Immunity to credible deviations from the truth*

by

Salvador Barberà[†], Dolors Berga[‡], and Bernardo Moreno[§]

April 2016

Abstract: We introduce a new notion of non-manipulability by groups, based on the idea that some of the agreements among a set of potential manipulators may be credible, while others may not. The derived notion of immunity to credible manipulations by groups is intermediate between individual and group strategy-proofness. We show that our new concept has bite, by applying it to the analysis of a large family of public good decision problems in separable environments, where there exist many attractive strategy-proof rules that are, however, manipulable by groups. In these environments we show that some of these rules are indeed, immune to credible group manipulations, while others are not. We provide characterization results that separate these two classes.

Journal of Economic Literature Classification Numbers: D71.

Keywords: Strategy-proofness, Group strategy-proofness, Credibility, Implementation, Voting.

*S. Barberà acknowledges financial support through grants ECO2014-53051-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 ECO2013-45395-R and 2014-SGR-1360. B. Moreno acknowledges financial support from grants SEJ4941, SEJ-5980, and ECO2014-53767. D. Berga and B. Moreno thank to the MOMA network under project ECO2014-57673-REDT.

[†]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, 17071 Girona, Spain. E-mail: dolors.berga@udg.edu

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

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, Berga, and Moreno 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 a new notion of immunity to credible group manipulations and characterize some subclasses of social choice rules that satisfy this property within specific environments. Specifically, we say that a group 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. And we say then that a rule is then immune to credible group manipulations if no set of agents can find a profitable deviation away from the truth that is credible.[2] 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.

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, Sonnenschein, and Zhou (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 in 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 references, also called generalized single-peaked. We build on Moulin (1980), Border and Jordan (1983) and Barberà, Gul, and Stachetti (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 consequences are also examined and proofs included

[1] We use this expression loosely here. Formal definitions of the environments we refer to are given in the next section.

[2] One could define credibility in other terms, some of which are equivalent, others not. This discussion is postponed to Section 4.

in the Appendix. Section 5 concludes with some final remarks.

2 The model and definitions: immunity and credibility

Let $N = \{1, ..., n\}$ 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 denote agent i's *preferences* on A. As usual, P_i and I_i denotes the strict and indifference preference relation induced by R_i, respectively. A *preference profile* $R_N = (R_1, ..., R_n) \in \mathcal{U} \times ... \times \mathcal{U} = \mathcal{U}^n$ 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 $f : \mathcal{U}^n \to A$.

Let us define some incentive related properties of social choice functions. The best known nonmanipulability 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 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.[4] But, of course, in many other

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

[4] See Le Breton and Zaporovhets (2009), Barberà, Berga, and Moreno (2010), and Barberà, Berga, and Moreno (2016).

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 he 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 $f(R'_C, R_{N|C}) R_i f(\overline{R}_i, R'_{C\setminus\{i\}}, R_{N\setminus C})$ for all $i \in C$ and all $\overline{R}_i \in \mathcal{U}$.*

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

Immunity to credible profitable deviations means that no profitable deviation of any coalition is credible at any profile. Observe that group strategy-proofness implies immunity to credible profitable deviations as defined above. However, in general the converse implication fails (see Proposition 1 below). Moreover, as Lemma 1 shows, immunity to credible profitable deviations implies strategy-proofness. And strategy-proofness implies immunity to credible profitable deviations by singletons.

Lemma 1 *Any social choice function f on $\mathcal{U}^n$ that is immune to credible profitable deviations is strategy-proof.*

Proof. By contradiction, let $R_N \in \mathcal{U}^n$, $i \in N$, and $R'_i \in \mathcal{U}$ such that $R'_i \neq R_i$ and $f(R'_i, R_{N\setminus\{i\}}) P_i f(R_N)$ and R'_i be such that it is a best deviation for agent i (which, as already stated, we assume to exist). By immunity to credible profitable deviations, there exists $\overline{R}_i \in \mathcal{U}$, such that $f(\overline{R}_i, R_{N\setminus\{i\}}) P_i f(R'_i, R_{N\setminus\{i\}})$ which contradicts that R'_i is a best deviation for i. ■

3 Applications

In this section 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 that are individually but not group strategy-proof and showing that some of them satisfy our new requirement but others do not. Specifically, we do that in a context where alternatives are multidimensional and preferences are separable.

Here is our framework. Let $\mathcal{K}=\{1,...,K\}$ be a finite set of $K \geq 2$ coordinates and for each $k \in \mathcal{K}$, let $B_k = [a_k, b_k]$ with $a_k < b_k$ be an integer interval.[5] Our *alternatives* are K-dimensional vectors in $B = \prod_{k=1}^{K} B_k$. To stress the role of a set of coordinates k_S, we will write $x = \left(x_{k_S}, x_{\mathcal{K}\setminus k_S}\right) \in B$. We endow B with the L_1-norm. That is, for any $x \in B$,

$$\|x\| = \sum_{k=1}^{K} \mid x_k \mid .$$

Given $x, y \in B$, the *minimal box containing* x *and* y is defined by

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

We restrict attention to the case where individual preferences are antisymmetric and thus, have a unique best alternative that we denote by $\tau(R_i)$.

We now impose a restriction on preferences which is a natural extension of single-peakedness to the multidimensional setting.

Definition 6 *A preference $R_i \in \mathcal{U}$ is (multidimensional)* ***single-peaked*** *if for any $z, y \in B$, if $y \in MB(z, \tau(R_i))$ then yR_iz.*

Let $\mathcal{S} \subset \mathcal{U}$ be the set of (multidimensional) *single-peaked* preferences on B. Under this preference restriction $\tau(R_i) = (\tau_1(R_i), ..., \tau_K(R_i)) \in B$ where $\tau_j(R_i)$ is the best (or top) alternative of R_i in dimension j.

It is known in the literature (Barberà, Gul, and Stacchetti 1993) that the class of multidimensional Generalized Median Voter Schemes (GMVS) are the only strategy-proof social choice functions in our setting, where multidimensional GMVS can be written as K unidimensional GMVS, one for each dimension. In this paper we restrict attention to a particular subclass of GMVS that is an K-dimensional extension of what Moulin (1980) called generalized Condorcet winner rules.

For each $k \in \mathcal{K}$, let $P_k = \{p_k^1, ..., p_k^{n-1}\}$ be an ordered list of $n-1$ values in B_k where $p_k^1 \leq ... \leq p_k^{n-1}$. If all values are the same we call the list degenerate. In what follows, we shall use K lists of such values, one for each dimension, as definitional parameters.

Definition 7 *We say that $f : \mathcal{S}^n \to B$, $f = (f_1, ..., f_K)$ is a generalized Condorcet winner rule if for any profile $R_N \in \mathcal{S}^n$, for any $k \in \mathcal{K}$, $f_k(R_N) = med\{\tau_k(R_1), ..., \tau_K(R_n), p_k^1, ..., p_k^{n-1}\}$, where $P_k(f) = \{p_k^1, ..., p_k^{n-1}\}$ is a list of parameters in B_k.*[6]

We remark that the rules we just defined are the only anonymous, onto, strategy-proof rules in this context.

[5] Note that for $K = 1$ any strategy-proof rule is also group strategy-proof. Hence, it is also immune to credible profitable deviations.

[6] The notation *med* denotes the median(s) of an ordered list. In the present definition this will be unique.

