barcelona gse
graduate school of economics
RESEARCH NETWORK

Strategy-proof social choice

Salvador Barberà

February 2, 2010

Barcelona Economics Working Paper Series

Working Paper nº 420

Strategy-proof Social Choice

Salvador Barberà
MOVE, Universitat Autònoma de Barcelona and Barcelona GSE
Edifici B, 08193 Bellaterra, Spain
E-mail: salvador.barbera@uab.cat

January 29, 2010

Many people have helped, directly and indirectly, to write this survey, and my co-authors in particular. Some colleagues and friends read it at some point and provided advise and criticism. They include Matt Jackson, Michel Le Breton Jordi Massó, Hervé Moulin, Klaus Nehring, Szilvia Pappai, Clemens Puppe and John Weymark. None of them is responsible for the final shape and contents of this piece, as I could only very partially take their comments into account. I received very valuable assistance from Conxa Iglesias, Miguel Angel Alcobendas and Diego Caramuta at different stages of the project.
Salvador Barberà gratefully acknowledges support from the Spanish Ministry of Science and Innovation through grant "Consolidated Group-C" ECO2008-04756, from the Generalitat de Catalunya, Departament d'Universitats, Recerca i Societat de la Informació through the Distinció per a la Promoció de la Recerca Universitària and grant SGR2009-0419 and from the Barcelona GSE Research Network.

Contents

1 The purpose and limitations of this survey

Surveys may have many different uses. I hope this survey on strategy-proofness in social choice is able to fulfill some of them.

A possible use is informing the general reader about the type of questions that are addressed by the experts in the area, and the kind of answers that they look for. In trying to meet this objective I have provided a narrative that may be read as an introduction to the topic[1]. It also suggests to the reader some "boxes" where to classify the many papers on the subject. The examples used to illustrate this narrative are treated more extensively that the rest of articles that are also mentioned. This, I hope, allows for this piece to be more than a list of papers. But the choice of illustrations is largely personal, includes some of my own work along with other contributions, and is not meant to signal that the results in question are more important than others that receive a more succinct treatment.

A second use for a survey is, I think, to provide a perspective of what has been done in the past. How did the topic come to be attractive to enough people as to deserve special attention, when and why did this happen, how did this interest evolve. Every generation may need to rediscover some themes that were already treated by earlier ones, and this is surely a guarantee that important issues are not forgotten and get to be analyzed from new angles. But there is also some tendency to forget earlier works, and economics at large suffers, I think, from an inability to accumulate established knowledge. In that respect, I provide an account of how an old remark that voting methods were manipulable was picked up in the early seventies and became part of a very successful move to incorporate the analysis of incentives as an essential part in the study of economic mechanisms. The attention here will be essentially restricted to the issue of strategy-proofness in the context of social choice, but I hope it becomes clear that the subject is part of a larger picture, and an important part indeed. I hope that the survey provides the reader with a good overview of the path followed until the beginning of this century.

A third important use of a survey is to suggest important directions for future work. In that respect, I must admit that the contributions of the last years are treated here in much less detail than those of the preceding

[1]Part of this narrative is borrowed from my previous paper "An Introduction to Strategy-proof Social Choice Functions", (2001a).

decades, and that the survey is not a good substitute for reading the last news on the topic, even if I mention enough recent references to provide the hungry reader with a good start. In fact, there are new developments that only mention slightly. I will go back and refer to them in the final remarks (section 11). I have even left out some essential work on mechanism design that was born and grew in parallel to the contributions in social choice theory that I emphasize here. Again, the reader will find some comments on this literature in the text, and some more at the end. My only confort when admitting these limitations is that any survey gets old, and therefore this last role is more ephemeral than the other two that I tried to serve primarily.

2 Introduction. A few historical notes

Voting rules have been used since ancient times, and the actors involved (candidates, voters and designers) have certainly been aware ever since of the many possibilities that arise to affect their results through strategic behavior, even when the rules are formally respected.

We have evidence of this awareness through the writings of those thinkers who discussed and analyzed voting systems in the past, and I shall mention a few, following the historical materials contained in Black (1948) and in McLean and Urken (1995).

Already in Roman times Pliny the Younger (A.D. 105) discussed beautifully the possibility of what we would call today agenda manipulation: it may be that the sequence in which different proposals are put to vote has an impact on the vote's outcome, thus giving the chairman, or whoever chooses the order of votes a clear occasion to act strategically to his advantage.[2]

The organization of the Church gave rise to many occasions to vote (see, for example, Gaudemet (1979)), and one of the contributions of Ramon Llull (1283) was to propose, in separate texts, the use of those rules that have come to be known as Borda's and Condorcet's methods. While not elaborating on the strategic aspects, Llull's remarks offer evidence of his awareness of the possibility that agents might not behave straightforwardly, and also of the possibility of influencing the behavior of voters through the design of appropriate rules and procedures. When describing the method to nominate candidates to become abbess, Llull (1283) requires from the participants that,

[2] In fact, as noted by McLean and Urken, the case described by Pliny is one where the order of vote does not actually matter.

before voting secretly, they "all....should take an oath to tell the truth". Likewise, in a later writing, where Llull (1299) proposes open, public voting, he demands that "all voters take an oath that they will elect the better and more suitable candidate". Yet, even then he insists that one of the virtues of the method is that if the voters "do not choose the best, it will be obvious to everyone in the chapter that they are choosing the worse candidate and perjuring themselves without any color of an excuse". To add, later on that "those who choose openly are so placed as to be in disgrace with their colleagues if they choose badly. Those who elect in secret are not".

