Soc Choice Welfare (2001) 18: 619–653

Social Choice and Welfare

© Springer-Verlag 2001

An introduction to strategy-proof social choice functions

Salvador Barberà

CODE and Departament d'Economia i d'Història Econòmica, Universitat Autònoma de Barcelona, Belleiterra, 08193 Barcelona, Spain (e-mail: salvador.barbera@uab.es)

Received: 21 March 2001/Accepted: 9 June 2001

1 Introduction

Establishing which collective choices respond best to the values of concerned individuals is a central problem in economics, ethics and politics. In order to take these values into account, they must be known, either directly or indirectly. Some decision-making processes recognize this explicitly and require citizens to vote, as in political elections, or to express their preferences otherwise, as when applying for admission to public universities or internships in hospitals. In other cases, the reference to interested individuals is more indirect. It is not always clear whether and how decision makers integrate the preferences of their constituencies into their calculations, but they cannot be completely insensitive to them.

Knowledge of individual preferences is a primary example of asymmetrically distributed private information. In an extreme version, we may claim that each individual knows her preferences perfectly, and that nobody else has any information about them. Less extremely, it is easy to accept that we all know about our own preferences better than anybody else. There is another sense in which information about preferences is private: agents are entitled to hold any opinion, and to sustain it as their own, even if others know it is not. Therefore, individual preferences must be the guide to good collective decisions, but knowledge about them is typically dispersed among individuals, and

This text is based on invited lectures delivered at the Nato Advanced Study Institute, Center for Game Theory in Economics, State University of New York at Stony Brook (July 1997). I am grateful to William Thomson for the opportunity to participate, and to Matthew Jackson and Antonio Nicoló for their comments on different versions of this text.

not obviously available to the decision-makers or to the institutions which determine social outcomes.

Can we guarantee that agents will provide the decision-maker with accurate information regarding their preferences, or that they will behave within a given institution in a way that reveals their actual preferences? Not necessarily! Even in ancient times people who voted were aware of the possibility that not revealing one's true preferences might be a superior strategy than just voting straightforwardly. Farquharson (1969) based one of the early contemporary monographs on strategic voting on a letter by Plinius the Young, who discusses a debate in the Roman Senate where three groups of voters were equally split in favor of three alternative courses of action. Plinius himself, fearing that his worse preferred outcome could come about, was considering to join the supporters of the decision that was intermediate for him, in order to at least guarantee himself that much. Writers of the enlightenment period in France, like Borda and Condorcet, who anticipated the importance of voting in democratic societies, defended specific methods by going well beyond their mere description: they actually defended their specific methods over others by providing sound discussions of their properties and their respective advantages. These founders of the modern theory of voting were also well aware of the possibility that voters might not only be led by their true preferences, but also by other calculations, when filling their ballots. Getting closer to ourselves, let us use some introspection: anyone who has sometimes rooted for a loser must have debated whether to actually vote for the hopeless candidate, just to make the point, or else select the least undesirable candidate among the potential winners, and cast a "useful" vote.

So, we learn from the books and from our own experience that it may be hard to know about the actual preferences of agents, even if we are informed about their actions, unless we can get a hold of the actual connections between what people want and what people do. How can we get such a hold? To begin with, notice that people with the same objectives may act differently depending on the rules under which society operates. I may find that my best action is to vote sincerely under majority rule, but prefer to be devious under point voting. Hence, if we want to control for the connections between desires and actions, we must be specific about the institutions through which society is making decisions. In fact, knowing that many well known rules are manipulable does not exempt us from asking a first and very natural question: could there be some voting rule, or some other institution for collective decision making, under which all agents would always find it best to act straightfordwardly and reveal their true preferences? We are not interested in naive, myopic agents, but on sophisticated ones; could it be that, after careful scrutiny of all opportunities, all agents would always conclude that their best action is to be truthful? If and when they exist, rules under which this would happen will be called strategy-proof, or non-manipulable. The aim of this article is to explain under what circumstances it may be possible to design strategy-proof rules, and how they would be like.

Before we jump to this main objective, let me elaborate a bit more on the

connections between individual preferences and individual actions. Strategy-proof rules, if they exist, will elicit the true preferences of agents. If, on top of that, the rule's choices are efficient from the point of view of declared preferences, then they will be truly efficient, since people will have declared their true preferences. And even if the chosen actions are suboptimal, we can still rely on meaningful preferences to make any welfare statements which might apply. There are other, more sophisticated procedures for arriving at efficient decisions, while not necessarily learning about the true preferences of agents. When asked to act within an institution, agents are allowed a number of strategies. Social outcomes are determined once each agent picks a strategy, through some outcome function. Under voting rules, strategies are ways to fill the ballots. Under market rules, they may include the possible decisions to supply or demand, or the possibility to fix prices, or some other actions. Sometimes, strategies are simple enough that they can be identified with the simple expression of preferences. For example, a ballot may be seen as a way to describe my preferences over candidates. Similarly, my willingness to pay for some good reveals an important part of my preferences. Then, for worlds where my set of preferences are essentially the same as my set of actions, a strategy-proof rule is one where, it is always a dominant strategy for me to use the strategy that reveals my actual preferences. To be dominant, a strategy must not only be the best response to some set of actions by other agents: it must be a best response to any set of actions that others may take. This is such a strong requirement that dominant strategies might well not exist, in many cases.

Then, strategy-proofness cannot be achieved, but there is still some hope to link the actions of rational agents with the set of socially desirable outcomes, by designing mechanisms that implement these outcomes. Combinations of individual strategies can still be in equilibrium even if agents do not have dominant strategies. There are different notions of equilibrium, and what we'll say next applies to any of them. But, for the sake of argument, consider Nash-equilibria, that is, combinations of strategies having the internal property that each one of them is a best response to the others. Fix the preferences of agents, and consider the Nash equilibria associated with the game that a given institution gives rise to, once the agent's preferences are set. Are the outcomes of the institution satisfactory, given the agent's preferences? If yes, check the same for another game, resulting from the same rules (the same institutions) but under another set of preferences. Are the Nash-equilibrium outcomes again satisfactory, given the new preferences? If so, proceed again. Should the answer always be positive for a family of preference profiles, we then can say that the institution at hand implements in Nash equilibria the performance criterion under which we have evaluated, in each case, whether the outcomes were satisfactory. Here is a more indirect and sophisticated way to connect the actions of individuals with their consequences: just design mechanisms under which the actions of agents (truthful or not, assuming that this word makes sense here) lead to desirable outcomes at equilibrium.

The purpose of implementation theory is to study when it is possible to implement some performance criteria by means of appropriate sets of strategies and outcome functions, and eventually to describe the types of institutions that would emerge from such a designing exercise. This issue also contains a primer to implementation theory and mechanism design due to Matthew Jackson (2001). Implementation theory anticipates the sophistication of agents and creates institutions where essentially all strategic efforts to beat the rules are self defeating. The quest for strategy-proof rules is part of this large intellectual building, but it is probably its most quiet room: we look for the case where people will find it rational to convince themselves that engaging in these self-defeating exercises is just not worth their effort.

This paper is meant as an introduction to the literature on strategy proofness. It is not a survey: the exposition concentrates on a handful of models. Suggestions for further reading are given in the concluding remarks, and the references include just a sample of the vast literature on the topic. My selection tries to be suggestive of the types of questions, the kinds of results to expect and the techniques of proof that appear in this literature. The paper is structured as follows. In Sect. 2, I state an important theorem, establishing that manipulations are essentially unavoidable unless the preferences of agents over alternatives are restricted. I also provide a guide to some of the alternative proofs for this important result. Section 3, then, reviews a family of models where it is natural to assume that the preferences of agents will be restricted, and this allows for the existence of nontrivial strategy-proof social choice functions. The basic restriction I consider is single-peakedness on a line, which naturally arises in many political and economic contexts. I also discuss extensions of this concept to more complicated sets of alternatives. Sections 2 and 3 concentrate on cases where all agents are concerned by the complete description of the alternatives as a whole, and they are all allowed to have similar preferences. For example, an alternative may be a political candidate, and then anyone can consider the candidate to be first, or last, or middle. By contrast, Sects. 4 and 5 consider models where agents are essentially concerned only on those aspects of the global choices that concern them directly. For example, in an exchange economy, the social alternatives are the matrices that represent the allocation of all resources among all agents, but each one of them can selfishly care only about the row in this matrix that refers to his own consumption. Hence, my preferences on the global alternatives will be based on my private component and will come from a different set of preferences on the global alternatives than those of other agents. In this type of environments it still makes sense to consider restricted domains of preferences, although the relevant restrictions will come from other motivations: they may include standard restrictions on economic preferences, like convexity or monotonicity. Section 4 considers rationing problems, and Sect. 5 considers exchange economies. In both cases I describe specific methods to allocate private resources to agents in a strategy-proof manner, for appropriate domain restrictions. Section 6 contains some final comments and refers to further readings.

2 Strategy-proofness for unrestricted domains: the Gibbard Satterthwaite Theorem

2.1 Preliminaries

A will be a set of *alternatives*, (finite or infinite). $I = \{1, 2, \ldots, n\}$ will be a finite set of *agents*. Agents in I will be assumed to have *preferences* on A. Preferences will be always complete, reflexive, transitive binary relations on A. R will stand for the set of all possible preferences on A. *Preference profiles* are n-tuples of preferences, one for each agent in $I = \{1, 2, \ldots, n\}$.

A *social choice function* on the domain $D_1 \times \cdots \times D_n \subset R^n$ is a function $f : D_1 \times \cdots \times D_n \to A$, where each D_i is considered to represent the set of preferences which are admissible for agent i.

What preferences are admissible, or interesting, or relevant, will change with the interpretation of A, the set of alternatives. Different economic situations will give rise to alternative setups, some of which will be considered along this paper.

We shall focus on social choice functions which are strategy-proof, or non-manipulable. *A social choice function $f : D_1 \times \cdots \times D_n \to A$ is manipulable* iff there exists some preference profile $(\succcurlyeq_1, \ldots, \succcurlyeq_n) \in D_1 \times \cdots \times D_n$, and some preference $\succcurlyeq' \in D_i$, such that

$$f(\succcurlyeq_1, \ldots, \succcurlyeq'_i, \ldots, \succcurlyeq_n) \succ_i f(\succcurlyeq_1, \ldots, \succcurlyeq_i, \ldots, \succcurlyeq_n)$$

The function f *is strategy-proof* iff it is not manipulable.

Given a social choice function f, denote by r_f the *range* of f. Given a complete preference relation $\succcurlyeq$ on the set A of alternatives, and a subset B of A, let $C(\succcurlyeq, B) = \{b \in B \mid \text{for all } c \in B, b \succcurlyeq c\}$. The set $C(\succcurlyeq, B)$ denotes the $\succcurlyeq$-maximal elements in B, and is interpreted as the set of alternatives that an agent endowed with preferences $\succcurlyeq$ would consider best out of those in B.

A social choice function f is *dictatorial* iff there exists a fixed agent i such that, for all preference profiles,

$$f(\succcurlyeq_1, \ldots, \succcurlyeq_n) \in C(\succcurlyeq_i, r_f)$$

Hence, a dictatorial social choice function is trivial, in that it does not really aggregate preferences of agents, but simply chooses one of the best elements of one and the same agent (when it is unique, this fully describes the rule; otherwise complementary criteria to break ties are allowed, but this hardly allows to consider the rule anything but trivial).