3.1 Choosing sets of candidates

The domain of separable preferences is important on the literature of strategy-proofness, since it has been proven to admit rich and attractive classes of non-manipulable social choice functions. Before engaging in a full analysis of those rules that are immune in a general framework, we consider the simple case, proposed in Barberà, Sonnenschein, and Zhou (1991), where in each dimension B_k can only take two values, say $a_k = 0$ and $b_k = 1$. Then, alternatives are vectors of zeros and ones, that we can interpret as the characteristic function of set of objects, where $x_k = 0$ means that the k-th object is not in the set, and $x_k = 1$ means that it does belong. We follow Barberà, Sonnenschein, and Zhou (1991) in interpreting this case as one where there exists a set $\mathcal{O}$ of K potential candidates or objects out of which we must choose the new members of a club or what items will form a shopping list and we adapt our notation accordingly. The following definition of separability which applies to this context, is equivalent to the one we define in the general case.[7]

Individual preferences are linear orders on the set $2^{\mathcal{O}}$ (including the empty set). Given any preference R on $2^{\mathcal{O}}$, we define the set of "good" objects $G(\mathcal{O}, R) = \{o_k \in \mathcal{O} : \{o_k\}P\varnothing\}$ and the set of "bad" objects $\mathcal{O}\backslash G(\mathcal{O}, R) = \{o_k \in \mathcal{O} : \varnothing P\{o_k\}\}$.

Definition 8 *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 $o_l \in G(\mathcal{O}, R)$.*

In words, adding a new good object to any set makes the union better than the original set and adding a bad object makes it worse. Now $\mathcal{S}$ denote the set of all separable preferences.

In this setting there exist strategy-proof social choice functions. In particular, the set of such functions that are anonymous and neutral coincides with the family of voting by quota rules, $f : \mathcal{S}^n \to 2^{\mathcal{O}}$ defined as follows:

Definition 9 *Let $q \in \{1, ..., n\}$. The social choice function f on $\mathcal{S}^n$ defined so that for any $R_N \in \mathcal{S}^n$,*

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

is called voting by quota q.

However, none of these voting by quota rules are group strategy-proof. And yet, we'll show that some of them are immune to credible profitable deviations, while others are not.

Before providing a characterization theorem allowing to distinguish between those rules that are immune and those that are not, we present two examples with 5 voters and 2 candidates. Notice that the following table is the set of all separable preferences when $K = 2$.

R^1	R^2	R^3	R^4	R^5	R^6	R^7	R^8
$\varnothing$	$\varnothing$	o_1	o_1	o_2	o_2	$\{o_1, o_2\}$	$\{o_1, o_2\}$
o_1	o_2	$\varnothing$	$\{o_1, o_2\}$	$\varnothing$	$\{o_1, o_2\}$	o_1	o_2
o_2	o_1	$\{o_1, o_2\}$	$\varnothing$	$\{o_1, o_2\}$	$\varnothing$	o_2	o_1
$\{o_1, o_2\}$	$\{o_1, o_2\}$	o_2	o_2	o_1	o_1	$\varnothing$	$\varnothing$

[7] See also Border and Jordan (1983), Le Breton and Sen (1999), Le Breton and Weymark (1999) who have analyzed a model with separable preferences in continuous multidimensional spaces.

Example 1 *Voting by quota 1: each agent declares her best set of objects and any object that is declared as good by some agent is selected.*
Consider the profile where $R_1 = R^3$*,* $R_2 = R^5$ *and for any other agent* $R_i = R^1$ *the outcome would be* $\{o_1, o_2\}$*, whereas* 1 *and* 2 *could vote for* $\varnothing$ *and get a preferred outcome.*
This proves that the rule is group manipulable. Notice, however, that after having agreed on voting for empty, any of the two agents could simply keep voting for their preferred candidate, and obtain an even better result, provided the other sticks to her announcement. Hence, this group manipulation will not be credible. We leave it to the reader to check that any other group manipulation under this rule will fail to be credible. Hence, in this example, voting by quota 1 is immune to credible group manipulation. As we shall see the result generalizes.

Example 2 *Voting by quota 3: each agent declares her best set of objects and any object that is declared as good by at least three agent is selected.*
Consider now the profile where $R_1 = R_2 = R^3$*,* $R_3 = R_4 = R^5$ *and* $R_5 = R^7$*. Then, the outcome would be* $\{o_1, o_2\}$*. Now, if agents* 1 *and* 2 *agree to vote for* $\varnothing$*, and so do agents 3 and 4, the coalition of these four agents can manipulate and have the outcome to be* $\varnothing$*, that they all prefer to* $\{o_1, o_2\}$*. Hence, the rule is group manipulable. Moreover, this particular manipulation is credible, because as long as the rest of deviators complies with the agreement, no single agent can profitable deviate from it. Hence, the rule is not immune to credible deviations in this case.*
Notice, however, that there would be other profitable deviations that would not be credible. For example, the one where only 1 and 3 agreed to drop their support to their preferred alternative.

In fact, we can prove the following general result.

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

Proof. To prove that voting by quota 1 is never subject to credible profitable deviations, notice that any profitable deviation by a group must involve agents who do not vote for some of the candidates they like (since they can always get them without anyone's help). In exchange these agents can get others not to vote for candidates that they dislike.
Let $R_N \in S^n$, C be a coalition that has a profitable deviation R'_C against R_N. Note that $f(R_N) \nsubseteqq f(R'_C, R_{N\backslash C})$ (otherwise, if $f(R_N) \subsetneqq f(R'_C, R_{N\backslash C})$, by quota 1, for any candidate $o \in f(R'_C, R_{N\backslash C}) \backslash f(R_N)$, $o \notin G(\mathcal{O}, R_i)$ for any $i \in N$. By separability, for any $i \in N$, $f(R_N) P_i f(R'_C, R_{N\backslash C})$ and R'_C could not be a profitable deviation, which is a contradiction). Thus, there exists a candidate o such that $o \in f(R_N) \backslash f(R'_C, R_{N\backslash C})$. Observe that for each such candidate $o \in f(R_N) \backslash f(R'_C, R_{N\backslash C})$, since f is voting by quota 1, there is at least one individual $i \in C$ such that $o \in G(\mathcal{O}, R_i)$ and $o \notin G(\mathcal{O}, R'_i)$. But now if i declares a preference $\overline{R}_i$ such that $G(\mathcal{O}, \overline{R}_i) = G(\mathcal{O}, R'_i) \cup \{o\}$, the outcome $f(\overline{R}_i, R'_{C\backslash\{i\}}, R_{N\backslash C}) = f(R'_C, R_{N\backslash C}) \cup \{o\}$, which is, by separability, strictly better for i under R_i than what he would get by following the agreed upon strategy. Therefore, no profitable deviation is credible under quota 1. A similar argument applies for quota n.

This already proves the proposition for the case $n = 2$ since there only the two extreme quotas can be used. From now on we treat the case $n > 3$.
To prove that any voting by quota rule q, $q \neq \{1, n\}$ violates immunity to credible profitable deviations we construct profiles against which there is a credible profitable deviation by some coalition. We begin by the case $K = 2$ and then argue that this can be embedded in a general profile presenting the same deviations whenever $K > 2$.

