NICEP Working Paper: 2017-05

On the different forms of individual and group strategic behaviour, and their impact on efficiency

Salvador Barberà

Dolors Berga

Bernardo Moreno

Nottingham Interdisciplinary Centre for Economic and Political Research
https://nicep.nottingham.ac.uk/
School of Politics, The University of Nottingham, Law & Social Sciences Building, University Park, Nottingham, NG7 2RD
ISSN 2397-9771

On the different forms of individual and group strategic behaviour, and their impact on efficiency

Salvador Barberà, Dolors Berga and Bernardo Moreno

NICEP Working Paper Series 2017-05
February 2017
ISSN 2397-9771

Salvador Barberà
Autonomous University of Barcelona
Salvador.barbera@uab.cat

Dolors Berga
University of Girona
Dolors.berga@udg.edu

Bernardo Moreno
University of Malaga
bernardo@uma.es

On the different forms of individual and group strategic behavior, and their impact on efficiency

by

Salvador Barberà*, Dolors Berga†, and Bernardo Moreno‡

February 13, 2017

*Corresponding author. MOVE, Universitat Autònoma de Barcelona, and Barcelona GSE. Mailing address: Departament d'Economia i d'Història Econòmica, Edifici B, 08193 Bellaterra, Spain. E-mail: salvador.barbera@uab.cat

†Departament d'Economia, C/ Universitat de Girona, 10; Universitat de Girona, 17003 Girona, Spain. E-mail: dolors.berga@udg.edu

‡Departamento de Teoría e Historia Económica, Facultad de Ciencias Económicas y Empresariales, Campus de El Ejido, 29071 Málaga, Spain. E-mail: bernardo@uma.es

Abstract: We survey a number of results regarding incentives and efficiency that have been recently added to the social choice literature, and we establish parallels and differences with the concepts and results obtained so far by co-utility theory. Our main purpose is to facilitate the convergence between researchers who end up dealing with similar issues in response to rather different motivations and backgrounds.

Journal of Economic Literature Classification Numbers: C70, D71, D82.

Keywords: Voting rules, strategy-proofness, group strategy-proofness, credible manipulation, efficiency, domain restrictions, co-utility.

1 Introduction

This paper is an attempt to approximate the work of social choice theorists to that of computer scientists, prompted by the emergence of the attractive concept of co-utility (see Domingo-Ferrer et al., 2016 and 2017).

Although surprising some decades ago, the connection between the economists' approach to mechanism design and the concerns of computer scientists is now a well-established fact, even if there are still communication barriers based on mutual ignorance and differences in language and motivation. But the appearance of new fields like that of computational mechanism design (see, for example, Nisan, 2007) witnesses the tendency to converge between disciplines that have bumped into similar problems from rather different starting points. We, the authors of the present essay, are economists working in the area of mechanism design, and mostly within the subfield of social choice. And, again, there is a growing literature on computational social choice (see Brandt et al., 2015 for an introduction) that enriches the interaction between the same two scientific communities. This literature has its own concerns and meeting points, but we are not aware of any previous effort to put together the advances regarding co-utility with results in our area that ring very similar bells. Hence, our purpose is to describe a line of work in social choice theory whose motivation and conclusions we can clearly describe from the economists' point of view, and to comment, much more tentatively, on the analogies and the still existing differences between the work that we survey and the research on co-utility. Although very fragmentary, we hope that this essay will be informative for computer scientists and also useful to economists, as a reminder of how worthy it is to connect our partial knowledge with that of other fields.

Let us spell out some of the limits of this essay. We shall present results whose main interpretation refers to voting systems, to be used for the collective choice of candidates, or the determination of levels of provision of public goods. Most of the issues we raise and of the concepts we introduce can be extended to other types of decisions involving private goods, but for the sake of simplicity we stick to the traditional social choice framework. This makes the comparison with co-utility a bit harder, since in principle that theory does not distinguish between these two subcases, and we explicitly exclude in the present work references to auctions, or to other very central concerns involving private goods like allocation of indivisible goods or matching. But we hope it provides a good enough starting point, and that it makes our presentation specific.

We also acknowledge that we shall not go as far as to provide a complete reconciliation between the results we present and the concepts, techniques and results of co-utility theory. Rather, we'll try to point out what we believe to be common concerns, and what we still feel are possible divergences.

The possibility of strategic voting has been discussed since ancient times (see the references to Pliny the Younger, Ramon Llull and later authors in the wonderful book by Mc Lean and Urken, 1995; a brief summary of that history is provided in Barberà, 2010). At the early ages of modern social choice theory, it was mentioned but not analyzed by Arrow (1963). Early contributors to its study were Farquharson (1969), Vickrey (1960) and Pattanaik (1976a, 1976b), but the real start of a thorough analysis of manipulation and strategy-proofness in voting came with the works of Gibbard (1973) and Satterthwaite