The following theorem establishes that all non trivial social choice functions on the universal domain of preferences are manipulable. We informally bunch up, under the term "trivial", two types of rules: those that are dictatorial, and those which only choose between two alternatives. Indeed, for the simple case where society must decide between only two alternatives, the majority rule, or any reasonable variant of it, are strategy-proof. But these rules break down dramatically when more than two choices are at stake, as expressed by the following

Theorem 1. *(Gibbard (1973), Satterthwaite (1975)) Any social choice function $f : \mathcal{R}^n \to A$, whose range contains more than two alternatives, is either dictatorial or manipulable.*

Notice that choosing by majority over two alternatives (with an appropriate tie-breaking rule) is a nondictatorial and non-manipulable social choice function. Because of this and other similar examples, Theorem 1 must be explicit about the requirement that there are at least three alternatives in the range. Another essential assumption of this theorem is that the social choice function is defined on the universal set of preferences over A.

2.2 Proofs of the Gibbard-Satterthwaite Theorem

Because the theorem is important, it has been the object of much attention, and many alternative proofs of it have been offered. We shall briefly outline several of them. To unify the discussion, we concentrate on the case where the set of alternatives is finite.

The earliest proof is due to Gibbard (1973), and it relies heavily on Arrow's impossibility theorem (1951). The latter refers to social welfare functions: that is, to rules which assign a transitive preference relation to each preference profile. It states that a social welfare function over the universal domain satisfying the properties of Pareto (P) and Independence of Irrelevant Alternatives (IIA) must be dictatorial (when there are at least three alternatives).

Gibbard's proof (and some variants of it, like a very elegant one due to Schmeidler and Sonnenschein 1978) run as follows. Start from a strategy-proof social choice function f with at least three alternatives in its range; construct (in a way to be described) an auxiliary rule, based on f, that assigns to each profile a binary relation on the alternatives in the range of f; prove that, under the given construction, this binary relation is transitive (if f is strategy-proof), and that the auxiliary rule w_f is thus a social welfare function: show that, again due to $f's$ strategy-proofness, w_f must also satisfy the conditions of Pareto and IIA; conclude (from Arrow's theorem) that w_f is dictatorial and (from the construction) that f must also be.

Different ways to define w_f from f can be used to make the above argument. Gibbard's is as follows: for any profile $(\succcurlyeq_1, \ldots, \succcurlyeq_n)$, and any two alternatives x and y in the range, construct a new profile $(\succcurlyeq_1^{xy}, \ldots, \succcurlyeq_n^{xy})$, where each agent i places x and y on the top of his ranking, while keeping the relative order of x and y as in $\succcurlyeq_i$, and also respecting the relative orders of any pair not involving x and y; calculate the outcome $f(\succcurlyeq_1^{xy}, \ldots, \succcurlyeq_n^{xy})$; if f is strategy-proof, we must get either x or y (this takes an easy proof); then, declare x socially preferred to y under profile $(\succcurlyeq_1, \ldots, \succcurlyeq_n)$ if x is the outcome of f for $(\succcurlyeq_1^{xy}, \ldots, \succcurlyeq_n^{xy})$, or y preferred to x if y comes out.

The above construction was initially proposed as a tool to prove the Gibbard-Satterthwaite Theorem. But, once set in place, the construction allows to establish a close connection between the set of social choice functions satisfying strategy-proofness and that of social welfare functions meeting

Arrow's conditions. This connection is very interesting, as emphasized by Satterthwaite (1975), and it holds not only under the universal domain assumption, but also for some restricted domains (like those where agents' preferences are single-peaked – see Barberà et al. 1993). However there are domains admitting nondictatorial strategy proof social choice functions that do not admit nondictatorial Arrowian social welfare functions. And there are domains where the converse holds. Hence, the close connection between the possibility of solving Arrow's aggregation problem, and that of finding a strategy-proof rule cannot be taken as a universal fact: it must be checked for each restricted domain.

A second interesting proof of the Gibbard-Satterthwaite theorem is based on a close examination of strategy-proof social choice rules for the two-person, three-alternative case, followed by a double induction on the number of agents and alternatives (Schmeidler and Sonnenschein 1978). Concentrating first on the 6×6 matrix corresponding to the combinations of strict preferences for the two agents, a number of short but subtle arguments lead to the conclusion that strategy-proofness only allows for social outcomes which always coincide with the preferred alternative of one of the two agents. Then, a simple reasoning extends the conclusion to general preferences (admitting indifferences), and induction does the rest. This proof emphasizes that the two person, three alternative case contains all the essential elements of the theorem, in a nutshell.

Finally, I would like to sketch a proof that was presented in Barberà and Peleg (1990), and has its roots in Barberà (1983). The preceding proofs only apply when the number of alternatives is finite. The proof I am about to present, although it is still proposed here for the finite case, can be easily adapted to cover the case with a continuum of alternatives. It is also a good starting point for the analysis of strategy-proof rules operating under restricted domains. Because of that, many of the results to be surveyed later are proven with techniques similar to those I will now present.

To be concise, I'll consider two-agent social choice functions, and assume that agents have strict preferences. We denote the set of all strict preferences by $\mathscr{P}$ (here again, the extensions to general preferences and to n agents are quite straightforward). The argument runs as follows.

- Given f, let $f : \mathscr{P} \times \mathscr{P} \to A$ be strategy-proof
- Given f, define the notion of an option set. This will be key to our proof. The options left for 2, given a preference P_1 for agent 1, are defined by

$$o_2(P_1) = \{x \mid \exists P_2, f(P_1 P_2) = x\}$$

Notice that this definition is relative to f. We should write $o_{2f}(P_1)$, but we omit the f for simplicity. These are the outcomes that 2 could obtain, by some declaration of preferences (truthful or not), should 1 declare preferences P_1.

The proof now proceeds along five elementary remarks. The first remark is that, *if f is strategy-proof, then for all preference profiles $f(P_1, P_2) = C(P_2, o_2(P_1))$*. This is just a rewording of the strategy-proofness condition, but

it allows us to think of functions satisfying this property as generated by a two stage process: agent one, by declaring her preferences P_1, narrows down 2's options to $o_2(P_1)$; then, agent 2 chooses her best alternative out of the options left by 1. (Clearly, the argument is symmetric; the roles of 1 and 2 could be reversed all along). Notice that, if agent 1 was a dictator, then $o_2(P_1)$ would be a singleton and coincide with $1's$ preferred alternative. On the other hand if 2 is a dictator $o_2(P_1) = r_f$ for any P_1, since 1's declaration is irrelevant to the function's outcome, and fixing it does in no way restrict the possible choice of 2.

Given this first remark, the proof of the Gibbard-Satterthwaite Theorem consists in showing that a strategy-proof social choice function must generate option sets $o_2(P_1)$ which always select a singleton (1's best alternative) or always leave all of r_f for 2 to choose from. This is easily proven through a sequence of additional remarks, which shed light on the structure of strategy-proof functions, and whose proofs are really simple. (The reader can try to prove them directly. If in need, turn to Barberà and Peleg, Sect. 2).

The second remark is that, for any P_1, $o_2(P_1)$ must contain the best element of P_1 in r_f. That is, agent 1 should always leave room for 2 to choose, eventually, 1's favorite outcome.

The third remark establishes that whenever $C(P_1, r_f) = C(P_1', r_f)$, then $o_2(P_1) = o_2(P_1')$. That is, only the "top" alternative for agent 1 in r_f can be relevant in determining the options that 1 leaves for 2.

The fourth remark is that, whenever the range of f contains at least three alternatives, then $o_2(P_1)$ must either be, for each P_1, equal to r_f or to $C(P_1, r_f)$.

The fifth and last remark concludes the proof by showing that, in fact, only one of the two possibilities above can hold. Either $o_2(P_1)$ is always equal to r_f, or it is always equal to $C(P_1, r_f)$. Hence, f must be dictatorial if it is strategy-proof, has at least three alternatives in its range (this plays a role in proving the fourth remark) and is defined on a universal domain (this is used to prove the last three remarks).

Like all important results, the Gibbard-Satterthwaite Theorem can be looked at from different angles. The proofs we have sketched correspond to different approaches, and each one of them has brought some new insights into the structure of strategy-proof social choice functions. The purpose of this section was to present these basic insights somewhat informally, and to encourage the reader to learn the G-S theorem in full detail, as a useful first step for the design of strategy-proof social choice rules.

3 Strategy-proofness and single-peakedness

3.1 The choice of linearly ordered alternatives, when individual preferences are single-peaked

The clear-cut conclusion of the Gibbard-Satterthwaite Theorem is obtained at some costs: one of them is the assumption of universal domain, according to which all possible preferences over alternatives are admissible for all agents.

In many cases, however, the nature of the social decision problem induces a specific structure on the set of alternatives and this structure suggests, in turn, some restrictions on the set of admissible individual preferences. It is then natural to investigate how much does the negative conclusion of the Gibbard-Satterthwaite Theorem change, when social choice functions are only required to operate on restricted domains of preferences. In the rest of this essay we'll study a number of special problems, each one giving rise to a specific structure for the set of alternatives and to some natural domain restrictions. In each of these instances, we'll investigate the possibility of defining nontrivial strategy-proof social choice functions, and try to characterize the sets of such functions when possible.

We'll first consider situations where alternatives can be linearly ordered, according to some criterion (from "left" to "right" in political applications, from smaller to greater according to some quantitative index, etc.) In this context, it makes sense to say that one alternative x is between two others, z and w, say. And it is sometimes natural to assume that the preference of agents over alternatives is single-peaked, meaning that (1) each agent has a single preferred alternative $\mathscr{T}(\succcurlyeq_i)$, and (2) if alternative z is between x and $\mathscr{T}(\succcurlyeq_i)$, then z is preferred to x (intuitively, this is because z can be considered closer than x to the ideal $\mathscr{T}(\succcurlyeq_1)$).

Single peaked preferences were first discussed by Duncan Black (1948) and they arise naturally in many contexts. Here is an example from location theory. Let the real line stand for a set of locations, let each agent be located at some point in the line, and let the alternatives be the locations where to place a facility. Take any location $\bar{l}$ and look at it from the point of view of an agent located at $\bar{a}$. Consider some l' which is between $\bar{l}$ and $\bar{a}$ (hence, closer to $\bar{a}$ than $\bar{l}$). If it is always the case that an agent in $\bar{a}$ would prefer the "closest" alternative l' to $\bar{l}$, which is farther away, then the preferences of this agent are single peaked. For another example, consider an agent who has preferences on the space $\Re^2_+$. Let points $(x, y) \in \Re^2_+$ stand for the amounts spent by government on services X and Y, respectively. The budget line $Z_1 + Z_2 = B$ for $Z_i \geq 0$ represents the different ways to distribute a budget B between there two types of expenditures. It is easy to check that any agent with monotonic and strictly convex preference on $\Re^2_+$ will rank the elements of the budget line single-peakedly.

To be specific, we'll concentrate on the case where the number of alternatives is finite, and identify them with the integers in an interval $[a, b] = \{a, a+1, a+2, \ldots, b\} \equiv A$. (All the results we describe also apply to the case where A is the real line, ordered by the $\succcurlyeq$ relation. In fact, that is the context of Moulin 1980, whose results we adapt here). We assume throughout that the preferences of all agents are single-peaked.

