ARTICLE IN PRESS

Mathematical Social Sciences ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Mathematical Social Sciences

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

Immunity to credible deviations from the truth

Salvador Barberà [a], Dolors Berga [b,*], Bernardo Moreno [c]

[a] MOVE, Universitat Autònoma de Barcelona, and Barcelona GSE. Departament d'Economia i d'Història Econòmica, Edifici B, 08193 Bellaterra, Spain

[b] Departament d'Economia, C/ Universitat de Girona, 10; Universitat de Girona, 17003 Girona, Spain

[c] Departamento de Teoría e Historia Económica, Facultad de Ciencias Económicas y Empresariales, Universidad de Málaga, Campus de El Ejido, 29071 Málaga, Spain

HIGHLIGHTS

- We define the notion of immunity to credible deviations.
- We discuss alternative versions of credibility.
- We single out immune rules with multidimensional alternatives and single-peakedness.
- We identify voting by quota 1 and n as the unique GCW immune rules.

ARTICLE INFO

Article history:
Received 15 March 2016
Received in revised form
11 November 2016
Accepted 15 November 2016
Available online xxxx

ABSTRACT

We study a notion of non-manipulability by groups, based on the idea that only some agreements among potential manipulators may be credible. The derived notion of immunity to credible manipulations by groups is intermediate between individual and group strategy-proofness. Our main non-recursive definition turns out to be equivalent, in our context, to the requirement that truthful preference revelation should be a strong coalition-proof equilibrium, as recursively defined by Peleg and Sudhölter (1998, 1999). We provide characterizations of strategy-proof rules separating those that satisfy it from those that do not for a large family of public good decision problems.

1. Introduction

In many contexts where the basic incentive property of strategy-proofness can be met by non-trivial social choice functions, it becomes natural to investigate whether some of them may not only be immune to manipulation by individuals, but can also resist manipulation by groups of coordinated agents. In previous work (Barberà et al., 2010, 2016) we have identified conditions under which, surprisingly, all social choice functions that are immune to manipulations by individuals will also be free from group manipulation. But this is not always the case. In particular, many interesting strategy-proof rules in separable environments[1] will indeed be group manipulable. In these cases, we shall argue that not all group manipulations represent an equally serious threat, because some strategic movements by coalitions are credible, while others are not. To make this point precise, we define several notions of immunity to credible group manipulations and characterize subclasses of social choice rules that satisfy them in specific environments. We concentrate especially in the following: we say that a deviation leading to a profitable improvement for a group is credible if no individual member of the group would gain from not following the agreed upon strategy under the assumption that all others stick to the agreement. Hence, a group manipulation is credible if the set of prescribed strategies for those individuals who plan to deviate are a Nash equilibrium in the induced game where these agents could use any other preference instead, while those of the rest of agents remain fixed.[2] And then we say that a rule is immune to credible group manipulations if no set of agents can find a profitable deviation away from the truth that is

* Corresponding author.
E-mail addresses: salvador.barbera@uab.cat (S. Barberà), dolors.berga@udg.edu (D. Berga), bernardo@uma.es (B. Moreno).

1 We use this expression loosely here. Formal definitions of the environments we refer to are given in Section 3.

2 Actually, the concept remains the same if the possible deviations of manipulators are limited to either following the prescription or revealing their true preferences. See Section 4 for a deeper discussion of this and related points.

http://dx.doi.org/10.1016/j.mathsocsci.2016.11.002
0165-4896/

credible. We illustrate the strength of our new definition, which is more demanding than individual but weaker than group strategy-proofness, by characterizing some families of rules in separable environments, and distinguishing between those that can meet our new requirement and those that cannot. The issue of credibility of group deviations has been formalized in alternative ways, one of which is based on the recursive definition of coalition-proof Nash equilibrium (see Bernheim et al., 1987). In fact, Peleg and Sudhölter (1999) applied this concept to the same environment that we analyze, and concluded that all rules that are strategy-proof in that environment are also coalition-proof. In the same paper, these authors (see also Peleg, 1998) define strong coalition-proofness, again recursively based. Surprisingly, this turns out to be equivalent, in our context, to our non-recursive concept of immunity. Let us remark again that our notion of immunity, and that of strong coalition-proofness, allows for a classification and characterization of different strategy-proof rules according to their degree of group manipulability.

After this Introduction we provide notation and definitions in Section 2. Section 3 presents characterization results in two specific contexts. We start with the problem faced by voters who must select a set of entrants to a club, as described in Barberà et al. (1991). We concentrate on quota rules: voters can support all candidates they like, and then those who receive at least a fixed number of votes, q, are chosen. In the domain of separable preferences, we show that rules based on quota 1 or n (where n is the number of voters) are immune to credible deviations, while all other rules in the class are not. Hence, very extreme distributions of power among voters are needed to guarantee immunity. We then turn to a general version of choice among multi-dimensional alternatives under separable preferences, also called multidimensional single-peaked. We build on Moulin (1980), Border and Jordan (1983) and Barberà et al. (1993). The cases we consider include the previous example and many more. We restrict attention to a large class of rules that are strategy-proof in this context, and again characterize those within the class that are immune to credible deviations by groups. Again, a requirement in the form of unanimity plays a crucial role in separating these rules for all the rest, those that are credibly manipulable. Section 4 discusses alternative definitions of credibility for group manipulations, establishes the equivalence of several apparently different formulations, and the differences with other potential definitions, whose proofs are also examined. Section 5 concludes with some final remarks. Appendix contains proofs that are not essential for the continuity of our arguments.