(1975). Their main result proved that any non-trivial social choice function (or voting rule, depending on the interpretation) is manipulable, as soon as we must choose from more than two alternatives and the preferences of agents are unrestricted. This result generated two large streams of literature. One concentrated on the idea that manipulation per se is nothing but the play of a game, and studied the consequences of having agents behave strategically. Implementation theory emerged from that, in all the richness that derives from considering a variety of possible games and applying to them different solution concepts. The other stream, where we place the results that follow, interpreted the negative result of Gibbard and Satterthwaite as an invitation to concentrate on the specifics of each relevant case, and to characterize those situations where one could expect restrictions on agent's preferences to hold and to allow for the existence of attractive strategy-proof rules. Two such examples are provided by environments where agents' preferences can be assumed to be single-peaked, or others where alternatives consist of collections of objects and preferences satisfy separability conditions. In these environments, and under the preference restrictions that they suggest, one can identify many strategy-proof social choice functions. And this rings a bell regarding co-utility's requirement that individual agents should follow a prescribed protocol. However, one can ask for more, and not all environments admitting strategy-proof rules allow for the additional requirement that groups of agents should also find it beneficial to abide to the rules. Group strategy-proofness, or some intermediate concept qualifying the joint behavior of voters, becomes an additional objective for the social choice theorist, as well as an analysis of the relative efficiency of those rules that satisfy such conditions. Until now co-utility has only demanded individual incentives to be appropriate, and not insisted directly on the strategic behavior of groups. Its concern with efficiency, however, relates to the possibility of avoiding joint manipulations by the set of all players, and is therefore quite connected with the type of group behavior that we shall study here. Hence our interest in exposing what we know as social choice theorists, to suggest how far the analogies with co-utility go, and to express the need we still feel for further approaches.

2 The general framework

Let A be the set of *alternatives* and $N = \{1, 2, ..., n\}$ be the set of *agents* (with $n \geqslant 2$). Let capital letters $S, T \subset N$ denote subsets of agents while lower case letters s, t denote their cardinality.

Let $\mathcal{R}$ be the set of complete, reflexive, and transitive orderings on A and $\mathcal{R}_i \subseteq \mathcal{R}$ be *the set of admissible preferences for agent* $i \in N$. A *preference profile,* denoted by $R_N = (R_1, ..., R_n)$, is an element of $\times_{i \in N} \mathcal{R}_i$. As usual, we denote by P_i and I_i the strict and the indifference part of R_i, respectively. Let $C, S \subset N$ be two coalitions such that $C \subset S$. We will write the subprofile $R_S = (R_C, R_{S \backslash C}) \in \times_{i \in S} \mathcal{R}_i$ when we want to stress the role of coalition C in S. Then the subprofiles $R_C \in \times_{i \in C} \mathcal{R}_i$ and $R_{S \backslash C} \in \times_{i \in S \backslash C} \mathcal{R}_i$ denote the preferences of agents in C and in $S \backslash C$, respectively.

A *social choice function* (or a *rule*) is a function $f : \times_{i \in N} \mathcal{R}_i \to A$.

All along the paper we will focus on rules that are non-manipulable, either by a single agent or by a coalition of agents. We first define what we mean by a manipulation and then we introduce the well known concepts of *strategy-proofness* and *group strategy-proofness*.

Definition 1 *A social choice function f is (strongly) group manipulable on $\times_{i\in N}\mathcal{R}_i$ at $R_N \in \times_{i\in N}\mathcal{R}_i$ if there exists a coalition $C \subset N$ and $R'_C \in \times_{i\in C}\mathcal{R}_i$ ($R'_i \neq R_i$ for any $i \in C$) such that $f(R'_C, R_{-C})P_i f(R_N)$ for all $i \in C$. We say that f is individually manipulable if there exists a possible manipulation where coalition C is a singleton.*

Definition 2 *A social choice function f is (weakly) group strategy-proof on $\times_{i\in N}\mathcal{R}_i$ if f is not (strongly) group manipulable for any $R_N \in \times_{i\in N}\mathcal{R}_i$. Similarly, f is strategy-proof if it is not individually manipulable.*

Definition 3 *A social choice function f is (weakly) efficient if for any $R_N \in \times_{i\in N}\mathcal{R}_i$, there is no alternative a such that $aP_i f(R_N)$ for all $i \in N$.*

Definition 4 *A social choice function f is strongly efficient if for any $R_N \in \times_{i\in N}\mathcal{R}_i$, there is no alternative a such that $aR_i f(R_N)$ for all $i \in N$ and $aP_j f(R_N)$ for some $j \in N$.*

Note that (weak) group strategy-proofness implies (weak) efficiency but the converse does not hold. For example, in the first application defined in Section 3, the rule choosing the mean of the peaks is an efficient rule, but it is manipulable.

In our definitions above, and at different points in the text, we qualify different notions as being weak or strong depending on whether strict domination occurs for all agents involved, or only for some. Since most of the work we are going to discuss assumes for simplicity that all agents have strict preferences, the qualification will not in general be needed here, although it is important in other contexts, as those where private goods are present, where it becomes quite essential. Another qualification we want to make regards the definition of efficient rules. Typically we consider social choice functions whose range covers the whole set of alternatives. When applied to rules that are not onto, the efficiency notion that we apply restricts attention to those alternatives that are in the range.

3 Possibility and impossibility results in designing strategy-proof social choice rules

We first present Gibbard and Satterthwaite's on the impossibility of designing non-trivial strategy-proof rules when individual preferences are non-restricted. After that, we present three results characterizing classes of strategy-proof rules that may be defined in different settings where agents' preferences are appropriately restricted given the type of problem under consideration. These results will provide reference points to discuss the connections between incentives and efficiency in the following section.

Definition 5 *A social choice function is dictatorial if there exists an agent $d \in N$ such that for any preference profile $R_N \in \mathcal{R}^n$, $f(R)$ is such that $f(R_N)R_d x$, $\forall x \in A$.*

Theorem 1 (*Gibbard, 1973 and Satterthwaite, 1975*) *Any social choice function f on $\mathcal{R}^n$ with at least three alternatives in the range is either manipulable or dictatorial.*

If we admit that dictatorial rules are trivial, and we also consider as such those that only select among two alternatives, the result expresses the impossibility of designing non-trivial rules operating on the universal domain of preferences.