Under these assumptions, there exist non trivial strategy-proof social choice functions. Here are some examples:

Example 1. *There are three agents. Allow each one to vote for her preferred alternative. Choose the median of the three voters.*

To see that the rule is not manipulable, consider the options of one agent, say 1, when the other two have already voted for some alternatives c and d (without loss of generality, let $c \leq d$). Then, 1 can determine any outcome between c and d, and none other (if $c = d$, then this is the outcome regardless of 1's vote). If 1's top alternative is in the integers interval $[c, d]$, then 1 gets her best without manipulating. If her top alternative is below c, then c is the outcome and, by single-peakedness, this is better for 1 than any outcome in $[c, d]$. Similarly, if the top for 1 is above d, d is 1's best option. Notice that the same rule would not be strategy-proof for larger domains, allowing preferences not to be single-peaked.

Example 2. *There are two agents. We fix an alternative p in $[a, b]$. Agents are asked to vote for their best alternatives, and the median of p, $\mathscr{T}_1$ and $\mathscr{T}_2$ is the outcome.*

Again, the median is well defined, because it is taken from an odd number of values: two of them are the agent's votes, while the third one is a fixed value. We'll call this value a phantom.

Example 3. *For any number of agents, ask each one for their preferred alternative and choose the smallest.*

This is another strategy-proof rule. Notice that the options left to any agent are those smaller than or equal to the smallest vote of others. Hence, if this agent's ideal is still lower, she can choose it. Otherwise, the outcome of voting for her best (which is the lowest vote of others) cannot be improved either.

Remark that this rule, which might appear to be quite different from the preceding ones, can in fact also be written as a median. To do so, when there are n agents, place $n - 1$ phantoms and n alternatives on the lowest alternative a. Then the function can be described as choosing the median between these $n - 1$ phantoms and the n alternatives supported by actual voters.

Up to here, those rules are anonymous: interchanging the roles of agents (along their votes) does not change the outcome. The following and last example describes a strategy-proof rule where different agents play different roles.

Example 4. *There are two agents. Fix two alternatives w_1 and w_2, $(w_1 \leq w_2)$. If agent 1 votes for any alternative in $[w_1, w_2]$, the outcome is 1's vote. If 1 votes for an alternative larger than w_2, the outcome is the median of w_2 and the votes of both agents. If 1 votes below w_1, then the outcome is the median of w_1 and the votes of both agents.*

Notice that this rule can also be described in other ways.

One way is the following. Assign values on the extended real line to the sets $\{1\}, \{2\}, \{1, 2\}$. Specifically, let $a_1 = w_1$, $a_2 = w_2$, $a_{1,2} = a$ (the lowest value in the range). Now, define the rule as choosing

$$f(\succcurlyeq_1, \succcurlyeq_2) = \inf_{S \in \{\{1,2\},\{1\}\{2\}\}} \left[\sup_{i \in S} (a_S, \mathscr{T}(\succcurlyeq_i)) \right].$$

We shall state immediately that this formula generalizes. There are also other ways to write the same rule. These will be discussed in Sect. 3.2.

Moulin (1980) characterized the class of all strategy-proof social choice functions. Actually, he worked on the extended real line. He also assumed that the rules were only based on the preferred elements for each voter. This is an unnecessary assumption, because strategy proof social choice rules in these (and in many other) domains are restricted to only use information on what each agent considers best. This was proven in Barberà and Jackson (1994) in a context of public goods, and also in Sprumont (1991) in a context of allocation rules. As a result, we can express the structure of all strategy-proof social choice functions (defined on the full set of single-peaked preference profiles), even if the actual rules we discuss only use information about the peaks. Our adaptation of Moulin's characterization is as follows.

Construction. For each coalition $S \in 2^N \setminus \varnothing$, fix an alternative a_S. Define a social choice function in a such a way that, for each preference profile $(\succcurlyeq_1, \ldots, \succcurlyeq_n)$,

$$f(\succcurlyeq_1, \ldots, \succcurlyeq_n) = \inf_{S \subset N} \left[\sup_{i \in S} (a_S, \mathcal{T}(\succcurlyeq_i)) \right]$$

The functions so defined will be called *generalized median voter schemes*.

The values a_S, appear here just as parameters defining functions in this class. Their role becomes more clear under the alternative definition of generalized median voter schemes proposed in Sect. 3.2.

Theorem 2. *(Moulin, 1980) A social choice function on profiles of single-peaked preferences over a totally ordered set is strategy-proof if and only if it is a generalized median voter scheme.*

This characterization can be sharpened if we restrict attention to anonymous social choice functions. In this case, the only strategy-proof rules are those which are indeed based in calculating the medians of agents' votes and some fixed collection of phantoms.

Theorem 3. *(Moulin, 1980) An anonymous social choice function on profiles of single-peaked preferences over a totally ordered set is strategy-proof if and only if there exist $n+1$ points $p_1, \ldots, p_{n+1}$ in A (called the phantom voters), such that, for all profiles,*

$$f(\succcurlyeq_1, \ldots, \succcurlyeq_n) = med(p_1, \ldots, p_{n+1}; \mathcal{T}(\succcurlyeq_1), \ldots, \mathcal{T}(\succcurlyeq_n))$$

(A similar statement, with f defined with only $n-1$ phantoms, characterizes strategy-proof and efficient social choice functions).

3.2 An alternative definition of generalized voter schemes

Generalized median voter schemes are an important class of voting rules, and it will prove useful to provide a second definition of that class. This second

definition is equivalent to the one given above. It is useful when stating and proving some results. It also provides an alternative view on how these rules operate.

To motivate this new definition, let us first consider the case when we must choose among two alternatives only. A possible rule would be to choose 1 unless there is enough support for the opposite, in which case 0 will be selected. What do we mean by "enough support"? We could establish the list of coalitions that will get 0 if all their members prefer it to 1; and it is natural to require that, if a coalition can enforce 0, then its supersets are also able to. Such a family of "winning" coalitions will fully describe the rule; it corresponds to what is called a monotonic simple game.

We now explain how this same idea can be extended to cases where we must select among a finite set of values on the real line (as opposed to only two). Without loss of generality, we can identify these values with a list of integers, from a to b. Let each voter declare her preferred value. Now, we can start by asking whether a should be chosen. If "enough" people have voted for a, then let us choose a. To determine what we mean by "enough", we can give a list of coalitions $C(a)$. If all agents in one of these coalitions support a, then a if chosen. If not, go to $a+1$. Now ask the question whether "enough" agents support values up to $a+1$. That is, look at all agents who support either a or $a+1$, and check whether they form a group in the list $C(a+1)$. If they do, then choose $a+1$. If not, go to $a+2$, and check whether the agents who support a, $a+1$ and $a+2$ form a group in $C(a+2)$. If so, choose $a+2$; if not, proceed to $a+3$, etc. Given appropriate lists of coalitions $C(a)$, $C(a+1), \ldots, C(b-1)$, $C(b)$, the rules described above should lead us to choose some value between a and b, for each list of the agents' preferred values. These lists of coalitions will be called left coalition systems, because the first value to the left of the interval to get enough support is declared to be the choice. (One can similarly describe the rules by a set of right coalition systems, and then start by checking first whether b has enough support, then $b-1$, then $b-2$, etc. In this description, the first value to the right which gets enough support should be chosen). To complete the description of a left coalition system, we need to add a few requirements on the lists of values for $C(.)$, in order to guarantee that the above description makes sense. These requirements are that (1) if a coalition is "strong enough" to support an outcome, its supersets are too; (2) if a coalition is "strong enough" to support the choice of a given value, it is also "strong enough" to support any higher value; and (3) any coalition is "strong enough" to guarantee that the choice will not exceed the maximum possible value b. (Similar requirements must hold for right coalition systems). All of this is summarized by the following formal definitions. Definition 1 formalizes our description of left(right) coalition systems. Definition 2 describes how each of these coalition systems can be applied to produce a generalized median voter scheme. Notice that the parameters as in Moulin's definition of a generalized median voter scheme (Sect. 3.1) correspond to the minimum (or maximum) value of a at which coalition S appears in $C(a)$.

Definition 1. *A left (resp. right) coalition system on the integer interval* $B = [a, b]$ *is a correspondence* $\mathscr{C}$ *assigning to every* $\alpha \in B$ *a collection of non-empty coalitions* $\mathscr{C}(\alpha)$*, satisfying the following requirements:*

1. if $c \in \mathscr{C}(\alpha)$ and $c \subset c'$, then $c' \in \mathscr{C}(\alpha)$;
2. if $\beta > \alpha$ (resp. $\beta < \alpha$) and $c \in \mathscr{C}(\alpha)$, then $c \in \mathscr{C}(\beta)$; and
3. $\mathscr{C}(b) = 2^N \backslash \varnothing$ (resp. $\mathscr{C}(a) = 2^N \backslash \varnothing$).

We'll denote left coalition systems by $\mathscr{L}$, and right coalition systems by $\mathscr{R}$. Elements of $\mathscr{L}$ will be denoted by $l(\cdot)$, and those in $\mathscr{R}$ by $r(\cdot)$.

We can now proceed with our definition of generalized median voter schemes.

Definition 2. *Given a left (resp. right) coalition system* $\mathscr{L}$ *(resp.* $\mathscr{R}$*) on* $B = [a, b]$*, its associated generalized median voter scheme is defined so that, for all profiles* $(\succcurlyeq_1, \ldots, \succcurlyeq_n)$

$$f(\succcurlyeq_1, \ldots, \succcurlyeq_n) = \beta \text{ iff } \{i \mid \mathscr{T}(\succcurlyeq_i) \leq \beta\} \in \mathscr{L}(\beta)$$

and

$$\{i \mid \mathscr{T}(\succcurlyeq_i) \leq \beta - 1\} \notin \mathscr{L}(\beta - 1)$$

(respectively,

$$f(\succcurlyeq_1, \ldots, \succcurlyeq_n) = \beta \text{ iff } \{i \mid \mathscr{T}(\succcurlyeq_i) > \beta\} \in \mathscr{R}(\beta)$$

and

$$\{i \mid \mathscr{T}(\succcurlyeq_i) > \beta + 1\} \notin \mathscr{R}(\beta + 1)$$

Clearly, we could have just referred to either left (or right) coalition system as the primitives in our definitions. To every generalized median voter scheme we can associate one system of each type. Referring to both simultaneously will be useful later on.

Notice that, in order for these rules to be well defined, we only need the alternatives to be linearly ordered and the agents to have a unique maximal alternative. Whether or not the rules have good properties depends then on the domain of preferences over which thus operate.

Example 5. *Let* $B = [1, 2, 3]$, $N = \{1, 2, 3\}$. *Let* $\mathscr{L}(1) = \mathscr{L}(2) = \{S \in 2^N \backslash \varnothing : \# S \geq 2\}$.

Define f *to be the generalized median voter scheme associated with* $\mathscr{L}$. *Then, for example*

$$f(1, 2, 3) = 2$$

$$f(3, 2, 3) = 3$$

$$f(1, 3, 1) = 1$$

This is, in fact, the median voter rule.