In spite of having required his voters to "strip themselves of all sins" and "swearing an oath in the Lord's altar to elect the person that their free conscience shall duly judge best", Cusanus (1434) had to qualify his enthusiastic defense of the method he proposes. He said that no other method "can be conceived which is more holy, just, honest or free. For, by this procedure, no other outcome is possible, *if the electors act according to conscience* (my *italics*), than the choice of that candidate adjudged best by the collective judgement of all present."

More than three hundred years later, Jean-Charles de Borda (1784) also defended the method that took his name under the assumption that voters would sincerely express their preferences, and when criticized for that assumption just retorted that his election method was only meant to be used by honest men (see McLean and Urken, Chapter 1, footnote 10). Joseph Isidoro Morales (1797), who proposed the same method than Borda just a few years after, discussed much more explicitly the possibility of strategic behavior under different voting rules. "In the methods of election currently in use (...) an area lies open to private or personal injustice by the electors, as, depending on the situation, one, two, three or more of them can prevent the election of the most deserving candidate if they thus wish to contravene the course of justice. This system is so well known and occurs so often that an explanation of it is redundant". He then defended his proposed method (Borda's) along lines that were already expressed in Llull, but at much greater length: "in such an election, merit and justice are safeguarded by censorship of other electors in the case of a public election, and pangs of conscience if it is secret. Even if men's passions cause them to lean toward injustice, their pride will lead them to conceal it."

In spite of Morales' arguments, which run along a different line, Borda's rule is manipulable, in the sense to be used in this survey, which is standard in the contemporary literature. That had already been argued by Daunou

(1803), who pointed at the possibility of voters abusing the method by ranking the most dangerous opponents to their favorite candidate as being the worst candidates.

The most important predecessor of modern social choice theory, M.J.A.N. de Caritat, Marquis de Condorcet, wrote on the matter from two different perspectives. One, associated with the point of view that we now use in developments related to the "Condorcet Jury Theorem" (Condorcet, 1785). That is indeed the perspective that better explains the writings of his predecessors as well. In this perspective, it makes sense to speak about the "true" or "correct" social ordering[3], to refer to malicious deviations from it, and to consider the possibility of costs for those who are discovered to act maliciously (either imposed socially or by their own conscience). In other writings, Condorcet did propose the use of specific rules, and adopted a point of view much closer to that underlying the Arrowian Tradition, where there is no room for such thing as the "true" social order, as a separate entity from individual preferences. Condorcet was aware of the possibilities of manipulation arising from the simultaneous consideration of more than two alternatives, and that his insistence on the use of pairwise comparisons responds to this concern, at least in part. This, coupled with his awareness that these majority comparisons can lead to no conclusion, leads him to accept imperfect but operational methods. Take, for example, Condorcet (1792), where he actually accepts a second best, after arguing that a first best might not be attained, brings us very close to what a modern social choice theorist could end up presenting as an impossibility theorem. In his ideal method, and once all candidates are nominated, "each voter would then express his complete will, by making a comparative judgement between all the candidates taken two by two, and from the majority will on each comparison, we could deduce its general will. However, this method will often give an unsatisfactory result and will not always reveal which candidate the majority prefers, since there may sometimes be no such thing as a majority preference". Now, since we cannot choose the only method which usually reveals the candidates considered most worthy

[3]In the setup of the Condorcet Jury Theorem, there is not only a "true" ranking, but all voters are assumed to share it as a common goal, while having different information and perceptions regarding what the common ranking is. Voting rules must then be viewed as estimators of the common good, given the voters' revealed information. Even in this version of Condorcet's setup there is room for strategic behavior, this time directed to enhance the common interest (Austen-Smith and Banks (1996), Coughland (2000), Feddersen and Pesendorfer (1998), and Rata (2002)).

by a majority (...) we have had to choose the simplest and most practical one, the one that is least susceptible to factions and intrigue (...)".

I will take here a huge leap in time, since it took almost two hundred years after the first golden age of social choice theory[4] to have a general and precise statement confirming the extended suspicion that all rules are subject to some form of strategic manipulation.

The sparse nineteen century authors that worked on electoral systems concentrated on other topics. Likewise, the emphasis of contemporary social choice theory, as initiated by Arrow (1951, second edition 1963) was not on the strategic aspects.

Even then, all major authors in the fifties, sixties and early seventies were aware of the relevance of the issue, and elaborated on it to some extent.

Arrow devotes Section 2 of his introductory Chapter 1 in Social Choice and Individual Values to discuss some limitations of his analysis, and says that "The aspects not discussed may be conveniently described as the game aspects (....). Once a machinery for making social choices is established, individuals will find it profitable, from a rational point of view, to misrepresent their tastes by their actions, either because such misrepresentation is somehow directly profitable or, more usually, because some other individual will be made so much better off by the first individual's misrepresentation that he could compensate the first individual in such a way that both are better off than if everyone really acted in direct accordance with his tastes (...) Even in a case where it is possible to construct a procedure showing how to aggregate individual tastes into a consistent social preference pattern, there still remains the problem of devising rules of the game so that individuals will actually express their true tastes even when they are acting rationally" (Arrow 1963, page 7)[5].

[4] This is how McLean and Hurkens qualify the times of Borda and Condorcet.