Let n be odd. We distinguish three subcases.
(1) $q > \frac{n-1}{2} + 1$. Let R_N be as follows: the preferences of any agent i in a set of $\frac{n-1}{2}$ agents are such that $o_1 P_i \{o_1, o_2\} P_i \varnothing$, the preferences of any agent j in a different set of $\frac{n-1}{2}$ agents are such that $o_2 P_j \{o_1, o_2\} P_j \varnothing$, and the preferences of the remaining agent l is such that $\tau(R_l) = \{o_1, o_2\}$. Observe that $f(R_N) = \varnothing$. Let C be the coalition of all agents except agent l, let R'_C such that each agent $i \in C$, $\tau(R'_i) = \{o_1, o_2\}$. Observe that since $f(R'_C, R_{N \setminus C}) = \{o_1, o_2\}$, R'_C is a profitable deviation of C against R_N. Finally, R'_C is credible since no agent can change the outcome by a unilateral deviation since $n > 3$.
(2) $q = \frac{n-1}{2} + 1$. Let R_N be as follows: the preferences of any agent i in a set of $\frac{n-1}{2}$ agents are such that $o_1 P_i \{o_1, o_2\} P_i \varnothing$, the preferences of any agent j in a different set of $\frac{n-1}{2}$ agents are such that $o_2 P_j \{o_1, o_2\} P_j \varnothing$, and the preferences of the remaining agent l is such that $\tau(R_l) = \varnothing$. Observe that $f(R_N) = \varnothing$. Let C be the coalition of all agents except agent l, let R'_C such that each agent $i \in C$, $\tau(R'_i) = \{o_1, o_2\}$. Observe that since $f(R'_C, R_{N \setminus C}) = \{o_1, o_2\}$, R'_C is a profitable deviation of C against R_N. Finally, R'_C is credible since no agent can change the outcome by a unilateral deviation since $n > 3$.
(3) $q < \frac{n-1}{2} + 1$. Let R_N be as follows: the preferences of any agent i in a set of $\frac{n-1}{2}$ agents are such that $o_1 P_i \varnothing P_i \{o_1, o_2\}$, the preferences of any agent j in a different set of $\frac{n-1}{2}$ agents are such that $o_2 P_j \varnothing P_j \{o_1, o_2\}$, and the preferences of the remaining agent l is such that $\tau(R_l) = \varnothing$. Observe that $f(R_N) = \{o_1, o_2\}$. Let C be the coalition of all agents except agent l, let R'_C such that each agent $i \in C$, $\tau(R'_i) = \varnothing$. Observe that since $f(R'_C, R_{N \setminus C}) = \varnothing$, R'_C is a profitable deviation of C against R_N. Finally, R'_C is credible since no agent can change the outcome by a unilateral deviation since $n > 3$.

Let n be even. We distinguish two subcases.
(1) $q > \frac{n}{2}$. Let R_N be as follows: the preferences of any agent i in a set of $\frac{n}{2}$ agents are such that $o_1 P_i \{o_1, o_2\} P_i \varnothing$, the preferences of any agent j in a different set of $\frac{n}{2}$ agents are such that $o_2 P_j \{o_1, o_2\} P_j \varnothing$. Observe that $f(R_N) = \varnothing$. Let C be the coalition of all agents, let R'_C such that each agent $i \in C$, $\tau(R'_i) = \{o_1, o_2\}$. Observe that since $f(R'_C, R_{N \setminus C}) = \{o_1, o_2\}$, R'_C is a profitable deviation of C against R_N. Finally, R'_C is credible since no agent can change the outcome by a unilateral deviation.
(2) $q \leq \frac{n}{2}$. Let R_N be as follows: the preferences of any agent i in a set of $\frac{n}{2}$ agents are such that $o_1 P_i \varnothing P_i \{o_1, o_2\}$, the preferences of any agent j in a different set of $\frac{n}{2}$ agents are such that $o_2 P_j \varnothing P_j \{o_1, o_2\}$. Observe that $f(R_N) = \{o_1, o_2\}$. Let C be the coalition of all agents, let R'_C such that each agent $i \in C$, $\tau(R'_i) = \varnothing$. Observe that since $f(R'_C, R_{N \setminus C}) = \varnothing$, R'_C is a profitable deviation of C against R_N. Finally, R'_C is credible since no agent can change the outcome by a unilateral deviation.

This is easily extended to the case $K > 2$ by considering profiles where agents preferences are like the ones described in each case above for objects 1 and 2, while all the agents share exactly the same preferences concerning other objects for all cases analyzed (for example, $o_k \in G(\mathcal{O}, \widehat{R}_i)$ for each $o_k \in \mathcal{O} \backslash \{o_1, o_2\}$, each $i \in N$ and each individual preference $\widehat{R}_i$ used in the analyzed cases). ■

Our next proposition covers the case $n = 3$, which is not contemplated by the previous one.

Proposition 2 *When $n = 3$ and $K = 2$, any voting by quota rule is immune to credible profitable deviations. When $n = 3$ and $K \geq 3$, voting by quota 1 and 3 are the only rules in our class satisfying immunity to credible profitable deviations.*

Proof. Let $N = \{1, 2, 3\}$ and $K = 2$. For voting by quota 1 and 3 the same argument in Proposition 1 applies. Consider voting by quota 2. As already remarked in Barberà, Sonnenschein, and Zhou (1991) this rule is not only strategy-proof but also efficient. Thus the only coalitions with profitable deviations consist of two agents. Let R_N, $C = \{i, j\}$, and R'_C be a profitable deviation of C against R_N.
To be a profitable deviation, observe that, by separability and voting by quota 2, either (1) both candidates are chosen under $(R'_C, R_{N|C})$ but none under R_N, or (2) no candidate is chosen under $(R'_C, R_{N|C})$ but both are chosen under R_N, or (3) only one candidate is chosen under R_N and only the other candidate is chosen under $(R'_C, R_{N|C})$.
In the first case, for each candidate, one of the agents in C considered that candidate not good under R_i but good under R'_i. In the second case, for each candidate, one of the agents in C considered that candidate good under R_i but not good under R'_i. In the third case, what said in the second case holds for the candidate chosen under R_N and what said in the first case holds for the candidate chosen under $(R'_C, R_{N|C})$.
In the three cases, either declaring $\overline{R}_i$ such that a good candidate under R'_i not to be under $\overline{R}_i$, or supporting a bad one will be an individual profitable deviation with respect to $(R'_C, R_{N|C})$. Thus, R'_C is not credible.

Let $N = \{1, 2, 3\}$ and $K = 3$. For voting by quota 1 and 3 the same argument in Proposition 1 applies. To prove that voting by quota 2 violates immunity to credible profitable deviations we provide an example of a credible profitable deviation against a profile. Let R_N be as follows: the preferences of agent 1 are such that $\tau(R_1) = o_1$ and $\{o_1, o_2, o_3\} P_1 \varnothing$, the preferences of agent 2 are such that $\tau(R_2) = o_2$ and $\{o_1, o_2, o_3\} P_2 \varnothing$, and the preferences of agent 3 are such that $\tau(R_3) = o_3$ and $\{o_1, o_2, o_3\} P_3 \varnothing$. Observe that $f(R_N) = \varnothing$. Let $C = N$, and R'_N such that each agent $i \in C$, $\tau(R'_i) = \{o_1, o_2, o_3\}$. Since $f(R'_N) = \{o_1, o_2, o_3\}$, R'_N is a profitable deviation of C against R_N. Finally, R'_N is credible since no agent can change the outcome by a unilateral deviation.
This is easily extended to the case $K > 3$ by considering profiles where agents preferences are like the ones described in each case above for objects 1, 2, and 3, while all the agents share exactly the same preferences concerning other objects for all cases analyzed (for example, $o_k \in G(\mathcal{O}, \widehat{R}_i)$ for each $o_k \in \mathcal{O} \backslash \{o_1, o_2, o_3\}$, each $i \in N$ and each individual preference $\widehat{R}_i$ used in the analyzed cases). ■

3.2 The general case: choosing from a grid

We now analyze the case where the set of options in each dimension is not binary.

Proposition 3 *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 profitable deviations.*

Proof. Let f be a generalized Condorcet winner rule with two dimensions, say 1 and 2, for which $P_1(f)$ and $P_2(f)$ are not degenerate. Consider the median(s), $medP_1(f)$ and $medP_2(f)$ of these parameters' lists. These medians may be unique or consist of two contiguous points, say $med^-P_k(f) < med^+P_k(f)$, for each $k \in \{1,2\}$.
In all cases below, in any profile we will define the preferences of each agent in N concerning dimensions different from 1 and 2 to be the same and with top at some point x_k in B_k, $k \in \mathcal{K}\backslash\{1,2\}$.