Example 6. *Let now* $B = [1, 2, 3, 4]$, $N = \{1, 2, 3\}$. *Consider the right coalition system given by*

$$\mathscr{R}(4) = \mathscr{R}(3) = \mathscr{R}(2) = \{C \in 2^N \backslash \varnothing : 1 \in C \text{ and } 2 \in C\}$$

In that case, both 1 and 2 are essential to determine the outcome. Let g be the generalized median voting scheme associated with $\mathscr{R}$. *Here are some of the values of g:*

$$g(1, 4, 4) = 1$$

$$g(3, 3, 1) = 3$$

$$g(3, 2, 2) = 2$$

3.3 Strategy-proof social choice functions on K-dimensional sets, with generalized single-peaked preferences

The assumption that social alternatives can be represented by a set of linearly ordered values is a very fruitful one. It allows us to represent situations where society must choose among different locations along a road, or a river, or in some space that can be described by a single parameter. But sometimes a spatial location requires at least two coordinates to be properly described. Similarly, the choice of how much to spend on a given project is naturally represented by the costs of the different ways in which this project can be accomplished: this is another natural setting for our one-dimensional model. However, it is more often the case that we can choose among different projects, each one admitting several ways to be accomplished. Then again, a multi-dimensional representation of social alternatives would allow for a much richer representation of the choices open to society. You can think of those characteristics which are crucial to distinguish among alternatives. For example, when choosing among political candidates, you may decide that they can be fully described by their stand on economic, human rights and foreign policy issues, say. Then, candidates could be described by a three dimensional vector, whose first component would describe the candidate's position on the economic dimension, with the second and third standing for the candidate's stand on the other two issues. On each issue, that is, on each of the three dimensions, you should decide how the candidates' stands can be attached a value, from lowest to highest.

The following framework will allow us to formalize multi-dimensional social choices of a rather general sort.

Let K be a number of dimensions. Each dimension will stand for one characteristic that is relevant to the description of social alternatives. Allow for a finite set of admissible $B_k = [a_k, b_k]$ on each dimension $k \in [K]$. Now the set of alternatives can be represented as the Cartesian product $B = \Pi_{k=1}^{K} B_k$. Sets like this B are called K-dimensional boxes. Representing the set of social alternatives as the set of elements in a K-dimensional box allows us to describe many interesting situations. With two dimensions, we can describe location problems in a plane. We can describe political candidates by their positions on

different issues. We can describe alternative plans for a municipality, by specifying which projects could be chosen in each of the different dimensions of concern: schools, safety, sanitation, etc.

There still remains a number of limitations in this specification. One is that we keep assuming that the projects are linearly ordered within each dimension. Another one is that, by assuming that any point in the Cartesian product is a possible choice for society, we are implicitly saying that there are no further constraints on the choices faced by society. We shall later comment on how to relax these assumptions. But the multidimensional model can represent a variety of interesting situations. We first consider what can be said about strategy-proof rules in this setting and will then proceed to other, maybe more realistic ones. (Again, we proceed with a specification that assumes a finite set of alternatives. Similar results can be expressed in a continuous setting (see Border and Jordan 1983, Barberà et al. 1998b).

Before we proceed, we must be specific about the type of restrictions to impose on preferences over such sets of alternatives. We shall maintain the spirit of single-peakedness, by requiring every preference to have a unique top (or ideal) and then assuming that, if z is between x and $\mathscr{T}(\succcurlyeq_i)$, then z is preferred to x. But in order to make the "betweenness" relationship precise, we must take a stand. Following Barberà et al. (1993), we endow the set B with the L_1 norm (the "city block" metric), letting, for each $\alpha \in B$, $\|\alpha\| = \sum_{k=1}^{K} |\alpha_k|$. Then, the minimal box containing two alternatives α and β is defined as $MB(\alpha, \beta) = \{\gamma \in B \mid \|\alpha - \beta\| = \|\alpha - \gamma\| + \|\gamma - \beta\|\}$.

We can interpret that z is "between" alternatives x and $\mathscr{T}(\succcurlyeq_i)$, if $z \in MB(x, \mathscr{T}(\succcurlyeq_i))$. Under this interpretation, the following is a natural extension of single-peakedness.

Definition 3. *A preference $\succcurlyeq_i$ on B is generalized single-peaked iff for all distinct $\beta, \gamma \in B$, $\beta \in MB(\mathscr{T}(\succcurlyeq_i), \gamma)$ implies that $\beta \succ_i \gamma$.*

This definition collapses to that of standard single-peakedness when the set of alternatives is one-dimensional. It implies, and it is in fact equivalent to, the following two conditions: (a) the restriction of generalized single-peaked preference to sets of alternatives that only differ on one dimension is single-peaked, and (b) the projection of the best element on each of this sets is the best element within them.

One possible way to choose from K-dimensional boxes consists in using K (possibly different) generalized median voter schemes, one for each dimension. Then, if each agent is asked for her best alternative, the k^{th} component of her ideal can be combined with the k^{th} component corresponding to other agents, and used to determine a choice, by means of the specific generalized median voter scheme that is attached to this k^{th} component. Similarly, the values for any other component can also be computed, and the resulting K-tuple of values be taken as social outcome.

Formally, we can define (K-dimensional) generalized median voter schemes on $B = \Pi_{k=1}^{K} B_k = \Pi_{k=1}^{K} [a_k, b_k]$, as follows:

Let $\mathscr{L}$ (resp. $\mathscr{R}$) be a family of K left (resp. right) coalition systems, where

each $\mathscr{L}_k$ (resp. $\mathscr{R}_k$) is defined on $[a_k, b_k]$. The corresponding k-dimensional generalized median voter scheme is the one that, for all profiles of preferences on B, chooses

$$f(\succcurlyeq_1, \ldots, \succcurlyeq_n) = \beta \text{ iff } \{i \mid \mathscr{T}(\succcurlyeq_i) \leq \beta_k\} \in \mathscr{L}_k(\beta_k)$$

and

$$\{i \mid \mathscr{T}(\succcurlyeq_i) \leq \beta_{k-1}\} \notin \mathscr{L}(\beta_{k-1}),$$

for all $k = 1, \ldots, K$ (or respectively,

$$f(\succcurlyeq_1, \ldots, \succcurlyeq_n) = \beta \text{ iff } \{i \mid \mathscr{T}(\succcurlyeq_i) \leq \beta_k\} \in \mathscr{R}_k(\beta_k)$$

and

$$\{i \mid \mathscr{T}(\succcurlyeq_i) \leq \beta_{k-1}\} \notin \mathscr{R}(\beta_{k-1}))$$

Example 7. *We can combine examples 5 and 6 in the preceding section, and give an example of a generalized median voter scheme.*

Let $B = [1,2,3] \times [1,2,3,4]$, $N = (1,2,3)$. *Let* $\mathscr{L}_1$ *be as* $\mathscr{L}$ *in example 5. Let* $\mathscr{R}_2$ *be as* $\mathscr{R}$ *in example 6. Let h be the two-dimensional generalized median voter scheme associated to this coalition system. Then, for example,*

$$h((1,1),(2,4),(3,4)) = (2,1)$$

$$h((3,3),(2,3),(3,1)) = (3,3)$$

$$h((1,3),(3,2),(1,2)) = (1,2)$$

Moulin's theorem generalizes nicely to this context. We just need to add a condition on the social choice function, which is usually referred to as voters' sovereignty. This condition requires that each one of the alternatives should be chosen by the function, for some preference profile.

Theorem 4. *(Barberà et al. 1993) A social choice function f defined on the set of generalized single peaked preferences over a K-dimensional box, and respecting voters' sovereignty is strategy-proof iff it is a (K-dimensional) generalized median voter scheme.*

The theorem above applies to the general case where alternatives are elements of any K-dimensional box and voters' preferences are generalized single peaked. A specific instance of this general setup can help us to describe what we have learned. The example is interesting on its own, and it was studied in Barberà et al. (1991). Consider a club composed of N members, who are facing the possibility of choosing new members out of a set of K candidates. Are there any strategy-proof rules that the club can use?

We consider that the club has no capacity constraints, nor any obligation to choose any pre-specified number of candidates. Hence, the set of alternatives faced by the present members consists of all possible subsets of candidates: they can admit any subset. Because of that, it is natural to assume that

the preferences of voters will be defined on these subsets: every member of the club should be able to say whether she prefers to add S, rather than S', to the current membership, or the other way around.

What is the connection between this example and our n-dimensional model? Observe that, given K candidates, we can represent any subset S of candidates by its characteristic vector: that is, by a K-dimensional vector of zeros and ones, where a one in the I^{th} component would mean that the I^{th} candidate is in S, while a zero in the J^{th} component indicates that the J^{th} candidate is not in S. Hence, the set of all subsets of K candidates can be expressed as the Cartesian product of K integer intervals. Each of these intervals would only allow for two values now: $a = 0$, and $b = 1$. The "characteristics" of the alternatives are known once we know what candidates are in and what candidates are out. Therefore, choosing members for a club can be seen as a particular problem within our general class of K-dimensional choice problems.

What about strategy-proofness? We certainly should not expect a general positive answer unless we assume some restriction on preferences. Consider, for example, that there are two candidates x and y, and that I am a voter. I prefer x to y, but since these two candidates would always be fighting if both elected, I prefer nobody to be elected rather than both being in: the latter is my worst alternative. Suppose that, under some voting rule, y will be elected even if I don't support it, while x would only be elected if I add my support to that of other voters. Then, I might not support x, which I like, in order to avoid the bad outcome that both candidates are in! This type of manipulation is almost unavoidable, unless the preferences of voters are restricted in such a way that these strong externalities from having several candidates can be ruled out. One way to do it is by restricting attention to separable preferences.

To check whether a given preference order on sets of candidates is separable, say that a candidate is "good" if it is better to choose this candidate alone than choosing no candidate at all; otherwise, call the candidate "bad" (this, of course, refers to the given preference order). Now, we'll say that the overall order is separable, if whenever we add a "good" candidate g to any set S of candidates, the enlarged set is better than S, and whenever we add a bad candidate b to S, then the enlarged set is worse than S.

In Barberà et al. (1991), it is shown that there exists a wide class of strategy-proof social choice rules when the preferences of club members over sets of candidates are separable. In fact, this is a corollary of Theorem 4 above. This is because, when there are only two possible values for each dimension, the separability assumption we just stated is equivalent to the assumption of generalized single-peakedness for the general case. Then the class of strategy-proof rules we are looking for is the one formed by all possible generalized median voter schemes. But, as we already remarked at the beginning of Sect. 3.2, the left coalition systems corresponding to the case with only two possible values are given by committees, that is, by monotonic families of winning coalitions. As a result, here is the way to guarantee strategy-proofness in our clubs. For each candidate, determine what sets of voters will have enough

strength to bring in that candidate, if they agree to do so. Make sure that if a set is strong enough, so are its supersets. Then, ask each voter to list all the candidates that she likes. Choose all candidates that are supported by a coalition which is strong enough to bring him in. This is a full characterization. In particular, it contains a family of very simple rules, called the quota rules. Fix a number q between 1 and N. Let each agent support as many candidates as she likes. Elect all candidates which receive at least q supporting votes. These rules are not only strategy proof under separable preferences. They are also the only ones to treat all candidates alike (neutrality) and all voters alike (anonymity).