[5] Even after excluding the subject explicitly, Arrow referred to it in the text, by discussing the possibility of strategic voting in elections based on pairwise sequential contests. He cites the following example: "Let individual 1 have ordering x, y, z; individual 2, y, x, z; and individual 3, z, y, x. Suppose that the motions come up in the order y, z, x. If all individuals voted according to their orderings, y would be chosen over z and then over x. However, individual 1 could vote for z the first time, insuring its victory; then, in the choice between z and x, x would win if individuals 2 and 3 voted according to their orderings, so that individual 1 would have a definite incentive to misrepresent." The problem treated here is similar to, though not identical with, the majority game, and the complicated analysis needed to arrive at rational solutions there suggests strongly the difficulties of this more general problem of voting. (Arrow, 1951 *a*, pp. 80-81).

Contemporary to Arrow's first writings were those of Black (1948a, 1948b). He explicitly stated that under majority rule, no agent or group could clearly gain from preference misrepresentation when agent's preferences are single-peaked, but he also pointed out that some misrepresentations could lead to cycles, even if this restriction holds. This is because Black and most of the authors in the period were trying to reason about manipulability as a characteristic of binary decision processes, where the chosen outcome is not directly a function of the preference profile, but arises as the maximal element of a social binary relation. Moreover, Black's statement assumes that agents, when called to vote in sequence, can reveal preferences that are not single-peaked[6]. As a result, Black's defense of the strategic properties of majority under single-peakedness is less optimistic than that of later writers. But his contribution is essential. As we shall see, single-peakedness and similar conditions have played an important role in defining domains which admit strategy-proof rules.

Without yet proving any theorem, Vickrey (1960) expressed a neat conjecture about the structure of rules that might be strategy-proof. In a section devoted to "Strategic Misrepresentations of Preferences", he stated that "social welfare functions that satisfy the nonperversity and the independence postulates and are limited to rankings as arguments are (...) immune to strategy. It can be plausibly conjectured that the converse is also true, that is, that if a function is to be immune to strategy and be defined over a comprehensive range of admissible rankings, it must satisfy the independence criterion, although it is not quite so easy to provide a formal proof for this".

Luce and Raiffa's encyclopedic work on Games and Decisions (1957, pp. 359-362) also touched at the issue of strategic voting (Section 14.8), though concentrating only on the manipulation of decisions made by majority rule when alternatives are eliminated sequentially and majority is not transitive, as already discussed by Arrow. Their remarks point out that transitivity may also be necessary for non manipulability, and are thus nicely complementary to those of Vickrey.

The most elaborate analysis of strategic issues in voting at that early period was due to Farquharson (1969). Actually, this monograph was published longer after it was written, in the mid-fifties, and part of Farquharson's ideas were transmitted to the profession through his joint work with Dummett (Dummett and Farquharson (1962)). Much of Farquharson's monograph is

[6] This point was taken up later by Blin and Satterthwaite (1976).

devoted to the analysis of game theoretic equilibria of voting games, using in particular the noncooperative notion of sophisticated equilibrium and a more cooperative notion of collective equilibrium. But before engaging in such analysis, he discusses the possibility of what are now called strategy-proof rules, ones for which expressing the truthful preferences is always a dominant strategy for all players. He uses the term straightforward to refer to such rules. I quote: "The only circumstance in which a voter can make his choice of strategy with absolute confidence are those in which he can be sure that, whatever contingency eventuates, his strategy will give at least as desirable an outcome as any other strategy would have done (...). We adopt the term "straightforward" to describe a strategy which is thus unconditionally best (...). If a procedure affords a voter a straightforward strategy, we may transfer the epithet and say that the procedure itself is straightforward for him".

Farquharson studied, in his Chapter 7, the conditions on individual preferences under which binary procedures, based on the sequential elimination of alternatives by pairwise comparison, may be straightforward. And he concluded that "if there are three or more outcomes, no binary procedure can be straightforward for all possibility scales". This is already a strong and rigorous statement, even if it only covers a particular class of procedures. In a a recent article, Dummet (2005) has described the circumstances of his cooperation with Farquharson and how close they had been on proving the result which was later attained by Gibbard and Satterthwaite.

Murakami (1968, Chapter 4, Section 10) also discussed the issue of stability along similar lines than Dummett and Farquharson. He took one step in the direction conjectured by Vickrey, showing that satisfying a monotonicity property is necessary and sufficient for a social decision function based on pairwise comparisons to be stable.

A fundamental text in the development of social choice theory was Sen's "Collective Choice and Social Welfare" (1970). There again, the author was perfectly aware of the issues, and in fact devoted a section (Section 11.3, pages 192 and on) to discuss the specific question of sincere preference revelation. Sen's discussion is in line with those authors, like Vickrey (1960) and Murakami (1968) who discussed the importance of specific requirements on social choice procedures in order to control the extent of manipulation. Sen was aware of the technical difficulty involved in finding conditions that eliminate all possibilities for manipulation. He pointed out that "non negative response or even positive responsiveness is no guarantee against insincere voting being an efficient strategy" (page 194), and provided an example involving manipulation by groups. But he also qualified the importance of the issue. While admitting that "honest voting" is often not in a person's best interest, he added that: "this is a perfectly general difficulty, but its relevance will vary greatly with the system of collective choice. As Murakami has argued, with those collective choice systems that are non negatively responsive to individual preferences the scope of what voters can achieve by distorting their preferences is very limited".