But then, starting from this impossibility, one may ask whether there exist environments where it is natural to expect agents' preferences to be restricted, and where, given these restrictions, strategy-proof rules may be defined within these natural settings. Much work has been done in characterizing restricted domains of preferences where attractive and strategy-proof rules can be found. We present three cases where such environments have been identified and strategy-proof rules characterized: (1) the choice of candidates or levels of provision of pure public goods when agents evaluate alternatives in one dimension and exhibit single-peaked preferences, (2) the choice of multiple candidates to join a club, when voters' preferences are separable, and (3) the choice of levels for different characteristics of candidates, where these can be seen as points in a grid, and the preferences of voters satisfy a generalized version of single-peakedness and separability. The latter case subsumes the preceding two, but presenting each case in turn will facilitate our exposition.

3.1 One dimensional decisions under single-peaked preferences

Let A be an interval in $\mathbb{R}$, and consider a linear order of A, say $>$. We now define the concept of a single-peaked preference on A (given $>$).

Definition 6 *A preference $R_i \in \mathcal{R}_i$ is* ***single-peaked*** *on A if (1) there exists $\tau(R_i) \in A$ such that $\tau(R_i)\, P_i y$ for any $y \in A$, and (2) for any $y, z \in A$, $[z < y < \tau(R_i)$ or $z > y > \tau(R_i)]$ then $y P_i z$.*

Let $\mathcal{SP}$ be the set of individual single-peaked preferences on A

Definition 7 *A social choice function $f : \mathcal{SP}^n \to A$ is a generalized Condorcet-winner rule if, for any profile $R_N \in \mathcal{SP}^n$, $f(R_N) = median\{\tau(R_1), ..., \tau(R_n), p^1, ..., p^{n-1}\}$, where $P(f) = \{p^1, ..., p^{n-1}\}$ is a list of parameters in A.*

We will consider rules satisfying the additional property of anonymity. Formally,

Definition 8 *A social choice function $f : \mathcal{SP}^n \to A$ is anonymous if for any $R_N \in \mathcal{SP}^n$ and for any permutation of agents $\sigma : N \to N$, $f(R_N) = f(R_{\sigma(N)})$.*

That is, anonymity says that the name of the agents involved in the social decision does not matter; only the preference profile matters, not the particular opinion of an individual.

Theorem 2 *(Moulin, 1980) A social choice function f on $\mathcal{SP}^n$ is anonymous, onto, and strategy-proof if and only if it is a generalized Condorcet-winner rule.*

This is not the most general characterization result in the single-peaked context, but we add the condition of anonymity for simplicity and because anonymous rules are of special interest. The reader may appreciate that choosing the median voter's best alternative is a special case of those rules that satisfy strategy-proofness in the single-peaked domain.

3.2 Choosing multiple candidates for a club

Let $A = 2^{\mathcal{O}}$, where $\mathcal{O} = \{o_1, ..., o_K\}$ is a finite set of objects. Let individual preferences be linear orders on $2^{\mathcal{O}}$ (including the empty set). Define the set of "good" objects as $G(K, R) = \{o_k \in \mathcal{O} : \{o_k\} P \varnothing\}$, and the set of "bad" objects as $K \backslash G(K, R) = \{o_k \in \mathcal{O} : \varnothing P \{o_k\}\}$.

Definition 9 *R is an individual separable preference on $2^{\mathcal{O}}$ if and only if for any set T and any object $o_l \notin T$, $T \cup \{o_l\} PT$ if $o_l \in G(K, R)$.*

Let $\mathcal{S}$ denote the set of all separable preferences.

Definition 10 *Let $q \in \{1, ..., n\}$. The social choice function f on $\mathcal{S}^n$ defined so that for any $R_N \in \mathcal{S}^n$,*

$$f(R_N) = \{o_k \in \mathcal{O} : |\{i : o_k \in G(K, R_i)\}| \geq q\}$$

is called voting by quota q.

These rules turn to be important because they have been obtained as the unique class satisfying desirable properties in some problems. One of these properties is neutrality.

Definition 11 *A social choice function $f : \mathcal{S}^n \to 2^{\mathcal{O}}$ is neutral if, for any $R_N \in \mathcal{S}^n$ and for any permutation of objects $\mu : \mathcal{O} \to \mathcal{O}$, $f(\mu(R_1), \mu(R_2), ..., \mu(R_n)) = \mu(f(R_N))$, where for each $j \in N$, $\mu(R_j)$ is the preference obtained from R_j by permuting the objects according to μ.*

Neutrality is a general property that social choice functions may or may not satisfy. In general terms, it requires that all alternatives should be treated equally by the rule, and that their labelling is irrelevant. We use it here for simplicity, although more general results are also known. Notice that, in the case we consider here, alternatives are sets of objects, hence our notation.

Theorem 3 *(Barberà, Sonnenschein, and Zhou, 1991) A social choice function f on $\mathcal{S}^n$ is anonymous, neutral, onto, and strategy-proof if and only if it is a voting by quota rule.*

Notice again that this is not the most general characterization theorem for strategy-proof rules in that environment, but that restricting attention to neutral and anonymous rules allows us to point at particularly simple ones: those where each agent votes for those candidates that she deems good, and all candidates who get at least q votes are elected.