In this section we have provided a characterization of all strategy-proof social choice functions when alternatives can be described as K-tuples of integer numbers and agents' preferences are single-peaked. What if we allow for a richer class of preferences? Certainly no new rules will arise, but will all generalized median voter schemes still be strategy-proof? If not, we could claim single-peakedness to be a maximal domain admitting non-trivial, strategy-proof rules (notice that we can always express dictatorship as an extreme example within the class; hence our reference to nontriviality). In fact single-peakedness does not exactly fit the requirement (see Berga 1998; Barberà et al. 1998a). But the sharper results on the subject are just remarks on fine points, and the basic message one can derive from them is that the hunch that single-peakedness provides a maximal domain is not far off the mark.

3.4 Voting under constraints

Many social decisions are subject to political or economic feasibility constraints. Different feasible alternatives may fulfill different requirements to degrees that are not necessarily compatible among themselves. A community may have enough talent to separately run a great program for the fine arts, or a top quality kindergarten, but not to maintain both programs simultaneously at the same level of excellence. We can still model these constraints within our model, where alternatives are described by K-tuples of integer values, as long as we do no longer require the set of alternatives to be a Cartesian product. For example, if a firm must choose a set of new employees out of $K = \{1, 2, \ldots, k\}$ candidates, the alternative sets can be identified with the elements in the box $B = \Pi_{j=1}^{k}[0, 1]$. But if only three positions are open, and at least one of them must be filled, the feasible set – consisting of K-tuples with at least a nonzero and at most three nonzero components – is no longer a Cartesian product. Similarly, the location of two facilities in some pair of sites out of a set of five municipal plots $(p_1, p_2, \ldots, p_5)$ can be formalized as a choice from $[1, 5] \times [1, 5]$, excluding (by feasibility) the elements with the same first and second component.

Here is how I will formalize the distinction between feasible and conceivable alternatives. Start from any set Z. Let B be the minimal box containing Z. Identify Z with the set of feasible alternatives. Restrict attention to functions whose range is Z. Then by exclusion, interpret the elements of $B \backslash Z$ as

those alternatives that are conceivable but not feasible. Let the agents' preferences be defined on the set Z. Specifically, consider domains of preferences which are restrictions to Z of multidimensional single-peaked preferences on B, with the added requirement that the unconstrained maximal element of these preferences belongs to Z. (This is a limitation, since it rules out interpretations of our model under which preferences would be monotonic on the levels of characteristics. These levels cannot be such that, for all agents, the higher is always the better).

Two major facts can be established in this context (see Barberà et al. 1997b; also Barberà et al. 1998b for a version with a continuum of alternatives). One is that, regardless of the exact shape of the set of feasible alternatives, any strategy-proof social choice function must still be a generalized median voter scheme. Notice, then, that not all generalized median voter schemes will now give rise to well defined social choice functions, because some of these schemes, by choosing the values on different dimensions in a decentralized way, could recommend the choice of non feasible alternatives. Our second result characterizes the set of all generalized median voter schemes that are proper social choice functions, for any $Z \subset B$. This characterization is based on the intersection property, a condition which states that the decision rules operating on different dimensions will be coordinated to always guarantee the choice of a feasible alternative. Before stating it, let us remark that it is not a simple condition, but it provides a full characterization, and it can orient our research for strategy-proof rules for any specification of feasibility constraints.

All of the above is expressed in the following results (Barberà et al. 1997b)

Definition 4. *A generalized median voter scheme f on B respects feasibility on $Z \subset B$ if $f(\succcurlyeq_1, \ldots, \succcurlyeq_n) \subset Z$ for all $(\succcurlyeq_1, \ldots, \succcurlyeq_n)$ such that $\mathscr{T}(\succcurlyeq_i) \in Z$.*

Definition 5. *Let $Z \subset B$ and let f be a generalized median voter scheme on B, defined by the left coalition system $\mathscr{L}$ or, alternatively by the right coalition system $\mathscr{R}$. Let $\alpha \notin Z$ and $S \subset Z$. We say that f has the intersection property for (α, S) iff for every selection $r(\alpha_k)$ and $l(\alpha_k)$ from the sets $\mathscr{R}(\alpha_k)$ and $\mathscr{L}(\alpha_k)$, respectively, we have*

$$\bigcap_{\beta \in S} \left[\left(\bigcup_{k \in M^+(\alpha,\beta)} l(\alpha_k) \right) \cup \left(\bigcup_{k \in M^-(\alpha,\beta)} r(\alpha_k) \right) \right] \neq \varnothing$$

where $M^+(\alpha, \beta) = \{k \in K \mid \beta_k > \alpha_k\}$ and $M^-(\alpha, \beta) = \{k \in K \mid \beta_k < \alpha_k\}$.

We will say that f satisfies the intersection property if it is does for every $(\alpha, S) \in (B - Z, 2^K)$.

Theorem 5. *(Barberà et al. 1997b) Let f be a generalized median voter scheme on B, let $Z \subset B$, and f respect voters' sovereignty on Z. Then f preserves feasibility on Z if and only if satisfies the intersection property.*

Denote by $\mathscr{S}_Z$ the set of all single peaked preferences with top on $Z \subset B$. Let f be an onto social choice function with domain $\mathscr{S}_Z^n$ and range Z.

Theorem 6. *(Barberà et al. 1997b) If $f : \mathcal{S}_Z^n \to Z$ is strategy-proof, then f is a generalized median voter scheme.*

Theorem 7. *(Barberà et al. 1997b) Let $f : \mathcal{S}_Z^n \to Z$ be an onto social choice function. Then f is strategy-proof on $\mathcal{S}_Z^n$ iff it is a generalized median voter scheme satisfying the intersection property.*

3.5 A surprising twist. Back to the Gibbard-Satterthwaite Theorem

One may by now feel to be walking on very narrow grounds. We have specified the alternatives to be a subset of K-dimensional space. We have required the preferences to be single-peaked with their top on the pre-specified subset. We have seen that strategy-proofness requires to use very specific voting rules, satisfying a general and not always easy to interpret condition (the intersection property). The Gibbard-Satterthwaite Theorem is an elegant result, even if it only applies to a specific situation, where all conceivable preferences are admissible. Our last theorem can be interpreted either as a possibility or an impossibility theorem, depending on the range restriction. Indeed, when the set of alternatives is Cartesian, our theorems are quite positive. True, respecting strategy-proofness restricts us to choose among generalized median voter schemes, but these are quite versatile, and different ones can be chosen for different dimensions. On the other hand, for some special shapes of the range, the intersection property becomes highly restrictive, and only very special rules are eligible. Moreover, our theorems apply to preferences which are restrictions to feasible sets of more general preferences, which in turn we assumed to be single-peaked on the minimal box containing our feasible alternatives, and to have their best element within this set. Hence, while the universal domain assumption is quite invariant to the specification of alternatives (modulo their total number), our domain restriction are specific, varying with the set of alternatives under consideration.

Because of all these ifs and buts, it is particularly pleasant to remark that our theorem is, in fact, a very general one, and includes the Gibbard-Satterthwaite as a corollary. The apparent specificity can be otherwise interpreted as a source of versatility, as allowing us to cover many different environments, and the one envisaged by Gibbard-Satterthwaite Theorem in particular.

Consider any finite set of alternatives, with no particular structure. We can always identify them with the k unit vectors in a k-dimensional space. The minimal box containing them is the set $B = \Pi_{k=1}^{K}[0, 1]$. Since no third element in the set of unit vectors U is "between" any other two, any arbitrary order of these unit vectors can be obtained as the restriction to U of a preference with peak on U which is single-peaked on B. Hence, our last theorem applies to social choice functions defined on all preferences over U, with range U. Any strategy-proof social choice function must be a generalized median voter scheme satisfying the implications of the intersection property. These implications are that the same scheme must be used for all dimensions, and that it must be dictatorial. This is the Gibbard-Satterthwaite Theorem. It is not a

separate entity, but the consequence of a much large characterization involving special shapes for the range, specific domain restrictions, and the general structure of the strategy-proof rules.

4 Agent-specific preferences: Rationing and exchange

In the preceding sections, we have studied situations where agents do care for alternatives in ways that are potentially the same for all. The same preferences on A can be held by all agents; unanimity is not required, but it is not precluded.

In other cases, the views of agents are necessarily conflictive. When we must split a dollar, distribute a bunch of desirable objects, decide who will perform an indivisible and unpleasant task, alternatives have to specify how much each one of us may get, or who is to work. Then, alternatives which are best for an agent will typically rank low for others, and unanimity is not to be expected. The sets of admissible preferences for agents over alternatives will not be common, but specific to each agent. In addition it will often be natural to assume that different alternatives assigning the same consequences to one agent are indifferent to her, even if they affect others in quite varied forms. This is the assumption of selfishness, which involves a particular form of specific preferences. Many different collective choice situations are well described by models where the set of agent's preferences over alternatives are specific. The analysis of strategy-proof social choice functions in these context is more intricate that for cases as those considered till now. This is partly due to the fact that unanimity can play a much weaker role in proofs. Another added complication, which we have skipped until now, is that strict preferences over the whole set of alternatives may not be admissible.

I will illustrate the analysis of strategy-proof rules in the specific preferences case by describing results for two closely related models.

4.1 Strategy-proof rationing

We'll now consider cases where a group of agents must share a task or a good. Examples include division of a job among individuals who have collectively agreed to complete it, distributing assets among creditors in a bankruptcy, sharing the cost of a public project or the surplus of a joint venture, or rationing goods traded at fixed prices. Since shares of the total task, or of the total amount of good, are the specific objects of choice, individuals are assumed to have preferences on shares.

Notice here that the alternatives are the distributions of the total among all agents. Since we'll be modelling situations where each agent only cares about her share of the total, preferences will not be common. Agent i will be indifferent among any two alternatives that give her the same share, but j will not be, if these two alternatives give her different amounts. We shall not be insisting on this, and just refer to agents' preferences over their own shares, rather than over complete alternatives. But it is worth making the point here, since the fact will make a difference on the results. We'll examine the class of

problems where preferences of agents are selfish and single-peaked over their own shares. This is well justified if we think of a reduced model, where the task assignment carries some reward, or the share of good one obtained must be paid for. It is then perfectly natural to prefer some amount of the task or good over all other amounts (and their accompanying rewards/costs), and to consider other amounts the better the closer to their ideal. In fact, this would be a consequence of assuming convex, increasing preferences in the effort/reward on good/cost space, and of having the agent choose her share on a convex bounded set.

Formally, for finite set of agents $N = \{1, \ldots, n\}$, allotments will be n-tuples a in the set $A = \{a \in [0, 1^n] \mid \sum_{i \in N} a_i = 1\}$. Preferences being selfish and continuous, they can be identified with continuous utility functions on $[0, 1]$, denoted by $u_i, u_i', u_j, \ldots$. These utility functions will represent single peaked preferences. That is, for each u_i there will be some $x^* \in [0, 1]$ such that, for any $y, z \in [0, 1]$,

$$x^* < y < z \Rightarrow u_i(x^*) > u_i(y) > u_i(z),$$

and,

$$x^* > y > z \Rightarrow u_i(x^*) > u_i(y) > u_i(z).$$

Denote by S the set of all continuous single-peaked utility functions on $[0, 1]$. We'll be interested in allotment rules of the form

$$f : S^n \to [0, 1]^n$$

with

$$\sum_{i \in N} f_i(u) = 1 \quad \text{for all } u \in S^n$$

Notice that u stands for a profile of preferences, $(u_1, \ldots, u_n)$. The value of $f_i(u)$ is the share that goes to i under preference profile u, given rule f.