Another important book from that period was Fishburn's "The Theory of Social Choice" (1973). It also discusses the issue of voting strategy, again in the specific context of sequential voting methods (pages 97-99), but mostly concentrates on other topics.

By this time Allan Gibbard came up with a framework and a result which gave new strength to all preceding remarks on strategic voting, and started a much more systematic study of the issue. We turn now attention to this fundamental contribution.

3 Strategy-proof social choice functions for unrestricted domains: the Gibbard-Satterthwaite theorem

3.1 Statement

It was Alan Gibbard (1973) who first published a precise theorem within a framework that allowed for a sharp and unambiguous statement that all non-trivial social decision functions are manipulable. The same result was proved independently by Mark Satterthwaite (1973) in his doctoral dissertation, and has thus become to be known as the Gibbard-Satterthwaite theorem.

Let us begin by a statement of this celebrated result[7].

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.

[7] Unfortunately, social choice theory has not developed a unified notation, or a unified set of denominations for its basic constructs. Terms like voting schemes, social decision functions, social choice functions, and the like are often used to name the same construct, and each one may be used to denote a different one in some other article. I will not try to use the words of each of the authors, but try to be somewhat consistent within the survey.

Preferences will be always complete, reflexive, transitive binary relations on A. We denote individual preferences by $\succcurlyeq$, $\succcurlyeq'$, $\succcurlyeq_i$, etc. The corresponding symbols $\succ$, $\succ'$, $\succ_i$will stand for the strict part of the relation. $\mathcal{R}$ 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 $D_1 \times \ldots \times D_n \subset \mathcal{R}^n$ is a function $f : D_1 \times \ldots \times D_n \to A$, 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* $f : D_1 \times \ldots \times D_n \to A$ *is manipulable* iff there exists some preference profile $(\succcurlyeq_1, \ldots, \succcurlyeq_n) \in D_1 \times \ldots \times D_n$, an agent i and some preference $\succcurlyeq_i \in D_i$, 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* on $D_1 \times \ldots \times D_n$ iff it is not manipulable.

Different reasons why this is an interesting property of social choice mechanisms will be discussed along the text. Let us first introduce a few comments here. As already noted by Gibbard,"... to call a voting scheme manipulable is not to say that, given the actual circumstances, someone is really in a position to manipulate it. It is merely to say that, given some possible circumstances, someone could manipulate". From the point of view of voters, it may pay to learn about others under a manipulable social choice function, but not under a strategy-proof one. As for the consequences of manipulation, if they occur, there may be many, but the possible loss of efficiency is particularly worrisome from the point of view of the designer. Social choice functions which would always select an efficient outcome if voters provide truthful information may end up recommending an inefficient alternative, after voters distort their preferences in order to manipulate.

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 in its domain,

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

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. Much of the work surveyed here consists in examining how this conclusion may change when the domain of preferences is restricted in different ways.

Let me briefly comment on the definition of manipulability and on the implications of the Gibbard-Satterthwaite theorem. Since I elaborate on the topic at different points, these initial remarks are just introductory, and a warning about too simplistic a reading of the result.

First of all, notice that the definition of manipulability assumes that one agent could gain from misrepresenting his preferences if (1) he knew for sure how others would vote, and (2) that their vote would not change as a result of the misrepresentation. Moreover, the definition predicates that such an opportunity for gain would arise at some profile. Therefore, there are many qualifications to make about the actual implications of manipulability. Certainly, the fact that a social choice function is manipulable does in no way imply that agents will engage extensively in changing their declared preferences. This is because they cannot always gain. Also because they may not know whether they can at any given situation, either for ignorance about the other agent's preferences, actions or strategies. And, as different authors have pointed out, also because the potential beneficiary of a manipulation may have difficulties in determining/computing the implications of his actions under complex rules. Yet, as soon as an agent understands that a social choice function is manipulable, he has an interest in knowing whether the conditions to take advantage from it do hold. It may be in his benefit to know about the other agent's preferences, intentions, strategies, and about the intricacies of the rule. By contrast, a rational agent who knew to be operating under a strategy-proof rule would know that there is no point in wasting resources in such quests, since the added knowledge would not yield any potential gain. Agents under strategy-proof rules will derive no advantage from acting strategically, that is, from conditioning their behavior on that of others. By contrast, the possibility of manipulation is an invitation to attach value to all the considerations above, even in those cases where, in the final analysis, agents could decide not to misrepresent their preferences for one reason or other.

3.2 The impact of the Gibbard-Satterthwaite theorem

In fact, this statement, which is standard today, did not appear in that exact form in the initial work of any of the two authors after whom it is named. Gibbard's theorem was stated as a corollary of a more general statement, regarding game forms. We shall elaborate further on that larger framework. Satterthwaite's original work used a different but equivalent notion of manipulability (Satterthwaite, footnote 5). A statement and proof of the result in the terms that became more widely know was due to Schmeidler and Sonnenschein (1976), who also used the term Gibbard-Satterthwaite theorem for the first time.

Gibbard's proof was based on connecting the structure of strategy-proof social choice functions and that of Arrowian social welfare functions. We shall elaborate on this connection later on, but point it out here to stress how ripe the situation was for the theorem to appear, after the different conjectures and results that had been stated on such connection by authors like Vickrey, Sen and Murakami, along with the negative partial statements already proved by Farquharson. Gibbard (1973) remarks that "The theorems in that paper should come as no surprise (...)Since many voting schemes in common use are known to be subject to manipulation, writers on the subject have conjectured, in effect, that all voting schemes are manipulable (...) A result such as the one given here, then, was to be expected. It does not, however, turn out to be easy to prove from known results".