3.3 Choosing from a grid

Let $\mathcal{K}$= $\{1, ..., K\}$ be a finite set of $K \geq 2$ coordinates, and $B_k = [a_k, b_k]$, $a_k < b_k$ be an integer interval of possible values for each of the coordinates. Then, $A = \prod_{k=1}^{K} B_k$ is the set of alternatives which are K-dimensional vectors. We endow A with the L_1-norm, that is, for any $x \in B$, $\|x\| = \sum_{k=1}^{K} | x_k |$. Given $x, y \in B$, the *minimal box containing x and y* is defined by

$$MB(x, y) = \{z \in B : \|x - y\| = \|x - z\| + \|z - y\|\}.$$

Definition 12 *A preference $R_i \in \mathcal{R}_i$ is (multidimensional)* ***single-peaked*** *if f (1) there exists $\tau(R_i) \in A$ such that $\tau(R_i) P_i y$ for any $y \in A$, and (2) for any $z, y \in B$, if $y \in MB(z, \tau(R_i))$ then $y R_i z$*

Let $\mathcal{S} \subseteq \mathcal{R}_i$ be the set of (multidimensional) single-peaked preferences on A. Define $\tau(R_i) = (\tau_1(R_i), ..., \tau_K(R_i)) \in A$, where $\tau_k(R_i)$ is the best (or top) alternative of R_i in dimension k.

As announced, notice that this environment includes the two previous ones as special cases, where only one dimension is allowed, or else only two possible values are possible for each dimension. In fact, the following definition encompasses the characteristics of the two types of mechanisms that we have characterized in the previous results.

Definition 13 *A social choice function $f : \mathcal{S}^n \to A$, $f = (f_1, ..., f_K)$ is a (multidimensional) generalized Condorcet-winner rule if for any $R_N \in \mathcal{S}^n$, $k \in \mathcal{K}$, $f_k(R_N) = median\{\tau_k(R_1), ..., \tau_k(R_n), p_k^1, ..., p_k^{n-1}\}$, where $P_k(f) = \{p_k^1, ..., p_k^{n-1}\}$ is a list of parameters in B_k.*

Note that in the one-dimensional case, that is for $K = 1$, these rules precipitate to the rules in Definition 7.

Theorem 4 *(Barberà, Gul, and Stacchetti, 1993) A social choice function f on $\mathcal{S}^n$ is anonymous, onto, and strategy-proof if and only if it is a generalized Condorcet-winner rule.*

We close this Section by emphasizing that, while the rules we have presented are all strategy-proof, they differ very much regarding their efficiency, for the case in which agents vote straightforwardly. The generalized Condorcet-winner rules are efficient in the one-dimensional case. What is more: not only they are immune to profitable deviations by the set of all agents, but also to profitable deviations by groups of agents: In the language we have already presented in Section 2, they are group strategy-proof, which implies they are efficient.

The following example shows that voting by quota q is an inefficient rule in the separable domain, and a fortiori, that multidimensional generalized Condorcet-winner also fail to satisfy efficiency.

Example 1 *Let us use quota 2 for electing subsets of a three candidate election with three voters, in a separable environment. Let $N = \{1, 2, 3\}$, $K = 3$ and consider voting by quota 2. Let R_N be as follows: the preferences of agent 1 are such that $\tau(R_1) = \{o_1, o_2\}$ and $\varnothing P_1 \{o_1, o_2, o_3\}$, the preferences of agent 2 are such that $\tau(R_2) = \{o_2, o_3\}$ and $\varnothing P_2 \{o_1, o_2, o_3\}$, and the preferences of agent 3 are such that $\tau(R_3) = \{o_1, o_3\}$ and $\varnothing P_3 \{o_1, o_2, o_3\}$. Observe that $f(R_N) = \{o_1, o_2, o_3\}$. This outcome is clearly inefficient.*

The reader will notice that the example can be generalized for any possible quota q, any $K \geq 2$ and any number of agents.

4 Equivalence between individual and group strategy-proofness

It is clear from their definitions that group implies individual strategy-proofness, but not the other way around. Yet in many interesting applications, once we have a function that satisfies the weak requirement, it also meets the stronger one. For other equally interesting applications, the equivalence does not hold. Let us provide some examples. In public goods settings, the ones we shall deal with in this essay, generalized Condorcet-winner rules are both individual and group strategy-proof under single-peaked preferences and one dimensional spaces. Both properties also hold for properly chosen social choice functions in one dimensional cases when preferences are single-dipped. And although we shall not present these results here, the equivalence between individual and group strategy-proofness also holds for interesting classes of social choice functions defined in contexts with private goods. This is the case, for example, for settings involving matching, rationing, house allocation cost sharing or simple auctions.

In contrast, other interesting environments admitting strategy-proof mechanisms allow for these mechanisms to be manipulable by groups. These include two of the examples we present here for the case of public goods, involving multi-dimensional choices, and, in the case of environments with private goods and quasi-linear preferences, the well-known family of Vickrey-Clarke-Groves mechanisms. In fact, in our examples group strategy-proofness is violated in a strong manner, and the only rules that are strategy-proof do not even satisfy efficiency.

In view of this dichotomy, we have investigated whether the equivalence, when it holds, needs a case-by-case explanation, or is the result of some common features underlying the different models. In what follows, we present a result that identifies such common features for the case of public goods, and proves that they precipitate the equivalence. A similar result also holds for many apparently diverse environments in private goods settings (Barberà, Berga, and Moreno, 2016).