Some standard requirements, like efficiency and anonymity, can be applied to allotment rules. Efficiency requires that the selected allotment be Pareto efficient at each preference profile. When coupled with the requirement that preferences are single-peaked it is equivalent to the following: at each preferences profile, agents that do not get exactly their ideal point must either all get less than what they wished, or all get more.

Anonymity is a property of symmetric treatment for all agents: for all permutations π of N (π is a function from N onto N) and $u \in S^n$, $f_{\pi(i)}(u^\pi) = f_i(u)$, where $u^\pi = (u_{\pi^{-1}(1)}, \ldots, u_{\pi^{-1}(n)})$. As we shall discuss, anonymity may or may not be an attractive property of allotment rules, depending on the a priori rights of the agents involved.

Finally, in our context, strategy-proofness can be written as the requirement that, for all $i \in N$, $u \in S^n$ and $v_i \in S$,

$$u_i(f_i(u)) \geq u_i(f_i(u_{-i}, v_i))$$

An elegant result due to Sprumont (1991) provides a full characterization of allotment rules satisfying the three requirements above. Actually, only one rule can satisfy all three simultaneously.

Theorem 8. *(Sprumont 1991) An allotment rule is efficient, strategy-proof and anonymous if and only if it is the uniform rule f^* defined by*

$$f_i^*(u) = \begin{cases} \min[x^*(u_i), \lambda(u)] & \text{if } \sum_{i \in N} x^*(u_i) \geq 1 \\ \max[x^*(u_i), \mu(u)] & \text{if } \sum_{i \in N} x^*(u_i) \leq 1 \end{cases}$$

where $\lambda(u)$ *solves* $\sum_{i \in N} \min[x^*(u_i), \lambda(u)] = 1$ *and* $\mu(u)$ *solves* $\sum_{i \in N} \max[x^*(u_i), \mu(u)] = 1$.

In order to relate this result to previous ones, as well as to understand its possible extensions to the non anonymous case, let us take a second look at the case where only two individuals must share. This case does not capture all the features of the problem, but gives us some interesting hints. With two agents, the allotment is fully described by a_1 since $a_2 = 1 - a_1$. Hence, the preferences of agent 2 can be expressed as preferences on a_1, as well, by letting $\tilde{u}_2(a_1) = u_2(1 - a_1)$. Clearly, $\tilde{u}_2$ is continuous and single-peaked whenever u_2 is. The allotment problem is now reduced to choosing a single point in $[0, 1]$ when both agents have preferences which are single-peaked over the same variable. We have already seen that anonymity and strategy-proofness force us to use the rule that chooses medians among the agents' peaks and one phantom. By symmetry, this phantom must be at $\frac{1}{2}$ in our case. It is easily seen that this is exactly the uniform rule for this simple case.

We can interpret the rule as giving each agent the implicit right to guarantee herself the (one half-one half) distribution. From this guaranteed level, mutually desired improvements can be achieved. A similar interpretation for the n-person case would start by guaranteeing the egalitarian share ($\frac{1}{n}$ of the total) to each agent. Changing these guaranteed levels, while keeping the possibility of mutually consented changes away from them, would be a natural way to eliminate anonymity while keeping efficiency and strategy-proofness. In particular, for the two person case, this would be equivalent to maintain the median rule, but have a phantom at any point $p \in [0, 1]$ different than $\frac{1}{2}$, thus guaranteeing agent 1 the share of p, and the agent 2 the share $1 - p$.

But, why drop anonymity at all? The reason is that, in many situation, people may have different rights or entitlements: these may be respected, while allowing agents who do not want to use them to pass on their rights and allow others to enjoy what they don't need. Age, seniority, previous contribution, all are examples of criteria calling for possibly non-symmetric treatment of agents, while efficiency and strategy-proofness are still desirable. Surprisingly, there is only one anonymous rule satisfying efficiency and strategy-proofness,

but there is a continuum of non-anonymous rules with the two latter properties. One of the apparent reasons, is that implicit rights can vary; moreover, they can change quite independently in the cases of excess demand from these of excess supply. To see this, let us take a final look at the uniform allotment rule, and at its possible modifications. This time we can look at an example, with $n = 5$ agents with ideal points $x_1^* = \frac{3}{20}$, $x_2^* = \frac{2}{20}$, $x_3^* = \frac{5}{20}$, $x_4^* = \frac{6}{20}$, $x_5^* = \frac{14}{20}$. The outcome prescribed by the uniform rule can be reached through the following algorithm (see Sönmez 1994):

Step 1. Determine whether $\sum_{i \in N} x^*(u_i)$ equals, exceeds, or falls short of 1. If $\sum_{i \in N} x^*(u_i) = 1$, then allot shares equal to the ideal points. If $\sum_{i \in N} x^*(u_i) > 1$, allot their ideal points to those agents who demand no more than $\frac{1}{n}$. If $\sum_{i \in N} x^*(u_i) < 1$, allot their ideal point to those agents who demand at least $\frac{1}{n}$. In our case, $\sum_{i \in N} x^*(u_i) > 1$, and agent 1 and 3's ideal points are less than $\frac{1}{5}$. Thus, $a_1 = \frac{3}{20}$ and $a_3 = \frac{2}{20}$.

Step 2. Determine the remaining number of agents to be allotted and the remaining share to be allotted. Say, there are k agents and an amount s to be shared. Perform the same procedure as in Step 1, letting s replace 1 and considering only the k agents. Iterate this step until all the k' remaining agents have ideal points exceeding (or falling short of) $\frac{s'}{k'}$.

In our case, $k = 3$ and $s = \frac{15}{20}$. Agent 2 is allotted $a_2 = \frac{5}{20}$. There are now $k' = 2$ agents remaining with $s' = \frac{10}{20}$. Each has an ideal point which exceeds $\frac{s'}{k'} = \frac{5}{20}$.

Step 3. Allot the remaining $\frac{s'}{k'}$ each.

In our case $a_4 = a_5 = \frac{5}{20}$.

We conclude that agents are allotted the shares $\left(\frac{3}{20}, \frac{5}{20}, \frac{2}{20}, \frac{5}{20}, \frac{5}{20}\right)$, which corresponds to the outcome of the uniform rule with $\lambda(u) = \frac{5}{20} = \frac{1}{4}$.

The above description suggests possible ways to create new non anonymous allotment rules in similar ways (and thus with good chances to still buy strategy-proofness and efficiency).

1. Rather than have $\frac{1}{n}$ as a starting reference point, choose any collection of shares q_i such that $\sum_{i \in N} q_i = 1$.
2. Rather than having the same reference point for the cases of $\sum_{i \in N} x^*(u_i) < 1$ and $\sum_{i \in N} x^*(u_i) > 1$, choose different reference points q_i^L and q_i^H.

3. Let the reference levels depend on the share remaining in each iteration of Step 2 (with enough qualifications on the form of this dependence, in order to preserve strategy-proofness).

The above remarks can lead to a characterization of wide classes of efficient and strategy-proof allotment rules. This is done in Barberà et al. (1997a), although the article also presents examples which indicate the need for qualifications of the suggested steps for technical reasons. But the essence is in what we have described: there are many reasonable and quite satisfactory ways to design allotment rules, if we can expect preferences on shares to be single-peaked.

Unfortunately, this property cannot be expected to hold for agent's preferences on richer types of alternatives, and in particular in the traditional case of exchange economies, where more than one good is to be distributed. We discuss this in the next session.

5 Strategy-proof exchange

One of the most classical models in economics is that of an exchange economy. There are n consumers holding initial endowments of l private goods. No production takes place. Consumers can exchange among themselves and reallocate the existing amounts of goods. This model emphasizes that preference diversity is an important basis for the existence of mutually advantageous trade among economic agents. Because of that, it is also an important testing ground for questions on preference revelation. The cost of strategy-proofness in exchange economies is efficiency. It has been shown that in exchange economies strategy-proof social choice functions which are efficient are also dictatorial. Hurwicz (1972) proved that result for two agents and two goods, for functions satisfying the added requirement of individual rationality with respect to the initial endowment. Zhou (1991) proved that this negative result holds for two people even without the assumption of individual rationality. Serizawa (1998) has recently extended Hurwicz's result to economies with any finite number of agents.

These negative results are important, because they point at some unavoidable trade-offs. But, if we want to go beyond, and perform any kind of second best analysis, it is worth pursuing matters a little further. Suppose we can characterize all the social choice functions that are strategy-proof in exchange economies, as we already have done in voting contexts. Clearly, no reasonable rule within this class will be efficient (we exclude dictatorial rules as unreasonable). But some may be more efficient than others, or less inefficient. This may also be qualified in reference to some additional information, regarding the number of agents, the distribution of preferences, or any other relevant parameters.

I will report on some of the existing characterization results for strategy-proof social choice rules in exchange economies. They are important because

they give us a catalog of those mechanisms which can satisfy strategy-proofness in full. Then, we may want to choose among them those that satisfy other interesting properties to some satisfactory extent. Of course, one could start by characterizing the set of rules that satisfy some alternative properties, and then select among them those which are "closest" to satisfy strategy-proofness. While this is also possible, it raises the question of what we mean by "approximate strategy-proofness." We shall not go deeply in that direction, but this is a good moment to mention the issue, in connection with the possibility of using the Walrasian mechanism, or some procedure related to it, and expect them to have good incentive properties, especially for economies with a large number of agents. Roberts and Postlewaite (1976) provide conditions under which the gains from manipulating the Walrasian mechanism become small as the economy grows large. However, small gains will still justify deviations by maximizing agents, and these deviations may have meaningful impacts when aggregated across a large population. Jackson (1992) and Jackson and Manelli (1997) investigate the size and impact of these deviations on the final equilibrium outcome, relative to the truthful one. They show that, under general conditions, each agent's deviations, as well as their aggregate impact, will again become small as the economy grows large. Let me also mention two related papers, one by Cordoba and Hammond (1998), the other by Kovalenkov (1997). Rather than concentrate on the Walrasian mechanism, which is manipulable for any finite economy, the latter papers describe variants of this mechanism that would be strategy proof (although not always balanced) for finite economies. Then, each of these strategy proof mechanisms are shown to be "approximately balanced" and "approximately Walrasian" when the number of agents is large. These papers nicely complement the previous ones in their attempt to capture the incentive properties of the Walrasian mechanism. One set of papers tends to support the statement that, for large economies, the Walrasian mechanism will perform approximately as if it was strategy-proof. The other set supports the statement that, again for large enough economies, some strategy-proof mechanisms will become approximately Walrasian.

As announced, we'll consider economies with l private goods and n consumers. The endowment of goods is denoted by $e = (e^1, \ldots, e^n) \in R_+^{nl}$. An allocation is a list $x = (x^1, \ldots, x^n)$ of goods received by each agent, and the set of balanced allocations constitutes the set of alternatives to choose from

$$A = \left\{ x \in R_+^{nl} \mid \sum_i x^i = \sum_i e^i \right\}.$$

To keep within our general framework, agents should be endowed with preferences over the set of alternatives, that is, on the set of full allocations. But since we'll limit ourselves to the analysis of situations when preferences are selfish, we'll resort to the traditional formulation in general equilibrium theory, where agents are attributed preferences over the set of admissible consumption vectors (elements in R_+^l, in our case). We assume that the preferences satisfy some further restrictions of convexity and monotonicity. But

selfishness itself is also a restriction: all together, these conditions on preferences define the restricted domain for which we'll discuss the possibility of strategy-proof social choice functions. The preferences of agent i are represented by a utility function $u^i : R^l_+ \to R$. U denotes the set of all continuous, strictly quasi-concave and increasing u^i's.