In fact, two important articles on the subject also appeared in 1973, both written independently from Gibbard's article and Satterthwaite's thesis. One is due to Pattanaik (1973), entitled "On the Stability of Sincere Voting Situations". This work, and subsequent articles on the subject, qualify Pattanaik's contributions among the most salient in the study of manipulation, as he essentially proved the impossibility of non manipulable rules within the Arrowian framework. Yet, by concentrating on group decision rules, rather than the simplest framework of social choice functions, Pattanaik had to face different technical problems, including the need to break ties in case of social indifferences, which made his results harder to communicate to larger audiences. Another important paper was Zeckhauser's "Voting Systems, Honest Preferences and Pareto Optimality" (1973). That paper considers a different framework than Gibbard and Satterthwaite's, by allowing lotteries as outcomes and letting individuals state their preferences over lotteries. We shall refer to it later, along with other results in this framework. What I just want to stress here is that the paper elaborates on the connections between voting and economic systems, and explicitly discusses the trade-off between strategy-proofness and efficiency, two conditions that we shall see to be very hard to encompass by nontrivial decision rules[8].

This insistence that the times were ripe for such a result is not to minimize the importance of the Gibbard-Satterthwaite theorem. Rather, it is to wonder about how important it is for a good result to appear in the right form, at the right moment.

The moment was right, not only because of the internal developments in social choice theory that I have already outlined, but also because of the larger trends in economics. The popularization of the Gibbard-Satterthwaite theorem coincided with other important discoveries in public economics: the

[8]For further comments, see section 6.5.3.

Clarke-Groves mechanisms (Clarke (1971), Groves (1973), Groves and Loeb (1975)), Green and Laffont (1979), which provided a sharp solution to the free rider problem as stated by Samuelson (1954), and Hurwicz's (1973, 1977) analysis of the incentives dimension within his general framework for the study of mechanism design. It also coincided with a renewal in the ability of game theory to provide sharp analytical tools to analyze the strategic behavior of agents. All of a sudden, then, the incentives for agents to behave according to the set rules of the economic and political game become a matter of priority in the research agenda of economists and political scientists. A clear picture of this state of affairs is provided by J. J. Laffont's editorial introduction to the book "Aggregation and Revelation of Preferences"(1979).

The form of the theorem was also crucial for its instant popularity. It is a sharp statement, formulated within a simple but general framework, and referring to a single type of strategic behavior, leaving aside other complications. Previous authors had been getting close to state that all social decision rules were manipulable, and to connect the condition of strategy-proofness with some of Arrow's conditions. But many of these statements were marred by the fact that authors had in mind a framework where social aggregation resulted in a social binary relation, which might in turn lead to social cycles (with ill defined choices), or to social indifference (with more than one choice). The simpler framework used by Gibbard and Satterthwaite allowed to focus on essentials[9].

Gibbard had to express the difference with great force. "Neither voting schemes nor game forms allow ties. Both take single outcomes as values, and for a good reason. In questions of manipulability, the final outcome is what matters (...). In this respect, a voting scheme differs from an Arrow "constitution", which it resembles in all other aspects".

By concentrating on social decision functions, that initial result gained in transparency and conveyed its message with maximal effectivity. Likewise, considering manipulability only, and leaving aside other important forms of strategic behavior, contributed to sharpen the results. And stating the theorem for functions defined on a universal domain was also essential to get its strong negative conclusion.

Much of the literature that followed can be seen as a sequence of quali-

[9] In fact, this framework had already been proposed by Farquharson, who then did not spend too much time on the issue of straightfordwarness and moved to the analysis of more game-theoretical questions.

fications regarding Gibbard and Satterthwaite's choice of framework. These results will sometimes show the robustness of the theorem, sometimes show that it does not hold under certain alternative frameworks. But there is no doubt that this theorem marks the start of several important lines of research. I will point at several of these directions in the next section, and will concentrate mostly on those more closely related to social choice theory in the rest of the paper.

Before anything else, we should look into the different proofs of that seminal result.

3.3 Proofs of the 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.

3.3.1 Proofs based on the connection with Arrowian social welfare functions.

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 referred to a wide framework, involving game forms. We concentrate here in its adaptation to social choice functions. In its simple form, the argument we provide, which captures Gibbard's essential insight, was popularized by Schmeidler and Sonnenschein (1974, 1978), and it runs 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.

A similar proof is due to Gärdenfors (1977, also see 1976). Instead of using Arrow's theorem, he resorted to an analogous result of social choice functions for multiple agendas due to Hanson (1969).

Batteau, Blin and Monjardet (1981) provided a proof that stresses the connection between strategy-proof rules and Arrowian social welfare functions, by showing that the distribution of power underlying both types of rules must have the same structure. Specifically, the family of "preventing sets" underlying a strategy-proof rule must be an ultrafilter, which is the structure of "decisive sets" in Arrowian functions.

3.3.2 Proofs by inspection and further induction

A second interesting approach to prove the Gibbard-Satterthwaite theorem is based on a close examination of strategy-proof social choice rules for three alternatives and two persons with linear preferences, followed by an extension to weak orders and by a double induction on the number of agents and alternatives. Induction had already been used by Satterthwaite (1973) in his dissertation. Schmeidler and Sonnenschein (1978) provided a simple and elegant proof along these lines. 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 (see Sen (2001)). This proof emphasizes that the two person, three alternative case contains all the essential elements of the theorem, in a nutshell. Another proof along similar lines, but limited to the two person, three alternative case, is provided by Feldman (1980).