Assume first that for each $k \in \{1,2\}$, $med^-P_k(f) \neq med^+P_k(f)$. This can only happen if n is odd and thus the number of parameters is even. Consider a partition of N into three sets, A, A', and l where l is a singleton and such that $\#A = \#A'$. Let the projections of R_N in dimensions 1 and 2 be as follows. For agents in A, let the $k-$dimensional top be $med^+P_k(f)$ for $k \in \{1,2\}$. For agents in A', let the $k-$dimensional top be $med^-P_k(f)$ for $k \in \{1,2\}$. Agent l has the $1-$dimensional top at $med^+P_1(f)$ and the $2-$ dimensional top at $med^-P_2(f)$. Also assume for agents in $A \cup A'$ that $(med^-P_1(f), med^+P_2(f), x_{\mathcal{K}\backslash\{1,2\}})P_i(med^+P_1(f), med^-P_2(f), x_{\mathcal{K}\backslash\{1,2\}})$. Observe that $f_k(R_N) = \tau_k(R_l)$ for each $k \in \{1,2\}$ and $f_k(R_N) = x_k$ for each $k \in \mathcal{K}\backslash\{1,2\}$. This is because, for each $k \in \{1,2\}$, $\tau_k(R_l)$ tie-breaks when computing f_k as the median of all tops and parameters in B_k. Let $C = A \cup A'$ and let R'_C be such that for each agent $i \in C$, $\tau_1(R'_i) = med^-P_1(f)$, $\tau_2(R'_i) = med^+P_2(f)$.[8] Observe that $f_k(R'_C, R_{N\backslash C}) = \tau_k(R'_i)$ for $k \in \{1,2\}$, and $f_k(R'_C, R_{N\backslash C}) = x_k$ for each $k \in \mathcal{K}\backslash\{1,2\}$. This is because, for each k, $f_k(R'_C, R_{N\backslash C})$ is the top for individual preferences in $(R'_C, R_{N\backslash C})$ for $n-1$ agents and it coincides with $med^-P_1(f)$ in dimension 1 and with $med^+P_2(f)$ in dimension 2. By definition, this shows that R'_C is a profitable deviation of C against R_N.
Moreover, for each dimension $k \in \{1,2\}$, since $f_k(R'_C, R_{N\backslash C})$ is either $med^-P_k(f)$ or $med^+P_k(f)$ and, given that $n > 3$, there are at least two parameters smaller or equal than $f_k(R'_C, R_{N\backslash C}) = med^-P_k(f)$ or greater or equal than $f_k(R'_C, R_{N\backslash C}) = med^+P_k(f)$.
Therefore, $f_k(R'_C, R_{N\backslash C})$ receives at least $n+1$ total votes for each $k \in \{1,2\}$. Hence, the profitable deviation R'_C is credible and f is not immune.

Assume now that for at least some $k \in \{1,2\}$, $med^-P_k(f) = med^+P_k(f) = medP_k(f)$. Remember that $med^-P_k(f) \neq med^+P_k(f)$ can only hold if n is odd and therefore the number of parameters is even. Because of that in the case where the above equality holds for only one of the two dimensions but not for the other can only happen when n is odd. This distinction is used along the rest of the proof because in one case a partition will only use two sets of agents A, A' while in other cases we will have to add a singlet l to it.

[8] In words, to define R'_C notice that by changing their vote the agents in A vote for the tops of those in A' in dimension 1, while agents in A' vote for the top of those in A in dimension 2.

For n odd, let A, A', and agent l be the elements of a partition of N such that $\#A = \#A' = \frac{n-1}{2}$. For n even, let, A, A' a partition of N such that $\#A = \#A' = \frac{n}{2}$. Let R_N be as follows. The preferences of agents in A are such that in the dimension 1 the top is either $med^+P_1(f)$ when $med^-P_1(f) \neq med^+P_1(f)$, or $medP_1(f)$, otherwise. In dimension 2 the top is either $med^+P_2(f)$ when $med^-P_2(f) \neq med^+P_2(f)$, or the highest parameter strictly smaller than $medP_2(f)$ if it exists and $med^-P_2(f) = med^+P_2(f)$, or the lowest parameter strictly greater than $medP_2(f)$, otherwise. The preferences of agents in A' are such that in dimension 1 the top is either $med^-P_1(f)$ when $med^-P_1(f) \neq med^+P_1(f)$, or the highest parameter strictly smaller than $medP_1(f)$ if it exists and $med^-P_1(f) = med^+P_1(f)$, or the lowest parameter strictly greater than $medP_1(f)$, otherwise. In dimension 2 the top is either $med^-P_2(f)$ when $med^-P_2(f) \neq med^+P_2(f)$, or $medP_2(f)$, otherwise.

Preferences of agent l (only required if n is odd) are defined as follows: R_l is such that in dimension 1 agent l's top is either $\tau_1(R_j)$, $j \in A$ when $med^-P_1(f) \neq med^+P_1(f)$, or the highest parameter strictly smaller than $medP_1(f)$ if such parameter exists or the lowest parameter strictly greater than $medP_1(f)$, otherwise. In dimension 2 the top of agent l is either either $\tau_2(R_i)$, $i \in A'$ when $med^-P_2(f) \neq med^+P_2(f)$, or the highest parameter strictly smaller than $medP_2(f)$ if such parameter exists or the lowest parameter strictly greater than $medP_2(f)$, otherwise.

From now on let $i \in A'$ and $j \in A$. We also assume that for any agent $m \in A \cup A'$, $(\tau_1(R_i), \tau_2(R_j), x_{\mathcal{K}\backslash\{1,2\}})\ P_m(\tau_1(R_j), \tau_2(R_i), x_{\mathcal{K}\backslash\{1,2\}})$. Observe that $f_k(R_N) = \tau_k(R_l)$ if $med^-P_k(f) \neq med^+P_k(f)$, and $f_k(R_N) = medP_k(f)$ otherwise for each $k \in \{1,2\}$, and that $f_k(R_N) = x_k$ for each $k \in \mathcal{K}\backslash\{1,2\}$. This is because, for each $k \in \{1,2\}$, $\tau_k(R_l)$ tie-breaks when computing f_k as the median of all tops and parameters in B_k in the case where only for one $k \in \{1,2\}$, $med^-P_k(f) = med^+P_k(f) = medP_k(f)$ and thus n is odd. And for each $k \in \{1,2\}$, $medP_k(f)$ tie-breaks when computing f_k as the median of all tops and parameters in B_k in the case where for both $k \in \{1,2\}$, $med^-P_k(f) = med^+P_k(f) = medP_k(f)$. Let $C = A \cup A'$ and let R'_C be such that for each agent $j \in A$, $\tau_1(R'_j) = \tau_1(R_i)$ and $\tau_k(R'_j) = \tau_k(R_j)$ for each $k \in \mathcal{K}\backslash\{1\}$, and for each $i \in A'$, $\tau_2(R'_i) = \tau_2(R_j)$ and $\tau_k(R'_i) = \tau_k(R_i)$ for each $k \in \mathcal{K}\backslash\{2\}$. Observe that $f_k(R'_C, R_{N\backslash C}) = \tau_k(R'_i)$ for $k \in \{1,2\}$, and $f_k(R'_C, R_{N\backslash C}) = x_k$ for each $k \in \mathcal{K}\backslash\{1,2\}$. This is because, for each k where $med^-P_k(f) = med^+P_k(f)$, $f_k(R'_C, R_{N\backslash C})$ is the top for individual preferences in $(R'_C, R_{N\backslash C})$ for n agents. For each k where $med^-P_k(f) \neq med^+P_k(f)$, $f_k(R'_C, R_{N\backslash C})$ is the top for the preferences for $n-1$ agents in $(R'_C, R_{N\backslash C})$ and coincides either with $med^-P_k(f)$ or $med^+P_k(f)$. By definition, this shows that R'_C is a profitable deviation of C against R_N.