5.1 Two agent, two good exchange economies

This is a particular case which turns out quite easy to analyze completely, and ties in the specific preference case with the common preference one.

To get a feeling for the general results, let us first consider a specific set of rules that will result in a strategy-proof social choice function (the rules define a game under which declaring the truthful preferences is a dominant strategy; the social choice rule is then the one assigning to each preference profile the outcome of this game under truthful strategies).

Example 8. *Fix a positive price p, and allow each agent i to select her best alternative out of the set* $B(p, e_i) = \{(x_{i1}, x_{i2}) \,|\, px_{i1} + x_{i2} = pe_{i1} + e_{i2}\}$. *This describes the supply/demand of both the agents for both goods. If both agents have excess demand/supply of the same good, the final allocation is e. If the excess demand/supply allow for mutually advantageous trade at price p, then the prescribed allocation is the one where the agent who is less inclined to trade maximizes her utility.*

This rule of voluntary trading at fixed prices and with rationing on the short side offers no advantage to manipulation. Since prices are fixed and the rationing rule is not sensitive to the size of unsatisfied demands, it is best for all agents to express what they want. The associated social choice rule is clearly strategy-proof. In what follows, we'll describe other strategy-proof rules for exchange economies. But the essential insights can always be referred back to this simple example.

Remark that, in this simple case, we can identify the exchange economies problem with one of choosing the level of a public good (hence connecting the problem of common preferences with those of specific preferences). This is because, once we have fixed a price ratio p and endowments e, choosing the level of one good for one agent, say x^1_1, fully determines the levels of x^1_2, x^2_1, x^2_2. Moreover, the preferences of agent 1 over values of x^1_1 compatible with allocation on the budget line are single peaked (because u^1 is quasi-concave and monotonic), and so are the preferences of agent 2 over the same x^1_1 values (which automatically determine $2's$ consumptions). Therefore, our allocation problem reduces to the choice of one value on a totally ordered set, with two agents and single-peaked preferences. Our rule above can be simply rewritten as one picking the median between the best values of x^1_1 for 1 and 2 and the value e_{11} of 1's endowment for good 1. With e_{11} as phantom voter, this is one of the median voter rules we have already identified in Part 1!

Trading at one fixed price has some features that are essential to any strategy-proof rule. Others can be dispensed with, to get a some general result.

Remark that any budget line corresponding to a positive price ratio defines a diagonal set within the set of allocations, in the following sense:

Definition 6. *A set $D \subset A$ is diagonal iff for each agent i and for all $x, y \in D$, $(x \neq y)$, $x^i \not\geq y^i$ and $y^i \not\geq x^i$.*

Diagonality of the range rules out the possibility of some agent i getting more of all the goods in one allocation in the range than i would get at another allocation also in the range.

In our case, the budget line corresponding to the fixed price is, indeed, the range of our social choice function. Diagonality of the range is necessary for a social choice function on 2×2 exchange economies to be strategy-proof. The use of only one price is not, as shown by the following examples.

We begin by a numerical example.

Example 9. *(See Fig. 1) Agent 1 is endowed with ten units of each of the two goods and agent 2 is endowed with five units of each of the two goods. Agent 1 may offer to buy good one at a price of 2 (units of good two per unit of good one) and sell good one at a price of $\frac{1}{2}$. If, for instance, agent 1 finds buying 3 units of good one most preferred (u^1 in Fig. 1), then agent 1's dominant strategy is to offer to buy up to 3 units of good one. If agent 2 has the utility function u^2*

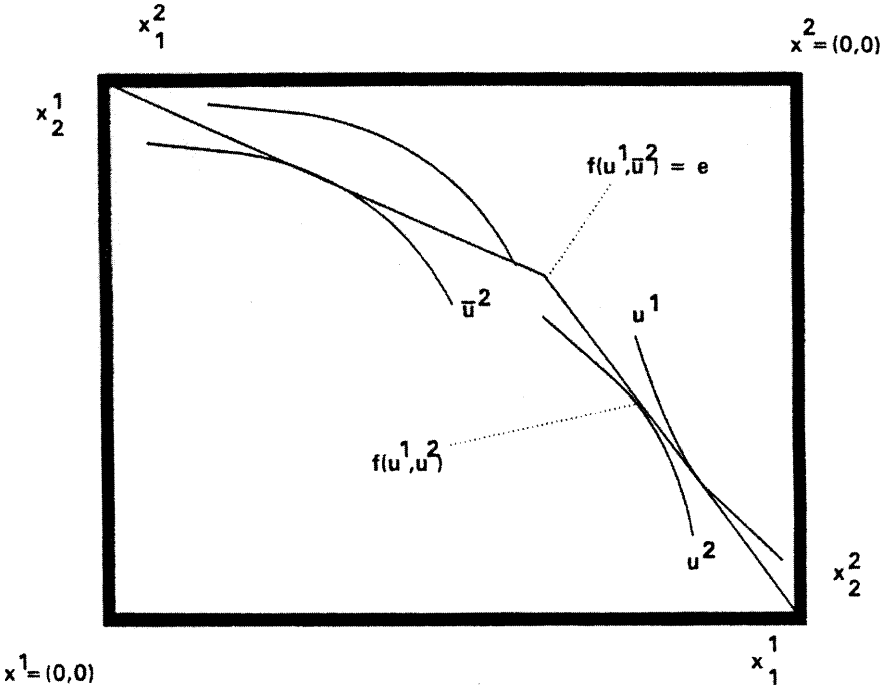

Fig. 1.

in Fig. 1, then his or her dominant strategy is to offer to sell up to 2 units of good one (at a price of 2) and buy up to 1 unit of good one (at a price of $\frac{1}{2}$*). In this case, the outcome of the fixed-price procedure would be that agent 2 sells 2 units of good one to agent 1 at a price of 2. The final allocation for u is (12, 6) for agent 1 and (3, 9) for agent 2. If instead, agent 2 has the utility function* $\bar{u}^2$ *(see Fig. 1 again), then he or she will not offer to sell good one, but will offer to buy up to 2 units of good one. In this case, no goods are exchanged and the final allocation is the initial endowment.*

The principles underlying the case above can be rewritten in a more general form. This is the purpose of our next example.

Example 10. *A two-price rule. Endow agents with any initial endowments* e^i*. Select one of the two agents, say 1. Choose two prices for the first good in terms of the second good. Interpret the first (and lowest) as the price at which agent 1 can offer to sell good 1; the second (and highest) as the price at which he can offer to buy the same good. Given her preferences, agent 1 can then choose to offer some amount of the first good (up to her endowment), or else post an offer to buy some amount (only one of these choices will be optimal, given that preferences are quasi-concave and the selling price is not higher than the lower one; agent 2, given these prices, might have had two best strategies – one of them selling and the other buying – but we do not allow this to be relevant). If agent 1 has made an offer to buy and 2 is ready to sell at the buying price, exchange takes place to the extent of the lowest willingness to transact (and up to 2's endowment). The same occurs if agent 1 wants to sell and 2 is ready to buy.*

Examples 9 and 10 implicitly describe, again, a strategy-proof social choice function. Their ranges are still diagonal. Also, the preferences of each agent over their option sets (given the declared preferences of the others) are single-peaked (our previous remark on two maxima already hinted that overall preferences of agent 2 on the whole range may no longer be single-peaked). In these two examples, as well as in Example 8, each agent can guarantee herself, by declaring the true preferences, that the social outcome will be at least as good as her initial endowment. The rules are individually rational. Subject to this qualification, the second class of rules we just described through Examples 9 and 10 (to be called the double fixed price rules) are the only ones to guarantee strategy-proofness in 2×2 exchange economies. Example 8, which is a special case where the selling and the buying prices coincide (to be called the single fixed price rule) constitutes the subclass which, in addition, respects anonymity (i.e., allows both agents to play symmetric roles), provided the initial endowments of both agents are also identical.

Theorem 9. *(Barberà and Jackson 1995) A social choice function on a two-good, two-agent economy is strategy-proof and individually rational iff it is the outcome of a double fixed price exchange mechanism. If, in addition, the function is anonymous, then it is the outcome of a single fixed price mechanism from the egalitarian endowment.*

The above theorem would need some qualifications to be rigorously stated (a more precise statement of the next theorem will cover this one as a special case). Some bounds on the traded amounts can be exogenously fixed, and a formal definition of what we have presented informally as examples would be in order. The interested reader is referred to Barberà and Jackson (1995). But the essentials are laid down.

Notice that all the functions covered by the theorem have very narrow ranges, and that these are formed by linear prices. This is a consequence of the fact that the domain of preferences includes all strictly quasi-concave utility functions and that both agents play a role in determining the outcome. Dictatorship is excluded by individual rationality. Since dictatorial rules are obviously strategy proof, a full characterization of rules satisfying this property would require to drop the individual rationality assumption (for this, we refer again to Barberà and Jackson 1995).

5.2 Two agents, l goods

Having learned about the 2×2 case, we can now ask whether the same basic ideas extend to exchange economies with more than two goods. We still retain the case of two agents, since this (plus our focus on distributing all the resources) allows us to describe the full allocation once we know what one agent gets.

Fixed prices were the key to strategy-proof exchange with two goods: the range is a line. But fixed prices, when there are more than two goods, describe budget sets whose boundaries are hyperplanes. However restrictive, these rules allow too much flexibility. Strategy-proofness cannot be satisfied if agents are allowed to express their preferences on such large sets. What happens is that, for two goods, the notion of fixed prices and fixed proportions are equivalent. For more than two goods, strategy-proof social choice functions can be based on limited trades, along some collection of fixed proportions satisfying some additional properties. In fact, they must be of this particular form if they are to satisfy individual rationality.

Let us express these ideas more formally, after an example.

Example 11. *There are two agents and three goods. Endowments are* $e^1 = e^2 = (5,5,5)$*, for a total* $(10,10,10)$ *resource vector. Agent 1 can buy units of any of the goods from agent 2, provided she pays one unit of each of the remaining two goods. Hence, agent 1 can offer multiples (but not combinations) of the trades* $(1,-1,-1)$, $(-1,-1,1)$, $(-1,1,-1)$. *The range of* f *in terms of* 1*'s final allocation, is* $r_f = \{x \mid \exists \gamma \in [0,1]$ *s.t.* $x' = \gamma(5,5,5) + (1-\gamma)(10,0,0)$*, or* $x' = \gamma(5,5,5) + (1-\gamma)(0,10,0)$*, or* $x' = \gamma(5,5,5) + (1-\gamma)(0,0,10)\}$. *If agent 1's must preferred point in the range is, say,* $(7,3,3)$*, then the allocation is 2's most preferred point from the convex combination of* $(5,5,5)$ *and* $(7,3,3)$ *(allocations are expressed in terms of what agent 1 gets. Then 2 gets* $(10,10,10) - x'$*).*

Notice that, given the structure of the range, agent 1 always has a unique most preferred point, and that all convex combinations of that point and the

endowment are preferred to all other points on any other segment. It is this property, that the agent who actually chooses the possible trades actually prefers these trades to any other, that makes the fixed proportion trading strategy-proof.