The use of induction on the number of alternatives is limitative, as it does not allow to extend the result to cases with an infinity of alternatives. On the other hand, the use of induction on the number of agents is quite suggestive of a fact that arises in many contexts: solving the problem of strategy-proofness for two-individual societies is a long step toward solving it for any society with a finite number of agents[10].

3.3.3 Proofs based on the necessity of strong monotonicity

Monotonicity properties can be predicated from social choice functions, social choice correspondences, social welfare functions or other models of social choice. They all try to reflect the idea that, if an alternative is chosen at some profile, then it must also be chosen at other profiles where that alternative has improved its position. Because the notion of "improving the position" is subject to different qualifications, there are several versions of monotonicity. A strong form of it turns out to be necessary for the strategy-proofness of social choice function. Moreover, social choice functions whose range has more than two alternatives can only satisfy strong monotonicity it if they are dictatorial. This is the line of argument developed by Muller and Satterthwaite (1977, 1985). See also Peleg (1984, page 33) and Moulin (1988, Section 9.1).

3.3.4 Proofs that emphasize the role of pivots

In two separate papers, Barberà (1980a, 1983a) provided new proofs of Arrow's and Gibbard and Satterthewaite's theorems which focused on the role of pivotal voters in collective decision-making. Essentially, an agent is a pivot at a preference profile if she can change the social outcome just by changing her preferences. The proofs consist, both for Arrow's and for Gibbard-Satterthwaite's, in first proving that only one agent can be pivotal at each preference profile (otherwise, a contradiction to Arrow's conditions would arise, or a manipulation would be possible). Then, it proceeds to show that the agent who is eventually pivotal at some profiles is always the same, and always pivotal: the dictator. By a very different type of reasoning that of Satterthwaite (1975), these early papers also pointed at the strong connections

[10] This is not the case, for example, for different forms of implementability, where the two-person case needs to be handled separately than that with more people.

between both theorems, in the context of unrestricted domains[11].

3.3.5 Proofs that build on the structure of strategy-proof rules and option sets

I would like to sketch a proof that was presented in Barberà and Peleg (1990), and has its roots in Barberà (1983a). While the preceding proofs only apply when the number of alternatives is finite, the proof I am about to present, although it is sketched here for the finite set of alternatives, two voter case, can be adapted to cover the case with a continuum of alternatives[12]. 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. As Sprumont (1995, pp. 98-99) pointed out, this "proof technique has been successfully applied to other domains of preferences over public alternatives... Moreover, the bulk of the recent literature on strategy-proofness in private commodity environments also follows (this) approach". As we shall see, the crucial notion in this approach is that of an option set. The role of this concept had been noticed by Laffond (1980), and by Chichilnisky and Heal (1981).

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 $\mathcal{P}$ (here again, the extensions to general preferences and to n agents are quite straightforward). The argument runs as follows.

- Let $f : \mathcal{P} \times \mathcal{P} \to A$ 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

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

[11] When domains are restricted, Arrow's conditions need not lead to the same conflicts that Gibbard-Satterthwaite's, and vice-versa. See Barberà (1996).

[12] The reader will see that the version I provide for a finite set of alternatives involves the use of concepts, like that of the second best alternative, that would not be well defined for the continuum , and need adaptation. For a careful analysis about the implications on the range and on the shape of option sets in the continuum case, see Le Breton and Weymark (1999).

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 (1990), Section 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).

To end this section, let me insist that strategy-proofness has implications on the structure of option sets. In the case where preferences are defined on a continuum, it also has implications on the structure of the range. On that, see Le Breton and Weymark (1999).

3.3.6 The Gibbard-Satterthwaite theorem as a corollary

In spite of its generality, the Gibbard-Satterthwaite theorem does appear as a corollary of even more general results, in different frameworks.

We shall see later that one can embed any finite set of alternatives into a multidimensional euclidean space. General results describing the structure of strategy-proof rules under different range structures and preference domains induce the Gibbard-Satterthwaite theorem as a special case (Barberà, Massó and Neme (1997), Barberà, Massó and Serizawa (1998), Nehring and Puppe (2002)).

An important model in mechanism design is the one where compensations via money transfers are allowed. This framework, which was introduced in the research on strategy-proof allocation initiated by Clarke (1971) and Groves (1973), does also provide a way toward the Gibbard-Satterthwaite theorem. This is based on the remark that no non-imposed decisive rule be implemented in dominant strategies without the use of transfers, which at some point violate the global feasibility constraints. This argument was made in an early and important paper of Roberts (1979), which is a starting point for many recent developments in the literature on rules involving compensation.

One of the intriguing aspects in the Gibbard-Satterthwaite result is that, even before it was formally proven, some authors did establish its close connection with Arrow's impossibility theorem (Vickrey (1960)). Satterthwaite (1975), and Kalai and Muller (1977) went a long way in establishing this connection (I elaborate further on this point in Section 10.1).

Recently, different authors (Geanakoplos (2001), Reny (2001)), insisted in the strong parallels between some of the proofs of these two results. Two papers by Eliaz (2004) and Barberà (2001b) have provided results which surpass this approach, by proving both results to be specific instances of larger, more abstract theorems on preference aggregation.