We end this section by stating the unifying result in our simple public goods setting. In the following section we shall turn attention to the properties of those rules that are strategy-proof but hold that property in environments where group manipulation remains a possibility.

Let $R_N \in \times_{i\in N}\mathcal{R}_i$, and y, z be a pair of alternatives. Denote by $S(R_N; y, z) \equiv \{i \in N : yP_iz\}$, that is, the set of agents who strictly prefer y to z according to their individual preferences in R_N. Moreover, define the concept of lower contour set. For any $x \in A$ and $R_i \in \mathcal{R}_i$, *the lower contour set of* R_i *at* x as $L(R_i, x) = \{y \in A : xR_iy\}$. Similarly, the strict lower contour set at x is $\overline{L}(R_i, x) = \{y \in A : xP_iy\}$.

Definition 14 *Given a preference profile $R_N \in \times_{i\in N}\mathcal{R}_i$ and a pair of alternatives $y, z \in A$, we define a binary relation $\succsim (R_N; y, z)$ on $S(R_N; y, z)$ as follows:*

$$i \succsim (R_N; y, z) j \text{ if } L(R_i, z) \subseteq \overline{L}(R_j, y)$$

Definition 15 *A preference profile $R_N \in \times_{i\in N}\mathcal{R}_i$ satisfies sequential inclusion for $y, z \in A$ if the binary relation $\succsim (R_N; y, z)$ on $S(R_N; y, z)$ is complete and acyclic.*

In Barberà, Berga, and Moreno (2010) we prove the following equivalence result.

Theorem 5 *Let $\times_{i\in N}\mathcal{R}_i$ be a domain such that any preference profile satisfies sequential inclusion. Then, any strategy-proof rule f on that domain is also group strategy-proof.*

Admittedly, sequential inclusion is a strong condition. Our emphasis here is on the fact that, however restrictive, it identifies environments and social choice rules for which the objectives of co-utility are met in a very strong sense. Incentives for individuals and for groups are perfectly aligned, and efficiency is attained.

While Theorem 5 qualifies the quite generalized belief that incentive compatibility and efficiency are always incompatible, the tension between these two objectives remains in other contexts. In the next section we study some of the consequences that derive from operating under strategy-proof mechanisms that are not immune to group manipulations.

5 A weaker notion to control manipulation by groups

The previous characterization results have proven that there is a large gap between those environments in which individual strategy-proofness implies group strategy-proofness and those where this equivalence does not hold. In the first case, a reconciliation between the two main objectives of co-utility is automatic, while in the second case it is problematic, to say the least.

We shall follow here Barberà, Berga, and Moreno (2017) and study conditions that are intermediate between individual and group strategy-proofness. That will allow us to classify the strategy-proof rules in the multidimensional case into two different classes, according to the credibility that one can attach to the manipulative actions of groups.

Definition 16 *Let f be a social choice function on $\times_{i\in N}\mathcal{R}_i$. Let $R_N \in \times_{i\in N}\mathcal{R}_i$ and $C \subseteq N$. A subprofile $R'_C \in \times_{i\in C}\mathcal{R}_i$ such that $R'_i \neq R_i$ $\forall i \in C$ is a* ***profitable deviation*** *of coalition C against profile R_N if $f(R'_C, R_{N\setminus C}) P_i f(R_N)$, for any agent $i \in C$.*

Definition 17 *Let f be a social choice function on $\times_{i\in N}\mathcal{R}_i$. Let $R_N \in \times_{i\in N}\mathcal{R}_i$ and $C \subseteq N$. We say that $R'_C \in \times_{i\in C}\mathcal{R}_i$ a profitable deviation of C against R_N is* ***credible*** *if for all $i \in C$ and all $\overline{R}_i \in \mathcal{R}_i$, then $f(R'_C, R_{N|C}) R_i f(\overline{R}_i, R'_{C\setminus\{i\}}, R_{N\setminus C})$.*

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 $f(\cdot, R_{N\setminus C})$.

Definition 18 *A social choice function f on $\times_{i\in N}\mathcal{R}_i$ is* ***immune to credible deviations*** *if for any $R_N \in \times_{i\in N}\mathcal{R}_i$ and $C \subseteq N$, there is no credible profitable deviation of C against R_N (that is, for any profitable deviation $R'_C \in \times_{i\in C}\mathcal{R}_i$ of C against R_N, there exists an agent $i \in C$ such that $f(\overline{R}_i, R'_{C\setminus\{i\}}, R_{N\setminus C}) P_i f(R'_C, R_{N|C})$ for some $\overline{R}_i \in \mathcal{R}_i$).*

Although formulated in a different manner, this concept of immunity turns out to be equivalent, in the contexts that we describe, to the notion of strong coalition-proofness proposed by Peleg and Sudhölter (1999). In their case, they described the concept of strong coalition proof equilibrium and then proposed that a social choice function be required to have truthful revelation as an equilibrium of their kind (see also Peleg, 1998).

Note that immunity implies strategy-proofness (assuming the existence of the best profitable deviations for single agents). However the converse does not hold in general.

Lemma 1 *Any social choice function f on $\times_{i\in N}\mathcal{R}_i$ that is immune to credible deviations is strategy-proof.*

By definition we can observe that any generalized Condorcet-winner rule on the single-peaked domain is both (group) strategy-proof and immune to credible deviations. Moreover, any voting by quota q is strategy-proof on the separable domain but only voting by quota 1 and n are immune to credible deviations, as we proved in Barberà, Berga, and Moreno (2017). On the (multidimensional) single-peaked domain, any multidimensional generalized Condorcet-winner is strategy-proof but only those whose list of parameters are degenerate in at least $K-1$ dimensions are immune to credible deviations, as we also proved in the mentioned paper.[1]

In the case of choosing candidates we have the following results.