We now provide, after some preliminaries, a more formal definition of fixed proportion trading.

Preliminaries. Given points a and b in A, we write ab to denote the set of points lying on the segment with endpoints a and b, so $ab = \{x \mid \exists \gamma \in [0,1], x = \gamma a + (1-\gamma)b\}$. We write $c \geq_i^* ab$ if $c^i \geq \gamma a^i + (1-\gamma)b^i$ for some $\gamma \in [0,1]$. Then $c \geq_i^* ab$ indicates that c lies above the segment ab from agent $i's$ perspective.

Given a set $B \subset A$ and a utility profile $u \in U^2$, let $T^i(B,u)$ denote the set of allocations in B which maximize u^i. This set is non empty if B is closed. A function t^i, which is a selection from T^i, is called a tie-breaking rule. A tie-breaking rule t^i is j-favourable at $B \in r_f$ if for any u, $t^i(B,u) \neq t^i(B,u^{-j},v^j)$ only if $v^j(t^i(B,u^{-j},v^j)) \geq v^j(t^i(B,u))$.

Definition 7. *A social choice function f defined on a two-agent exchange economy is the result of fixed proportion trading if r_f the range of f, is closed, diagonal and contains e, and there exists an agent i such that the following holds:*

1. *for all distinct x and y in r_f either $x \in ey$, $y \in ex$ or $e \geq_i^* xy$;*
2. *there exist tie-breaking rules t^i and t^j such that t^i is j-favourable at r_f and t^j is i-favourable at $ea \cap r_f$, for all $e \in r_f$*
3. *$f(u) = t^j(ea \cap r_f, u)$, where $a = t^i(r_f, u)$.*

Condition (1) assures that r_f lies along $k \leq l$ diagonal line segments, each having the endowment as an endpoint. If one chooses x from one segment and y from another segment, then $e \geq_i^* xy$. Condition (2) states that tie-breaking rules either are constant or choose in favor of the other agent. This condition only comes into play if the range is not connected, since then agents might have two possible utility maximizing choices. (This is an aspect that we have not emphasized in our previous informal discussion; it is needed for complete characterization and certainly complicates matters, but does not change anything essential). Condition (3) states that the outcome of f is agent j's most preferred point in the range, which lies between the endowment and agent $i's$ most preferred point in the range. We can now state

Theorem 10. *(Barberà and Jackson 1995) A two-person social choice function is strategy-proof and individually rational iff it is the result of fixed proportion trading.*

5.3 Three or more agents

With three or more agents, it is no longer the case that what one of them gets is determinant of the global allocation. This opens up new possibilities for strategy-proof rules, some of which are not necessarily attractive. For example, agent 1 might be offered to choose her best among some feasible baskets

of goods. Then, either agent 2 or agent 3 might get the remaining resources, with the beneficiary being determined, say, according to which one of the rejected baskets agent 1 declares to be her worst. This rule is clearly strategy-proof, since agent 1 guarantees herself the best attainable basket, and does not care who gets the rest; while the other two agents cannot help 1's choice of who will be the lucky one.

This is an example of a bossy social choice function, i.e., one where some essential part of the allocation is trusted to an agent who is unaffected by the choice, while affecting the utility of the others. Bossy functions were described by Satterthwaite and Sonnenschein (1981). They are usually considered unattractive, and efforts to characterize strategy-proof social choice functions have concentrated on finding rules not in this class, called non-bossy.

Barberà and Jackson (1995) provide a characterization of non-bossy strategy-proof social choice functions satisfying a version of anonymity and some additional technical conditions. Although the characterization becomes rather involved, it is in the spirit of the results we have described for two agent exchange economies: strategy-proofness requires a limited range of possible exchange, does not allow from trade to be exploited, and thus enters in conflict with efficiency.

6 Concluding remarks

We have described the characterizations of classes of social choice functions for different models; each model represents some family of economic or political situations where collective decisions must be adopted. The nature of these decisions suggests that alternatives may have a structure, which varies from one model to another; this structure of the alternatives, and the underlying situation we try to model, will usually suggest the class of admissible preferences relative to which our analysis of strategy-proofness takes place.

To emphasize the common thread in our approach, let us reconsider, for a last time, the notion of strategy-proofness.

Given a social choice function defined over a domain of preference profiles, we can naturally define a game form describing the strategic possibilities of agents who participate in this social choice. Preferences in the domain of the function are the possible strategies. Alternatives in its range are the outcomes. The social choices for each n-tuple of strategies define an outcome function. We can fix any profile of preferences in the domain and interpret them as the actual preferences of agents; each choice of profile, along with the previously defined game form, defines a game. The incentive properties of the social choice function are given by the solutions of the games in this class. Implementation theory deals with the connection between such solutions (taken to be the prediction of behavioral analysis) and the normative desiderata of the society, as expressed by the social choice function.

One particular important question within this framework is the following. Since every specific preference profile, once interpreted as the set of "actual"

preferences, defines one game in the class, and it is also an n-tuple of strategies in this game, is this n-tuple formed by dominant strategies for each player? If so, we say that the social choice function generating the class of games in question is strategy-proof. If not, the function is manipulable.

It is well known that games where all agents have dominant strategies are rare and it should come as no surprise that classes of games having all this very strong property are not easy to find. What we have illustrated is that, in spite of these small odds, it is possible to examine the question systematically for many different social choice functions, once we understand that these objects are defined on specific domains, and that the extent of the domain is a crucial element to determine the class of social choice functions that may satisfy strategy-proofness. In general terms, only rather trivial social choice functions can be strategy-proof when the domains of admissible preferences are "large". Yet, nontrivial social choice functions are known for economic and political analysis. Our examples have been chosen to illustrate both the positive and the negative side of this picture. Generalized median voter schemes, or allotment rules like the uniform rule and its non anonymous extensions, prove that it is worth examining each interesting environment without the prejudice that no interesting rule is strategy-proof. On the other hand, the Gibbard-Satterthwaite Theorem, or the characterization of the narrow and inefficient methods which are strategy-proof for exchange economies, remind us that, in many situations, rational agents will find themselves endowed with rich strategic possibilities, if they are ready to use their private information as a form to gain advantage over other participants in collective decision processes.

Let me finish by making it very clear that there is a vast literature on strategy-proofness, and the choice of examples has been biased. I have, in particular, concentrated only on the problem of disclosing the preferences of agents. While preferences are certainly an important part in the description of the relevant characteristics of agents, other aspects may be private information for these. We have skipped anything having to do with revealing one's abilities, an issue that becomes essential when studying productive processes. Even in worlds with trivial preferences we may need to elicit what people can do in order to know how to compensate them, how to allocate scarce resources among them, what production plans to shoot at. There is a lot done, and a lot still to do in this topic. There are also many results and some open questions in the narrow topic that I have focused on, by assuming that available alternatives are already fixed, and that preferences are all that matter in order to choose among them. The reader is referred to the surveys by Moore (1992), Sprumont (1995), Barberà (1996), and Moulin (1996) for further discussion.

References

Arrow K (1951) Social choice and individual values. Cowles Foundation Monograph, Yale University Press

Barberà S (1983) Strategy-proofness and pivotal voters: A direct proof of the Gibbard-Satterthwaite Theorem. Int Econ Rev 24(2): 413–418

Barberà S (1996) Notes on strategy-proof social choice functions. In: Arrow K, Sen A, Suzumura K (eds) Social choice re-examined. Macmillan, London. *French version*: Sur les fonctions de choix non manipulables. Revue d'Économie Politique 106(1): 1996
Barberà S, Gul F, Stacchetti E (1993) Generalized median voter schemes and committees. J Econ Theory 61: 262–289
Barberà S, Jackson M (1994) A characterization of strategy-proof social choice functions for economies with pure public goods. Soc Choice Welfare 11: 241–252
Barberà S, Jackson M (1995) Strategy-proof exchange. Econometrica 63: 51–87
Barberà S, Jackson M, Neme A (1997a) Strategy-proof allotment rules. Games Econ Beh 18: 1–21
Barberà S, Massó J, Neme A (1997b) Voting under constraints. J Econ Theory 76: 298–321
Barberà S, Massó J, Neme A (1998a) Maximal domains of preferences preserving strategy-proofness for generalized median voter schemes. Soc Choice Welfare 16: 321–336
Barberà S, Massó J, Serizawa S (1998b) Strategy-proof voting on compact ranges. Games Econ Beh 25: 272–291
Barberà S, Peleg B (1990) Strategy-proof voting schemes with continuous preferences. Soc Choice Welfare 7: 31–38
Barberà S, Sonnenschein H, Zhou L (1991) Voting by committees. Econometrica 59: 595–609
Berga D (1998) Strategy-proofness and single-plateaued preferences. Math Soc Sci 35: 105–120
Black D (1948) On the rationale of group decision making. J Pol Econ 56: 23–34
Border K, Jordan JS (1983) Straightforward elections, unanimity and phantom voters. Rev Econ Stud 50: 153–170
Cordoba JM, Hammond PJ (1998) Asymptotically strategy-proof Walrasian exchange. Math Soc Sci 36: 185–212
Farquharson R (1969) Theory of voting. Yale University Press, New Haven
Gibbard A (1973) Manipulation of voting schemes: A general result. Econometrica 41: 587–601
Hurwicz L (1972) On informationally decentralized systems. In: McGuire GB, Radner R (eds) Decision and organization. North Holland, Amsterdam
Jackson M (1992) Incentive compatibility and competitive allocation. Econ Lett 40: 299–302
Jackson M (2001) A crash course in implementation theory. This issue
Jackson M, Manelli A (1997) Approximately competitive equilibria in large finite economies. J Econ Theory 77: 354–376
Kovalenkov A (1997) On a 'folk' strategy-proof approximately Walrasian mechanism. Universitat Autònoma de Barcelona, Preprint
Moore J (1992) Implementation, contracts, and renegotiation in environments with complete information. In: Laffont JJ (ed) Advances in economic theory. Cambridge University Press
Moulin H (1980) On strategy-proofness and single peakedness. Publ Choice 35: 437–455
Moulin H (1996) Procedural cum endstate justice: An implementation viewpoint. Department of Economics, Duke University, memo
Roberts DJ, Postlewaite A (1976) The incentives for price-taking behavior in large exchange economies. Econometrica 44(1): 115–127
Satterthwaite M (1975) Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. J Econ Theory 10: 187–217
Satterthwaite M, Sonnenschein H (1981) Strategy-proof allocation mechanisms at differentiable points. Rev Econ Stud 48: 587–597

Schmeidler D, Sonnenschein H (1978) Decision theory and social ethics, issues in social choice. Dordrecht, D. Reidel Publishing Company
Serizawa S (1998) Inefficiency of strategy-proof rules for pure exchange economies. mimeo
Sönmez T (1994) Consistency, monotonicity, and the uniform rule. Econ Lett 46: 229–235
Sprumont Y (1991) The division problem with single-peaked preferences: A characterization of the uniform allocation rule. Econometrica 59: 509–519
Sprumont Y (1995) Strategy-proof collective choice in economic and political environments. Can J Econ XXVIII: 68–107
Zhou L (1991) Inefficiency of strategy-proof allocation mechanisms in pure exchange economies. Soc Choice Welfare 8: 247–257