4 Game forms and the question of implementation

4.1 Strategy-proofness in a larger picture

As I already pointed out, Gibbard's (1973) statement and proof of his results about manipulation do not only refer to social choice functions (which he called voting schemes), but to a more general framework. In his own words:

"A voting scheme is a special case of what I will call a game form. A game form (...) is a system which allows each individual his choice among a set of strategies, and makes an outcome depend, on a determinate way, on the strategy each individual chooses. A "strategy" here is the same as a pure strategy in game theory, and indeed a game form is a game with no individual utilities yet attached to the possible outcomes. Formally, then, a game form is a function g with a domain of the following sort. To each player 1 to n is assigned a nonempty set, $S_1, ...S_n$ respectively of strategies. It does not matter, for purposes of the definition, what a strategy is. The domain of the function g consists of all n-tuples $< s_1, ..., s_n >$, where $s_1 \in S_1, s_2 \in S_2, ..., s_n \in S_n$. The values of the function g are called outcomes. A voting scheme, it follows, is a game form such that, for each player, his set of strategies is the set of all orderings of a set Z of available alternatives, where Z includes the set X of outcomes. (...) For game forms alone (...) there is no such thing as manipulation. To manipulate a system, a voter must misrepresent his preferences. To talk about manipulation, then, we must specify not only a game form, but for each voter and preference ordering P we must specify the strategy which "honestly represents" P. Manipulability, then, is a property of a game form $g(s_1, ..., s_n)$ plus n functions $\sigma_1, ..., \sigma_n$ where for each individual k and preference ordering P, $\sigma k(P)$ is the strategy for k which honestly represents P.(...)What we can show is this: however we characterize honest voting in a system, the system, as characterized, will be manipulable (...) Here is the result (...) A strategy $s*$ is dominant for player k and preference ordering P of the set of outcomes if, for each fixed assignment of strategies to players other than k, strategy $s*$ for k produces an outcome at least as high in preference ordering P of the set of outcomes if, for each fixed assignment of strategies to players other than k, strategy s^* for k produces an outcome at least as high in preference ordering P as does any other strategy open to k.(...) A game form is straightforward if for every player k and preference ordering P of the outcomes, some strategy is dominant for k and P. The theorem on game forms says that no non-trivial game form is straightforward" (Gibbard, pages 588 to 591).

I quoted extensively from Gibbard to show that, indeed, his formulation was very broad, and opened the way to consider, with very small changes of framework, general issues of incentive compatibility and implementation.

For comparison and perspective, I will provide a framework for implementation theory, as described by John Moore (1992)[13] in a brilliant survey, written some twenty years later.

"Consider an environment with a finite set $\{1, ..., i, ..., I\}$ of *agents*, and a set A of feasible *outcomes*, with typical element a.

The profile of the agents' preferences over outcomes is indexed by the *state* θ, agent i has preference ordering $R_i(\theta)$ on the set A. Let $P_i(\theta)$ and $I_i(\theta)$ respectively denote the strict preference relation and the indifference relation corresponding to $R_i(\theta)$.

Each of the agents is assumed to observe the state θ, so there is *complete information* among the agents about their preferences over A.

The above formulation allows for any degree of correlation across the agents' preferences. Θ may, for example, comprise all possible vectors of preference orderings over A: the *universal* domain. Or there may be perfect correlation, in which case knowing one agent's preference ordering over A would be enough to deduce all the other agents'. We shall say that preferences in Θ have *independent domains* if agent i's set of possible preference orderings over A is fixed -independent of how the other agents $j \neq i$ happen to rank A.

A choice rule is a correspondence $f : \Theta \twoheadrightarrow A$ that specifies a non-empty choice set $f(\theta) \subseteq A$ for each state θ.

The implementation problem is as follows: does there exist a *mechanism,* or game form, g such that in any state θ, the set of equilibrium outcomes of g coincides with $f(\theta)$? If so, then g (fully) *implements* f. This is a general notion of implementation, in that we have left open the choice of equilibrium

[13] Note: I excerpt literally from Moore's exposition (pages 214-217), which has become standard.

concept.

A natural place to start is with a *revelation mechanism*, g^*, in which each agent i's strategy set comprises his set of possible preference orderings, $\{R_i(\theta) \mid \theta \in \Theta\}$: that is, each agent simply announces what are his preferences over A. (Arguably, revelation mechanisms make most sense if preferences in Θ have independent domains, because then any vector of preferences reported by the agents could in principle be the truth.) If, in each state θ truth-telling is an equilibrium, whose outcome is in $f(\theta)$, then g^* truthfully implements f. Notice that this is weaker than (full) implementation, because there may be other, untruthful equilibria in state θ whose outcomes are not in $f(\theta)$.

The most appealing notion of implementation is the one that makes the weakest assumptions about the agents' behavior: *implementation in dominant strategy* equilibrium.