Moreover, for the dimensions where $med^-P_k(f) = med^+P_k(f)$ there is a parameter at $f_k(R'_C, R_{N\backslash C})$. For the dimensions where $med^-P_k(f) \neq med^+P_k(f)$, $f_k(R'_C, R_{N\backslash C})$ is either $med^-P_k(f)$ or $med^+P_k(f)$ and, given that $n > 3$, there are at least two parameters smaller or equal than $f_k(R'_C, R_{N\backslash C}) = med^-P_k(f)$ or greater or equal than $f_k(R'_C, R_{N\backslash C}) = med^+P_k(f)$. Therefore, $f_k(R'_C, R_{N\backslash C})$ receives at least $n+1$ total votes for each $k \in \{1,2\}$. Hence, the profitable deviation R'_C is credible and f is not immune. ■

The following claims will be crucial for the proofs of immunity in the following propositions. In all the claims we consider a generalized Condorcet winner rule f, and we assume that R'_C is a profitable deviation of C against R_N. We will use the fact that by the separability condition of individual preferences R, for any x_k, y_k and fixed $x_{k'}, y_{k'} \in B_{k'}$ for $k' \in K \backslash \{k\}$, $(x_k, x_{K\backslash\{k\}})P(y_k, x_{K\backslash\{k\}})$ if and only if $(x_k, y_{K\backslash\{k\}})P(y_k, y_{K\backslash\{k\}})$.

Claim 1 *For any $k \in K$ and R'_C defined as above, $\tau_{k'}(R_i) \neq \tau_{k'}(R'_i)$ for some $k' \in K\backslash\{k\}$, and some $i \in C$.*

Claim 1 holds because unidimensional generalized Condorcet winners are group strategy-proof and by separability only the top in dimension k matters. In what follows we will refer to this claim by saying that f is group strategy-proof in each dimension k. Therefore, if $f_k(R'_C, R_{N\backslash C}) \neq f_k(R_N)$ there is at least one agent $i \in C$ such that $\left(f_k(R_N), x_{K\backslash\{k\}}\right) P_i \left(f_k(R'_C, R_{N\backslash C}), x_{K\backslash\{k\}}\right)$. Formally:

Remark If $f_k(R'_C, R_{N\backslash C}) \neq f_k(R_N)$ and $f_{k'}(R'_C, R_{N\backslash C}) = f_{k'}(R_N)$ for all $k' \in K\backslash\{k\}$ then there exists $i \in C$ such that $f(R_N)P_i f(R'_C, R_{N\backslash C})$.

In words, we say that there is an agent $i \in C$ who is losing according to R_i at $(R'_C, R_{N\backslash C})$ in dimension k. Since the original preferences R_i remains fixed in all proofs, we will informally say that an agent is losing at some profile in some dimension. Similarly, we could define the concept of an agent winning.

Claim 2 *There exist $k, k' \in K$ and $i, j \in C$ such that $\left(f_k(R_N), x_{K\backslash\{k\}}\right) P_i \left(f_k(R'_C, R_{N\backslash C}), x_{K\backslash\{k\}}\right)$, $\left(f_k(R'_C, R_{N\backslash C}), x_{K\backslash\{k\}}\right) P_j \left(f_k(R_N), x_{K\backslash\{k\}}\right)$, for any $x_{K\backslash\{k\}} \in B_{K\backslash\{k\}}$, $\left(f_{k'}(R'_C, R_{N\backslash C}), x_{K\backslash\{k'\}}\right) P_i \left(f_{k'}(R_N), x_{K\backslash\{k'\}}\right)$, and $\left(f_{k'}(R_N), x_{K\backslash\{k'\}}\right) P_j \left(f_{k'}(R'_C, R_{N\backslash C}), x_{K\backslash\{k'\}}\right)$, for any $x_{K\backslash\{k'\}} \in B_{K\backslash\{k'\}}$.*

To show it, consider for each dimension k the partition of agents in C between the ones winning (W_k) and the ones losing (L_k) at $(R'_C, R_{N\backslash C})$ in dimension k. Notice that, by definition of profitable deviation, each agent in C must be winning at $(R'_C, R_{N\backslash C})$ in some dimension k, and as remarked after Claim 1 there must be another agent in C losing at $(R'_C, R_{N\backslash C})$ in that dimension k. For each agent in L_k there exists a dimension k' where he wins. Take one agent in L_k, if some of the agents in W_k belongs to $L_{k'}$, the result holds. Otherwise, $W_{k'} \supsetneq W_k$. Take another agents in L_k and repeat the same argument. Since there is a finite number of agents we will obtain the result.

Claim 3 *Let k be such that $P_k(f)$ is degenerate. Let $i \in C$ and $j \in C$ be winning and losing at $(R'_C, R_{N\backslash C})$ in dimension k, respectively. Then, the profitable deviation R'_C is not credible.*

The sketch of the proof is as follows. Consider two cases. In the first case, in R_N all agents' k- dimensional top are placed in the same side of the parameters' unique position ($P_k(f)$ is degenerate). $f_k(R_N)$ is the top closest to the single parameter. For R'_C to be a profitable deviation, all agents with k-dimensional top in $f_k(R_N)$ must belong to C and for these agents, $\tau_k(R'_i)$ must be such that is even further from the single parameter. Note that all these agents would be losing at $(R'_C, R_{N\backslash C})$ in that dimension k. Then, any of these agents, say i, announcing $\overline{R}_i$ such that $\tau_k(\overline{R}_i) = \tau_k(R_i)$ and for each $k' \in \mathcal{K}\backslash\{k\}$, $\tau_{k'}(\overline{R}_i) = \tau_{k'}(R'_i)$ would be winning at $(\overline{R}_i, R'_{C\backslash\{i\}}, R_{N\backslash C})$ in dimension k and by separability he would be better off $f(\overline{R}_i, R'_{C\backslash\{i\}}, R_{N\backslash C})P_i f(R'_C, R_{N\backslash C})$ which means that R'_C is not a credible profitable deviation. In the second case, in both sides of the single parameter there is at least one agents' top given R_N. Thus, $f_k(R_N)$ is the single parameter. For R'_C to be a profitable deviation, all agents in one side of the single parameter belongs to C and they should announce $\tau_k(R'_i)$ in the other side of the single parameter. Note that all these agents would be losing at $(R'_C, R_{N\setminus C})$ in that dimension k. Then, any of these agents, say i, announcing $\overline{R}_i$ such that $\tau_k(\overline{R}_i) = \tau_k(R_i)$ and for each $k' \in \mathcal{K}\setminus\{k\}$, $\tau_{k'}(\overline{R}_i) = \tau_{k'}(R'_i)$ would be winning at $(\overline{R}_i, R'_{C\setminus\{i\}}, R_{N\setminus C})$ in dimension k and by separability he would be better off $f(\overline{R}_i, R'_{C\setminus\{i\}}, R_{N\setminus C}) P_i f(R'_C, R_{N\setminus C})$ which means that R'_C is not a credible profitable deviation.