Proposition 1 *Let $n>3$. Then, voting by quota 1 and n are the only voting by quota rules satisfying immunity to credible deviations.*[2]

In the case of choosing from a grid, here are the results.

Proposition 2 *Let $n>3$. Let f be a generalized Condorcet-winner rule. If f is defined by lists of parameters that are non-degenerate in at least two dimensions, then f is not immune to credible deviations.*

Proposition 3 *Let $n\geq 2$. Let f be a generalized Condorcet-winner rule. If f is defined by lists of parameters that are degenerate in at least $K-1$ dimensions, then f is immune to credible deviations.*[3]

Since the notion of credibility may have different expressions, we shall offer now some alternatives to our main definition, in addition to the already mentioned possibility of reformulating it in terms of strong coalition-proof equilibria.

[1] A list of parameters is degenerate if they are all equal.

[2] The case $n=2$ gives the same results as $n>3$. The case $n=3$ requires special treatment. We obtain that when $n=3$ and $K=2$, any voting by quota rule is immune to credible deviations. When $n=3$ and $K\geq 3$, voting by quota 1 and 3 are the only voting by quota rules satisfying immunity to credible deviations.

[3] The case $n=3$ requires special treatment and the result obtained is the following: Any generalized Condorcet winner rule defined by lists of parameters such that are non-degenerate in two dimensions is immune to credible deviations. Any generalized Condorcet winner rule defined by non-degenerate lists of parameters in at least three dimensions is not immune to credible deviations.

The first variant will be one where, instead of letting agents in C to have any choice of preferences as a strategy, we restrict them to either use strategy R'_i or to revert to strategy R_i. The resulting notion of a credible deviation will be stronger than ours. However, as shown below the set of rules that are immune to credible deviations will be the same (after a minimal qualification) under either definition.

Definition 19 *Let f be a social choice function on $\times_{i\in N}\mathcal{R}_i$. Let $R_N \in \times_{i\in N}\mathcal{R}_i$ and $C \subseteq N$. We say that $R'_C \in \times_{i\in C}\mathcal{R}_i$ a profitable deviation of C against R_N is (type 1)* ***credible*** *if $f(R'_C, R_{N|C})R_i f(R_i, R'_{C\setminus\{i\}}, R_{N\setminus C})$ for all $i \in C$. A social choice function f on $\times_{i\in N}\mathcal{R}_i$ is* ***immune to*** *(type 1)* ***credible deviations*** *if for any $R_N \in \times_{i\in N}\mathcal{R}_i$, any $C \subseteq N$, there is no (type 1) credible profitable deviation of C against R_N.*

Proposition 4 *A social choice function f on $\times_{i\in N}\mathcal{R}_i$ is immune to credible deviations if and only if it is immune to (type 1) credible deviations.*

A second variant will require that in order to be (extensively) credible, deviation R'_C should be a Nash equilibrium for the game where all agents can play any preference, f being the outcome function. If f is strategy-proof, then again the set of immune rules will still be the same under either definition. Otherwise, the equivalence is not true.

Definition 20 *Let f be a social choice function on $\times_{i\in N}\mathcal{R}_i$. Let $R_N \in \times_{i\in N}\mathcal{R}_i$ and $C \subseteq N$. We say that $R'_C \in \times_{i\in C}\mathcal{R}_i$ a profitable deviation of C against R_N is (type 2)* ***credible*** *if $f(R'_C, R_{N|C})R_i f(\overline{R}_i, R'_{C\setminus\{i\}}, R_{N\setminus(C\cup\{i\})})$ for all $i \in N$ and all $\overline{R}_i \in \mathcal{R}_i$. A social choice function f on $\times_{i\in N}\mathcal{R}_i$ is* ***immune to*** *(type 2)* ***credible deviations*** *if for any $R_N \in \times_{i\in N}\mathcal{R}_i$, any $C \subseteq N$, there is no (type 2) credible profitable deviation of C against R_N.*

Proposition 5 *Any strategy-proof social choice function f on $\times_{i\in N}\mathcal{R}_i$ is immune to credible deviations if and only if it is immune to (type 2) credible deviations.*

A third variant of our definition of credibility would result from simply changing our original one, but ask the deviation to be a strong Nash, rather than a Nash equilibrium. The rationale for such proposal would be to allow for several agents to coordinate when defecting from the agreed upon joint manipulation.

Definition 21 *Let f be a social choice function on $\times_{i\in N}\mathcal{R}_i$. Let $R_N \in \times_{i\in N}\mathcal{R}_i$ and $C \subseteq N$. We say that $R'_C \in \mathcal{U}^c$ a profitable deviation of C against R_N is* ***strongly credible*** *if $f(R'_C, R_{N|C})R_i f(\overline{R}_S, R'_{C\setminus S}, R_{N\setminus C})$ for all $S \subseteq C$, for all $\overline{R}_S \in \times_{i\in S}\mathcal{R}_i$ and for some $i \in S$. A social choice function f on $\times_{i\in C}\mathcal{R}_i$ is* ***immune to strongly credible deviations*** *if for any $R_N \in \times_{i\in C}\mathcal{R}_i$, any $C \subseteq N$, there is no strongly credible profitable deviation of C against R_N.*

Proposition 6 *All generalized Condorcet-winner rules are immune to strongly credible deviations.*

After this digression showing the connections between different possible definitions of credibility and their consequences, the following comments refer to our main definition of immunity, as presented in Definition 17.