2. The model and definitions: immunity and credibility

Let $N = \{1, \ldots, n\}$ be the set of *agents* and A be the set of *alternatives*. *Preferences* are complete, reflexive, and transitive binary relations on alternatives. Let $\mathcal{U}$ denote such set of preferences. For $i \in N$, R_i denotes agent i's *preferences* on A. As usual, P_i and I_i denote the strict and indifference preference relation induced by R_i, respectively. A *preference profile* $R_N = (R_1, \ldots, R_n) \in \mathcal{U} \times \cdots \times \mathcal{U} = \mathcal{U}^n$ is a n-tuple of preferences on A. It can also be represented by $R_N = (R_C, R_{N\setminus C}) \in \mathcal{U}^n$ when we want to stress the role of coalition C in N. We call a subprofile of agents in C as $R_C \in \times_{i\in C}\, \mathcal{U} = \mathcal{U}^c$.

A *social choice function* (or *rule*) f on $\mathcal{U}^n$ is a function $f : \mathcal{U}^n \to A$.

At this point it is worth mentioning that although we define our main concept and state our results in Sections 2 and 4 assuming that the set of preferences is the same for all agents, all definitions and results would be correct and straightforwardly obtained if we allowed agents' sets of preferences to be different. We assume equal sets of preferences since this is the case of our application in Section 3.

Let us define some incentive-related properties of social choice functions. The best known non-manipulability axiom is that of strategy-proofness. In its usual form it requires the truth to be a dominant strategy for each agent. However, we provide a more general definition which encompasses strategy-proofness and also considers the option that several agents evaluate the possibility of joint deviations.

Definition 1. Let f be a social choice function on $\mathcal{U}^n$. Let $R_N \in \mathcal{U}^n$ and $C \subseteq N$. A subprofile $R'_C \in \mathcal{U}^c$ such that $R'_i \neq R_i$ for all $i \in C$ is a **profitable deviation** of coalition C against profile R_N (for f) if $f(R'_C, R_{N\setminus C}) P_i f(R_N)$ for any agent $i \in C$.

Profitable deviations are usually called (group) manipulations in the standard definitions of group and individual strategy-proofness. Throughout the paper we shall assume that among profitable deviations for single agents there is always one that is best.[3]

Definition 2. A social choice function f on $\mathcal{U}^n$ is manipulable at $R_N \in \mathcal{U}^n$ by coalition $C \subseteq N$ if there exists a profitable deviation of coalition C against profile R_N, say $R'_C \in \mathcal{U}^c$. A social choice function is **group strategy-proof** if it is not manipulable by any coalition $C \subseteq N$.

When we consider only deviations by single agent coalitions we have strategy-proofness.

Definition 3. A social choice function f on $\mathcal{U}^n$ is manipulable at $R_N \in \mathcal{U}^n$ by agent $i \in N$ if there exists a profitable deviation of agent i against profile R_N, say $R'_i \in \mathcal{U}$. A social choice function is **strategy-proof** if it is not manipulable by any agent $i \in N$.

Remark that, formally, strategy-proofness is a much weaker condition than group strategy-proofness in any of its versions. In many environments and in spite of this definitional gap, individual strategy-proof rules end up also being group strategy-proof.[4] But, of course, in many other situations this equivalence may not hold, and even when there are attractive strategy-proof rules, they are open to manipulation by groups. In this paper, we concentrate on a form of manipulation that is intermediate between those of individual and group strategy-proofness and that is based on the notion of credible profitable deviations, those where no agent in the deviating coalition can gain by not declaring those preferences she was supposed to use as part of the group strategy. Formally,

Definition 4. Let f be a social choice function on $\mathcal{U}^n$. Let $R_N \in \mathcal{U}^n$ and $C \subseteq N$. We say that $R'_C \in \mathcal{U}^c$ a profitable deviation of C against R_N is **credible** if for all $i \in C$ and all $\bar{R}_i \in \mathcal{U}$, then $f(R'_C, R_{N\setminus C}) R_i f(\bar{R}_i, R'_{C\setminus\{i\}}, R_{N\setminus C})$.

On other terms, a profitable deviation by C from $R_N = (R_C, R_{N\setminus C})$ is credible if R'_C is a Nash equilibrium of the game among agents in C, when these agents strategies are their admissible preferences and the outcome function is $f(\cdot, R_{N\setminus C})$.

3 The existence of a best deviation is guaranteed when the number of alternatives, and those of preferences are finite. Moreover, the condition will also hold under standard assumptions.

4 See Le Breton and Zaporozhets (2009) and Barberà et al. (2010, 2016).