Proposition 4 *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 profitable deviations.*

Proof. Consider f as in the statement. Let $R_N \in \mathcal{S}^n$, $C \subseteq N$, and $R'_C \in \mathcal{S}^c$ be a profitable deviation of C against R_N. By Claim 2, there exist $k, k' \in K$ and $i, j \in C$ such that

$$\left(f_k(R_N), f_{K\setminus\{k\}}(R'_C, R_{N\setminus C})\right) P_i \left(f_k(R'_C, R_{N\setminus C}), f_{K\setminus\{k\}}(R'_C, R_{N\setminus C})\right),$$

$$\left(f_k(R'_C, R_{N\setminus C}), f_{K\setminus\{k\}}(R'_C, R_{N\setminus C})\right) P_j \left(f_k(R_N), f_{K\setminus\{k\}}(R'_C, R_{N\setminus C})\right),$$

$$\left(f_{k'}(R'_C, R_{N\setminus C}), f_{K\setminus\{k'\}}(R'_C, R_{N\setminus C})\right) P_i \left(f_{k'}(R_N), f_{K\setminus\{k'\}}(R'_C, R_{N\setminus C})\right), \text{ and}$$

$$\left(f_{k'}(R_N), f_{K\setminus\{k'\}}(R'_C, R_{N\setminus C})\right) P_j \left(f_{k'}(R'_C, R_{N\setminus C}), f_{K\setminus\{k'\}}(R'_C, R_{N\setminus C})\right).$$

By hypothesis, either $P_k(f)$ or $P_{k'}(f)$ is degenerate (or both).[9] By Claim 3, R'_C is not a credible profitable deviation. ■

Proposition 5 *Let $n = 3$. Any generalized Condorcet winner rule defined by lists of parameters such that are non degenerate in two dimensions is immune to credible profitable deviations. Any generalized Condorcet winner rule defined by non degenerate lists of parameters in at least three dimensions is not immune to credible profitable deviations.*

Proof. To prove the first statement, let f be a generalized Condorcet winner rule with lists of parameters, denoted in each dimension k as $p_k^- \leq p_k^+$, and such that they are non degenerate in exactly two dimensions. To prove that f is immune to credible profitable deviations, let $R_N \in \mathcal{S}^3$, $C \subseteq N = \{1,2,3\}$, and $R'_C \in \mathcal{S}^c$ be a profitable deviation of C against R_N. If the profitable deviation is such that there is an agent that is winning in a dimension k for which $P_k(f)$ is degenerate, by Claim 3, R'_C could not be a credible profitable deviation. Thus, agents must be winning at $(R'_C, R_{N\setminus C})$ in dimensions k for which $P_k(f)$ is not degenerate. By Claim 2, there must be agents in C winning at $(R'_C, R_{N\setminus C})$ in the two dimensions with non degenerate list of parameters $P_k(f)$. Let's call them dimensions 1 and 2. By strategy-proofness, C has at least two agents. Without loss of generality, by anonymity and Claim 2, suppose that agents 1 and 2 belong to C and that agent 1 is winning while agent 2 is losing at $(R'_C, R_{N\setminus C})$ in dimension 1 and the opposite holds in dimension 2. Consider the following two cases depending on the size of the deviating coalition.

Case 1: $C = \{1, 2\}$.

[9]For $n = 2$ any list of parameters is degenerate in all dimensions since all parameters take the same value.

Since agent 3's preferences are fixed, we can assume now that we have 3 fixed parameters in each dimension, two of them different: p_k^-, p_k^+, and $\tau_k(R_3)$. Consider dimension 1. First, observe that

$$f_1(R_N), f_1(R'_{\{1,2\}}, R_3) \in \left[\min\left\{\tau_1(R_3), p_1^-\right\}, \max\left\{\tau_1(R_3), p_1^+\right\}\right].$$

Since agent 1 is winning and agent 2 is losing at $(R'_{\{1,2\}}, R_3)$ in dimension 1, then $\tau_1(R_1)$ must be strictly placed on one side of $f_1(R_N)$, while $\tau_1(R_2)$ must be weakly placed on the other side of $f_1(R_N)$. Moreover, for both $i \in \{1,2\}$, $\tau_1(R'_i)$ must be strictly in the same side of $f_1(R_N)$ and in fact in the same side as $\tau_1(R_1)$ is. Then, note that agent 2 announcing $\overline{R}_2$ such that $\tau_1(\overline{R}_2) = \tau_1(R_2)$ and for each $k \in \mathcal{K}\backslash\{1\}$, $\tau_k(\overline{R}_2) = \tau_k(R'_2)$ would be winning at $(\overline{R}_2, R'_1, R_3)$ in dimension 1 and by separability he would be better off $f(\overline{R}_2, R'_1, R_3) P_2 f(R'_{\{1,2\}}, R_3)$ which means that $R'_{\{1,2\}}$ is not a credible profitable deviation.
<u>Case 2</u>: $C = \{1,2,3\}$.
Since $R'_N \in \mathcal{S}^n$ is a profitable deviation of C against R_N, agent 3 is winning at R'_N in some dimension. Without loss of generality, suppose that agent 3 is winning at R'_N in dimension 1 (otherwise, by anonymity a similar argument would apply). We already assumed that agent 1 is winning and agent 2 is losing at R'_N in dimension 1. If $f_1(R_N) \in \left(p_1^-, p_1^+\right)$, then $\tau_1(R_2) = f_1(R_N)$ since agent 2 is losing at R'_N, and also both $\tau_1(R_1)$ and $\tau_1(R_3)$ must be strictly on a different side of $f_1(R_N)$. But then, R'_C would not be a profitable deviation where agents 1 and 3 are winning at R'_N in dimension 1. Thus, $f_1(R_N) \leq p_1^-$ or $f_1(R_N) \geq p_1^+$. Suppose $f_1(R_N) \leq p_1^-$ (a symmetric argument would apply for the other case). Observe first that for each $i \in \{1,2,3\}$, $\tau_1(R_i) \leq f_1(R_N)$. Since agent 2 is the only agent losing at R'_N in dimension 1, $\tau_1(R_2) = f_1(R_N)$ and $\tau_1(R_1), \tau_1(R_3) < f_1(R_N)$, and thus $f_1(R'_C, R_{N\backslash C}) < f_1(R_N)$. Then, note that agent 2 announcing $\overline{R}_2$ such that $\tau_1(\overline{R}_2) = \tau_1(R_2)$ and for each $k \in \mathcal{K}\backslash\{1\}$, $\tau_k(\overline{R}_2) = \tau_k(R'_2)$ would be winning at $(\overline{R}_2, R'_{\{1,3\}})$ in dimension 1 and by separability he would be better off $f(\overline{R}_2, R'_{\{1,3\}}) P_2 f(R'_{\{1,2,3\}})$ which means that $R'_{\{1,2,3\}}$ is not a credible profitable deviation.

To prove the second statement, let f be a generalized Condorcet winner rule with lists of parameters, denoted in each dimension k as $p_k^- \leq p_k^+$, and such that they are non degenerate in at least three dimensions. Assume they are dimensions 1, 2, and 3. To prove that f is not immune to credible profitable deviations, we provide an example of a credible profitable deviation against a profile. In any profile we will define the preferences of each agent in N concerning dimensions different from 1, 2, and 3 to be the same and with top at some point x_k in B_k, $k \in \mathcal{K}\backslash\{1,2,3\}$.
Let $R_N \in \mathcal{S}^3$ be as follows in dimensions 1, 2, and 3: define the preferences of agent 1 such that $\tau(R_1) = \left(p_1^+, p_2^-, p_3^-\right)$ and $\left(p_1^+, p_2^+, p_3^+, x_{\mathcal{K}\backslash\{1,2,3\}}\right) P_1 \left(p_1^-, p_2^-, p_3^-, x_{\mathcal{K}\backslash\{1,2,3\}}\right)$, the preferences of agent 2 such that $\tau(R_2) = \left(p_1^-, p_2^+, p_3^-\right)$ and $\left(p_1^+, p_2^+, p_3^+, x_{\mathcal{K}\backslash\{1,2,3\}}\right) P_2 \left(p_1^-, p_2^-, p_3^-, x_{\mathcal{K}\backslash\{1,2,3\}}\right)$, and the preferences of agent 3 such that $\tau(R_3) = \left(p_1^-, p_2^-, p_3^+\right)$ and $\left(p_1^+, p_2^+, p_3^+, x_{\mathcal{K}\backslash\{1,2,3\}}\right) P_3$ $\left(p_1^-, p_2^-, p_3^-, x_{\mathcal{K}\backslash\{1,2,3\}}\right)$. Observe that $f(R_N) = \left(p_1^-, p_2^-, p_3^-, x_{\mathcal{K}\backslash\{1,2,3\}}\right)$. Let $C = N$, and R'_N such that each agent $i \in C$, $\tau(R'_i) = \left(p_1^+, p_2^+, p_3^+\right)$. Since $f(R'_N) = \left(p_1^+, p_2^+, p_3^+\right)$, R'_N is a profitable deviation of C against R_N. Finally, R'_N is credible since no agent can change the outcome by a unilateral deviation. ■