The distinction between those social choice functions which are immune to credible deviations and those that are not is very relevant for our purpose of connecting our results with those in co-utility theory. In some sense, one may think that the most attractive rules are those that are not only strategy-proof but also immune to credible deviations. If we take as an example the choice of candidates to a club under separable preferences, and we push this point of view, then quota one or quota n would stand out as being especially attractive. However, let us turn the argument around and ask what would be the performance of the rules based on voting by quota q, for values of q higher than one and lower than n.

In that case, we know that there will be some cases where groups of agents can credibly coalesce for their joint profit.

Let us see what may happen in a simple case, which can be easily extended to other cases.

Example 2 (Example 1, continued) *Lut us use quota 2 for electing subsets of a three candidate election with three voters, in a separable environment. When the voter's preferences are R_N as in Example 1: that is, agent 1 likes o_1 and o_2 but hates o_3, 2 likes o_2 and o_3 but hates o_1, 3 likes o_1 and o_3 but hates o_2, and all of them prefer not having any candidate to have all of them, the outcome from sincere voting would be all candidates, while in fact all agents would prefer no candidate being elected. Hence, the rule clearly yields inefficient outcomes. One way in which the three voters could depart beneficially from sincere voting would be by none of them giving support to any candidate. Their unanimous vote for the empty set would not only give them a result that Pareto improves upon the outcome resulting from sincere voting, but is credible, because no agent could deviate from the agreed upon joint voting strategy and gain from it.*

As we already observed, the outcome of voting by quota 2 under the example's preferences is inefficient, because all agents prefer the empty set to the set with all three alternatives which would result from truthful preference revelation. We now remark that their unanimous vote for the empty set would not only give them a Pareto superior result, but also constitute a credible deviation from the truth, since no agent alone could profitably deviate from that joint strategy profile, which is a Nash equilibrium.

The example suggests that, if we select rules that are credibly manipulable, and allow groups of agents that have credible joint strategies to use them, the induced connection between individual preferences and final outcomes, after manipulation, define an efficient social choice function.

Moreover, notice that none of the agents can be distinguished from the others by observing the votes, and that none of the differential features of the voters are revealed.

Hence, we can generate a result that is very close to the desiderata of co-utility, as we shall comment more extensively in the conclusions below.

Here is a more general argument and construct in the same spirit. Take a rule f that is inefficient and admits credible profitable deviations by the grand coalition. Look at the rule f' defined as follows. For those profiles P where no group has a credible profitable deviation, let $f(P) = f'(P)$. Now, for those profiles where there are one or more groups that could credibly profitably deviate, choose a specific one among those that involve the largest number of deviants, and if necessary pick it in such a way that it gives maximal gains to the agent with the smallest indicator, in lexicographic order. Then, let the value of f' be the resulting one from applying the function f to the profile where the agents in the coalition have profitably deviated, while the rest of agents still vote truthfully. Notice that the resulting function is efficient.

6 Conclusions

In this section we comment on the analogies and the differences that we find between our results on the manipulability of social choice functions and the purposes and findings in the of co-utility context.

First of all, there is the obvious analogy that in both cases we are trying to devise a mechanism, represented as a game form that gives rise to a game for each specification of the agents' objectives.

We concentrate on the case of public decisions taken without the possibility of money compensation. Although this decision excludes important parts of the literature on mechanism design, like auction theory, it has the advantage of showing that the issues we are dealing with arise in even the most basic settings.

The two basic notions underlying the design of co-utile protocols can be qualified in several ways. The idea of self-enforcement will depend on the type of game we assume agents will play, and on the associated notion of equilibrium. Our results are expressed for simultaneous games, where protocols are very simple: all agents are required to submit their information at the same time in one shot. While this formulation may be less realistic than formulating an extensive form game involving sequential protocols, it also has its advantages. The subgame perfect equilibria of the latter are very sensitive to the order of play, while in our case we can identify environments where all agents can be endowed with dominant strategies, which in our view is a very satisfactory expression of self enforcement.

The use of strategy-proof mechanisms in extensive form is under study by an emerging literature on the concept of obvious strategy-proofness, defined by Li (2016). The paper by Arribillaga, Massó, and Neme (2016) on that subject is particularly close to our present concerns, since it applies to the type of environments we have restricted attention to. We have not been able yet to establish the full connections between that work and co-utility, but we want to call attention toward these very recent papers.

Now, regarding efficiency, there are also possible qualifications, but let us say that, in those cases where we can prove that strategy-proofness implies group strategy-proofness, we can nicely tie the two essential requirements of co-utility into one. This is because group strategy-proofness implies efficiency, when applied to the set of all agents. This is what happens when preferences are restricted to be unidimensional single-peaked.

The more interesting connections arise when we analyze those contexts where strategy-proof rules are not efficient, hence not group strategy-proof, as it happens in the case of multi-dimensional single-peaked environments, and in the more specific case of separable preferences. In that case, we have provided a characterization of those anonymous strategy-proof rules that are immune to profitable deviations, and those that, although strategy-proof, are not immune.

