. Therefore, for the equivalence of *dl-adjacent strategy-proofness* and *dl-strategy-proofness*, each pair of preference relations should be connected by a path that first move preferred alternatives to desired positions. E.g., for preference relations 123 and 321, the path $\{123, 132, 312, 321\}$ should be in the domain. We call this the preferred-alternatives-first (PAF) path property. By contrast, under the *ul*-extension, probabilities for less preferred alternatives give extremely high disutility. Thus, the sufficient condition for the *ul*-extension requires that any two preference relations be connected by a path that first move less preferred alternatives to desired positions; e.g., for preference relations 123 and 321, the path $\{123, 213, 231, 321\}$ should be in the domain. We call this the less-preferred-alternatives-first (LAF) path property. The PAF and LAF path properties each imply non-restoration. The universal domain satisfies all the three properties and there are other interesting domains with these properties. The domain of single-dipped

Please cite this article in press as: Cho, W.J. Incentive properties for ordinal mechanisms. Games Econ. Behav. (2015), http://dx.doi.org/10.1016/j.geb.2015.12.003