4 Some alternative formulations of credibility, and their consequences

We believe that our definition of a credible deviation is quite attractive. But others could be conceivable, and in this section we shall discuss other possible proposals, and relate them to ours.

To favor the comparison, let us go back to the interpretation of credibility that we already proposed after Definition 4. 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 $f(\cdot, R_{N\setminus C})$. Starting from this, we shall discuss, then, three possible variants of the credibility concept.

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, we'll show that the set of rules that are immune to credible deviations will be the same (after a minimal qualification) under either definition. This is expressed in Proposition 6 and it explains why we do not give new names to the concepts that derive from that approach.

A second variant will require that in order to be (extensively) credible, the deviation R'_C should be a Nash equilibrium for the game where all agents (whether or not they are part of C) can play any preference, and f is the outcome function. If the initial function f is assumed to be strategy-proof (an assumption that we do not need under our original definition), then again the set of immune rules will still be the same under either definition (see Proposition 7). However, the equivalence is not true if our f function is not a priori restricted to be strategy-proof, as shown in Example 3.

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. We'll show that under this definition, all of the rules we consider will be immune to credible profitable deviations (see Proposition 8).[10]

We now present formal arguments to make the preceding discussion more precise. We also state some results and their proofs are included in the Appendix.

Definition 10 *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 (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 $\mathcal{U}^n$ is* ***immune to*** *(type 1)* ***credible profitable deviations*** *if for any $R_N \in \mathcal{U}^n$, any $C \subseteq N$, there is no (type 1) credible profitable deviation of C against R_N.*

Proposition 6 *Any social choice function f on $\mathcal{U}^n$ is immune to credible profitable deviations if and only if f is immune to (type 1) credible profitable deviations.*

[10] The same will hold if instead allowing agents to use any preferences, they are only assumed to use their true and the manipulative one.

Definition 11 *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 (type 2)* ***credible*** *if $f(R'_C, R_{N|C}) R_i f(\overline{R}_i, R'_{C\setminus\{i\}}, R_{N\setminus C})$ for all $i \in N$ and all $\overline{R}_i \in \mathcal{U}$. A social choice function f on $\mathcal{U}^n$ is* ***immune to*** *(type 2)* ***credible profitable deviations*** *if for any $R_N \in \mathcal{U}^n$, any $C \subseteq N$, there is no (type 2) credible profitable deviation of C against R_N.*

Proposition 7 *Any strategy-proof social choice function f on $\mathcal{U}^n$ is immune to credible profitable deviations if and only if f is immune to (type 2) credible profitable deviations.*

The following example shows that the latter immunity concept does not imply strategy-proofness. Therefore, the concept may be useful to apply in contexts where strategy-proofness is not to be expected, but one may still be interested in discussing the diversity of manipulative actions by groups of voters.

Example 3 *Immunity to (type 2) credible deviations does not imply strategy-proofness under appropriately restricted domains.*
The following table is the set of all admissible separable preferences when $K = 2$, for agents 1 and 2:

R_1^1	R_1^2	R_1^3	R_1^4	R_2^1	R_2^2	R_2^3	R_2^4
$\varnothing$	o_1	o_1	$\{o_1, o_2\}$	$\varnothing$	o_2	o_2	$\{o_1, o_2\}$
o_1	$\varnothing$	$\{o_1, o_2\}$	o_1	o_2	$\varnothing$	$\{o_1, o_2\}$	o_2
o_2	$\{o_1, o_2\}$	$\varnothing$	o_2	o_1	$\{o_1, o_2\}$	$\varnothing$	o_1
$\{o_1, o_2\}$	o_2	o_2	$\varnothing$	$\{o_1, o_2\}$	o_1	o_1	$\varnothing$

Suppose the following social choice function:

f	R_2^1, R_2^2	R_2^3, R_2^4
R_1^3, R_1^4	o_1	o_2
R_1^1, R_1^2	o_2	o_1

Note that in the direct revelation game induced by this social choice function, no agent has a dominant strategy. Hence, the rule is not strategy-proof (thus, violating immunity to both credible and (type 1) credible profitable deviations). Also notice that the grand coalition has no profitable deviation. Hence, all profitable deviations involve a single agent, and for each one of them, the remaining agent can respond with a new profitable deviation. Hence, the social choice function is immune to (type 2) credible deviations, even if not strategy-proof.

Definition 12 *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* ***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 $\overline{R}_S \in \mathcal{R}^s$ and for some $i \in S$. A social choice function f on $\mathcal{U}^n$ is* ***immune to strongly credible profitable deviations*** *if for any $R_N \in \mathcal{U}^n$, any $C \subseteq N$, there is no strongly credible profitable deviation of C against R_N.*

Proposition 8 *All generalized Condorcet winner rules are immune to strongly credible profitable deviations.*

Let us say that we are aware that the idea of credibility may have other expressions. Credibility is invoked when defining subgame perfection in extensive form games, and is also a label for a specific form of implementation. We are also aware that, by combining dominant strategy and Nash equilibria, we may raise larger questions regarding the consistency of the players reasoning when forming coalitions. Addressing the consistency of using both types of equilibria within the same definition is a challenge for future work.

5 Final remarks

We have opened the way to study new concepts regarding the incentives of groups of agents to cooperate in manipulating social choice functions, and characterized some subclasses of rules that may satisfy the new requirements in separable environments.

The voting methods we have identified are interesting in several respects.

One interesting aspect is efficiency. It is clear that strategy-proof rules cannot be fully efficient unless they satisfy a strong notion of group strategy-proofness. Yet, those that satisfy our intermediate property have the interesting feature that any departure from their prescribed outcomes leading to an efficient one would not be credible. Thus, they are, in that sense, efficient up to credibility constraints.

Another interesting conclusion of our analysis is that those rules that imply extreme distributions of voting power are immune to credible deviations from truth-telling. One could think that this distribution is uneven or unfair. However, the class of Generalized Condorcet winner rules that are obtained when the definitional parameters are concentrated in a single point do coincide, in each dimension, with those characterized by Thomson (1993, 1999) as being the only methods that satisfying an attractive normative property. His property, that Thomson calls "welfare domination under preference replacement", requires that when one agent changes preferences and modifies the social outcome, all other agents' welfare must change in the same direction. Hence, we not only found exactly what are the conditions that allow immunity, but also discovered that they may be justified in terms of pre-existing normative concepts.

Finally, let us acknowledge that the treatment of strategic considerations by the different agents is somewhat asymmetric. Indeed, groups are allowed to form in order to manipulate, but our main concept of credibility only considers single-agent non-cooperative departures from cooperative agreements, à la Nash. This invites for further reflection regarding these and other issues of coalition formation, that we hope to keep developing in further work.