Now, from the point of view of mechanism design, one could claim that the interesting case is the one where agents have no incentive to deviate from the truth, and cannot from profitable coalitions either, since these rules would be strongly self enforceable: no gains could be obtained by not abiding to the rules. But the interesting twist, and the one that approaches our results to co-utility, comes from the opposite point of view. Consider a rule that is not immune to credible deviations among those that we have characterized: for instance take Example 1, the quota two rule for electing subsets of a three candidate election with three voters, in a separable environment. We know that the whole set of agents will be able to credibly manipulate in some cases: Their joint manipulation will take the form of a transaction, where every voter stops supporting some desirable candidates if others stop supporting others that are undesirable for her. Or, vice-versa, if everyone supports some undesirable candidate in exchange for others to help a candidate that they want to be selected. This type of manipulation involves a Pareto improvement, which is not surprising since we have already remarked that anonymous strategy-proof rules in that context are not efficient. Hence, if we change the equilibrium concept that we apply to the voting game in that example, and simply require that manipulations by agents must be a Nash equilibrium, then the mechanism is co-utile, at least in the sense that the equilibria of no other mechanism Pareto dominate those of the one we started with.

Moreover, notice that the type of improvement in question derives from a transaction between groups of agents who can gain from giving up some gains in exchange for some larger ones. And that in order to make this transaction a Nash equilibrium, the agents end up voting in a manner that does not allow to tell each agent from the others, and that none of the agents reveals their true identity. All of this is strongly reminiscent of the privacy preserving character of mechanisms that is sought after by co-utility.

Our general conclusion, then, is that knowledge of results in social choice has allowed us to better understand the ambitions of people working in co-utility, and that our still limited understanding of their goals has led us to re-evaluate some of our findings, and to understand that not all departures from the proposed rules by groups of agents must be treated as undesirable, if they end up approaching the collective results to efficiency. We hope that our attempt to build that bridge meets others, in the directions of interdisciplinary dialog.

Ackowledgements

S. Barberà acknowledges financial support through grants ECO2014-53052-P and SGR2014-515, and Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2015-0563). D. Berga acknowledges the support from grants ECO2016-76255-P, ECO2013-45395-R and 2014-SGR-1360. B. Moreno acknowledges financial support from grants and ECO2014-53767. D. Berga and B. Moreno thank the MOMA network under project ECO2014-57673-REDT.

References

Arribillaga, P., Massó, J., Neme, A.: Not All Majority-based Social Choice Functions Are Obviously Strategy-proof, mimeo UAB (2016)

Arrow, K. J. Social Choice and Individual Values, New Haven and London, Yale University Press (1963)

Barberà, S.: Strategy-proof Social Choice, in K. J. Arrow, A. K. Sen and K. Suzumura (eds.), Handbook of Social Choice and Welfare. Volume 2 . Netherlands: North-Holland, chapter 25, 731-831 (2010)

Barberà, S., Berga, D., Moreno, B.: Individual versus group strategy-proofness: when do they coincide?. J. Econ. Theory 145, 1648-1674 (2010)

Barberà, S., Berga, D., Moreno, B.: Group strategy-proofness in private good economies. American Economic Review 106, 1073–1099 (2016)

Barberà, S., Berga, D., Moreno, B.: Immunity to credible deviations from the truth. Forthcoming in Mathematical Social Sciences (2017)

Barberà S., Gul F., Stacchetti E.: Generalized Median Voter Schemes and Committees. J. Econ. Theory 61, 262-289 (1993)

Barberà, S., Sonnenschein, H., Zhou L.: Voting by committees. Econometrica 59, 595-609 (1991)

Brandt, F., Conitzer,V., Endriss, U., Lang, J., and Procaccia, A.D.: Introduction to Computational Social Choice. Handbook of Computational Social Choice, F. Brandt, V. Conitzer, U. Endriss, J. Lang and A.D. Procaccia (eds.), Cambridge University Press (2015)

Domingo-Ferrer, J., Sánchez, D., Soria-Comas, J.: Co-utility: Self-enforcing collaborative protocols with mutual help, Prog. Artif. Intell. 5, 105-110 (2016)

Domingo-Ferrer, J., Martínez, S., Sánchez, D., Soria-Comas, J.: Co-utility: Self-enforcing protocols for the mutual benefit of participants, mimeo (2017)

Farquharson , R.: Theory of Voting. Yale University Press (1969)

Gibbard, A.: Manipulation of Voting Schemes: A General Result, Econometrica 41, 587-601 (1973)

Li, S.: Obviously strategy-proof mechanisms, mimeo (2016)

Mc Lean, I., Urken, A. Classics of Social Choice. The University of Michigan Press (1995)

Moulin, H.: On Strategy-proofness and Single Peakedness. Public Choice 35, 437-455 (1980)

Nisan, N.: Introduction to mechanism design (for computer scientists). In N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani, editors, Algorithmic Game Theory, Cambridge University Press (2007)

Pattanaik, P. K.: Counter-threats and Strategic Manipulation under Voting Schemes. The Review of Economic Studies 43 (1): 11-18 (1976)

Pattanaik, P. K.: Threats, Counter-Threats, and Strategic Voting. Econometrica 44 (1): 91-103 (1976)

Peleg, B., Almost all equilibiria in dominant strategies are coalition-proof. Econom. Lett. 60, 157–162 (1998)

Peleg, B., Sudhölter, P. Single-peakedness and coalition-proofness. Rev. Econ. Design 4: 381–387 (1999)

Satterthwaite, M.: Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions. J. Econ. Theory 10, 187-217 (1975)

Vickrey, W.: Utility, Strategy and Social Decision Rules, Quarterly Journal of Economics 74, 507-35 (1960)