References

S. Barberà, D. Berga, and B. Moreno, Individual versus group strategy-proofness: when do they coincide?, Journal of Economic Theory 145 (2010): 1648-1674.
S. Barberà, D. Berga, and B. Moreno, Group strategy-proofness in private good economies, American Economic Review 106(4) (2016): 1073-99.
S. Barberà, F. Gul, E. Stacchetti, Generalized Median Voter Schemes and Committees. J Econ Theory 61 (1993): 262-289
S. Barberà, H. Sonnenschein, L. Zhou, Voting by committees, Econometrica 59 (1991): 595–609.
K. Border, J. Jordan, Straightforward elections, unanimity and phantom voters, Rev. Econ. Stud. 50 (1983): 153–70.
M. Le Breton, A. Sen, Separable preferences, decomposability and strategyproofness, Econometrica 67 (1999): 605–28.
M. Le Breton and J. Weymark, Strategy-proof social choice with continuous separable preferences, Journal of Mathematical Economics 32 (1999): 47–85.
M. Le Breton and V. Zaporozhets, On the Equivalence of Coalitional and Individual Strategy-Proofness Properties, Soc. Choice Welfare 33 (2009): 287-309.
H. Moulin, On Strategy-proofness and Single Peakedness, Pub Choice 35 (1980): 437-455.
W. Thomson, The replacement principle in public good economies with single-peaked preferences, Economics Letters 42 (1993): 31-36.
W. Thomson, Welfare-domination under preference-replacement: A survey and open questions, Soc Choice Welfare 16 (1999): 373-394.

Appendix

This Appendix contains the proofs of propositions stated in Section 4.

Proof of Proposition 6 By definition immunity to (type 1) credible profitable deviations implies immunity to credible profitable deviations. To prove the converse, let $R_N \in \mathcal{U}^n$, $C \subseteq N$, and $R'_C \in \mathcal{U}^c$ be a profitable deviation of C against R_N. Suppose that for all $i \in C$, $f(R'_C, R_{N|C}) R_i f(R_i, R'_{C\setminus\{i\}}, R_{N\setminus C})$. By Lemma 1 f is strategy-proof, thus $f(R'_C, R_{N|C}) I_i$ $f(R_i, R'_{C\setminus\{i\}}, R_{N\setminus C})$ for all $i \in C$. By immunity to credible profitable deviations, there exists $i \in C$ such that $f(\overline{R}_i, R'_{C\setminus\{i\}}, R_{N\setminus C}) P_i f(R'_C, R_{N|C})$ for some $\overline{R}_i \in \mathcal{U}$. By these two facts, for some $i \in C$, $f(\overline{R}_i, R'_{C\setminus\{i\}}, R_{N\setminus C}) P_i f(R_i, R'_{C\setminus\{i\}}, R_{N\setminus C})$ which contradicts strategy-proofness.

Proof of Proposition 7 By definition immunity to credible profitable deviations implies immunity to (type 2) credible profitable deviations. To prove the converse, let $R_N \in \mathcal{U}^n$, $C \subseteq N$, and $R'_C \in \mathcal{U}^c$ be a profitable deviation of C against R_N. Suppose that for all $i \in N$, $f(R'_C, R_{N|C}) R_i f(\overline{R}_i, R'_{C\setminus\{i\}}, R_{N\setminus C})$ for all $\overline{R}_i \in \mathcal{U}$. Thus, for all $i \in N$, $f(R'_C, R_{N|C}) R_i f(R_i, R'_{C\setminus\{i\}}, R_{N\setminus C})$. Since f is strategy-proof, $f(R'_C, R_{N|C}) I_i f(R_i, R'_{C\setminus\{i\}}, R_{N\setminus C})$ for all $i \in N$. By immunity to credible profitable deviations, there exists $i \in C$ such that $f(\overline{R}_i, R'_{C\setminus\{i\}}, R_{N\setminus C}) P_i f(R'_C, R_{N|C})$ for some $\overline{R}_i \in \mathcal{U}$. By these two facts, for some $i \in N$, $f(\overline{R}_i, R'_{C\setminus\{i\}}, R_{N\setminus C}) P_i f(R_i, R'_{C\setminus\{i\}}, R_{N\setminus C})$ which contradicts strategy-proofness.

Proof of Proposition 8 Let f be a generalized Condorcet winner rule with lists of parameters, denoted in each dimension k as $p_k^- \leq p_k^+$. To prove that f is immune to strongly credible profitable deviations, let $R_N \in \mathcal{S}^n$, $C \subseteq N$, and $R'_C \in \mathcal{S}^c$ be a profitable deviation of C against R_N. Since R'_C is a profitable deviation, there must exist at least one dimension k in which $f_k(R_N) \neq f_k(R'_C, R_{N\backslash C})$ and some agent $i \in C$ is winning. By Claim 1, there is an agent $j \in C\backslash\{i\}$ who is losing in that dimension k. Let A and A' be a partition of C such that $A = \{i \in C$: is winning in dimension $k\}$ and $A' = \{j \in C$: is losing in dimension $k\}$. Suppose, wlog, that $f_k(R'_C, R_{N\backslash C}) < f_k(R_N)$. By definition of A, for any $i \in A$, $\tau_k(R_i) < f_k(R_N)$. By definition of A', $\tau_k(R_j) > f_k(R'_C, R_{N\backslash C})$. We distinguish two cases:

Case 1. For any $j \in A'$, $\tau_k(R_j) \leq f_k(R_N)$. Since $f_k(R'_C, R_{N\backslash C}) < f_k(R_N)$, then for any $l \in C$, $\tau_k(R'_l) \leq f_k(R_N)$. Also, it must happen that for some $j \in A'$, $\tau_k(R_j) = f_k(R_N)$. Let $S = \{j \in A' : \tau(R_j) = f_k(R_N)\}$. Then for any $l \in S$, let $\overline{R}_l$ be such that $\tau_k(\overline{R}_l) = \tau_k(R_l)$ and $\tau_{k'}(\overline{R}_l) = \tau_{k'}(R'_l)$ for any $k' \in \mathcal{K}\backslash\{k\}$. Thus $f_k(\overline{R}_S, R'_{C\backslash S}, R_{N\backslash C}) = f_k(R_N)$ and by separability $f(\overline{R}_S, R'_{C\backslash S}, R_{N\backslash C})$ is preferred to $f_k(R'_C, R_{N\backslash C})$ for any $l \in S$. Which means that R'_C is not strongly credible.

Case 2. For some $j \in A'$, $\tau_k(R_j) > f_k(R_N)$. Let $S = \{j \in A' : \tau(R_j) \geq f_k(R_N)$ and $\tau_k(R'_l) \neq \tau_k(R_l)\}$. Since $f_k(R'_C, R_{N\backslash C}) < f_k(R_N)$, S is not empty. Then for any $l \in S$, let $\overline{R}_l$ be such that $\tau_k(\overline{R}_l) = \tau_k(R_l)$ and $\tau_k(\overline{R}_l) = \tau_k(R'_l)$. Then for any $l \in S$, let $\overline{R}_l$ be such that $\tau_k(\overline{R}_l) = \tau_k(R_l)$ and $\tau_{k'}(\overline{R}_l) = \tau_{k'}(R'_l)$ for any $k' \in \mathcal{K}\backslash\{k\}$. Thus $f_k(\overline{R}_S, R'_{C\backslash S}, R_{N\backslash C}) = f_k(R_N)$ and by separability $f(\overline{R}_S, R'_{C\backslash S}, R_{N\backslash C})$ is preferred to $f_k(R'_C, R_{N\backslash C})$ for any $l \in S$. Which means that R'_C is not strongly credible.