To discover what can be implemented in dominant strategy equilibrium (and other equilibrium concepts), a useful ground-clearing result comes from the *Revelation Principle*. This provides a set of necessary conditions —in effect, incentive constraints— which a choice rule must satisfy if it is to be (fully) implementable. In particular, consider the case where preferences in Θ have independent domains, and where the choice rule f is *single-valued* (i.e., where $f(\theta)$ is a single outcome for all θ). Then if f is (fully) implementable in dominant strategy equilibrium, it must also be truthfully implementable in dominant strategy equilibrium. To see why, replace the non-revelation mechanism g which (fully) implements f in dominant strategies by a revelation mechanism g^* which mimics it. That is, if, in state θ, the I agents choose the vector of (dominant) strategies $(s_1(\theta))$, say, in g, then announcing the truth in g^* leads to the same outcome: $g^*[R_1(\theta), .., R_I(\theta)] \equiv g[s_1(\theta), ..., s_I(\theta)]$. Clearly, for each agent i, announcing the truth $R_i(\theta)$ in g^* must be a dominant strategy, because $s_i(\theta)$ is a dominant strategy in g —hence g^* truthfully implements f in dominant strategy equilibrium, as required. However, in moving from g to g^*, we may admit new, unwanted, untruthful dominant strategy equilibria: g^* need not (fully) implement f.

Nevertheless, one sense that for a rich enough choice rule and associated mechanism g^*, there is unlikely to be a multiplicity of dominant strategy equilibria. In particular, there will only be a gap between (full) implementation and truthful implementation in dominant strategy equilibrium if there are indifferences in the agents' preference orderings. If Θ only *contains* strict preference orderings, then dominant strategies are essentially unique:

it can easily be shown that a choice rule f is (fully) implementable in dominant strategy equilibrium if and only if it is single-valued and truthfully implementable in dominant strategies. Laffont and Maskin (1982, pp. 42-43) present other conditions guaranteeing that if truth-telling is a dominant strategy equilibrium of the revelation mechanism g^*, then it is the only one.

Unfortunately, for dominant strategy implementation, the necessary conditions (the incentive constraints) provided by the Revelation Principle are very demanding. For a single-valued choice rule f to be truthfully implementable in dominant strategy equilibrium, it must be *strategy-proof*: for any agent i, if $\theta, \phi \in \Theta$ are such that $R_j(\theta) = R_j(\phi)$ for all $j \neq i$, then $f(\theta) R_i(\theta) f(\phi)$ and, symmetrically, $f(\phi) R_i(\phi) f(\theta)$.

This follows straight from the definition: if f is truthfully implementable, then in state θ agent i cannot gain from misreporting his preferences as $R_i(\phi)$, thereby changing the outcome from $f(\theta)$ to $f(\phi)$. Moreover, it is clear that, if preferences in Θ have independent domain, then strategy-proofness is also sufficient for f to be truthfully implementable in dominant strategy equilibrium: simply use the revelation mechanism $g^*[R_1(\theta), .., R_I(\theta)] \equiv f(\theta)$.

Under this general framework, the Gibbard-Satterthwaite theorem takes the following form:

Theorem 2 *(Gibbard, 1973; Satterthwaite, 1975). Suppose Θ includes all possible strict preference orderings over A. Then no single-valued choice rule f, whose range contains at least three distinct outcomes, can be truthfully implemented in dominant strategy equilibrium unless it is dictatorial."*[14]

It should be apparent from this formulation that the question of strategy-proofness is a limited one. The question of implementation can be asked with reference to many other equilibrium concepts, and that is what implementation theory has done, while also allowing for multivalued choice functions and restricted domains. There exist many expository articles, and a vast literature on the issue. Three good references are Moore (1992), Jackson (2001) and Maskin and Sjostrom (2002).

4.2 Strengthenings, weakenings and related definitions

Although I will not review results in implementation theory that result from substituting the requirement of dominant strategies for other standard game

[14] This is the end of my extensive quote from Moore.

theoretic equilibrium concepts, I need to mention some variants of the notion of manipulation, and some definitions close to that of manipulation which have appeared in the literature, in close connection to the basic ideas we analyze here.

As we have already seen, Farquharson (1969), and Dummett and Farquharson (1961) did analyze the concept of stability. A modified version of stability was used by Murakami (1968), criticized by Sen (1970) and by Pattanaik (1973, 1974). The latter author provided a new definition, which applies to group decision rules, resulting in a social preference relation for every preference profile, and allowing for the (weak) social preference to select more than one alternative. Besides the fact that stability is defined for a different class of objects than social choice functions, its main difference with strategy-proofness is that it contemplates the possibility of joint deviations by groups of agents. Similar possibilities of cooperative manipulation can be defined for social choice functions, and rules avoiding them are termed group strategy-proof.

In those contexts where only trivial strategy-proof social choice functions exist, there is little point insisting on a stronger requirements. However, there are contexts where the strategy-proofness of attractive rules can be attained, and then group strategy-proofness becomes an additional standard.

Let us provide some formal definitions, in the vein of those we used when presenting the notion of (individual) strategy-proofness.

A social choice function $f : D_1 \times \ldots \times D_n \rightarrow A$ is *group manipulable* iff there exists some preference profile $(\succcurlyeq_1, \ldots, \succcurlyeq_n) \in D_1 \times \ldots \times D_n$, a group of agents $S \subset I$ and preferences $\succcurlyeq_i' \in D_i$ for all agents $i \in S$, such that

$$f\left(\succcurlyeq_{I\setminus S}, \succcurlyeq_S'\right) \succ_i f\left(\succcurlyeq_1, \ldots, \succcurlyeq_i, \ldots, \succcurlyeq_n\right)$$

for all $i \in S$[15].

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

Clearly, this definition allows for many qualifications. A social choice function may be strategy-proof but manipulable by groups, and in this case the size of the groups can be relevant. Intuitively, rules that are manipulable by large groups only can be seen as more robust than other that can be manipulated by small groups. The extreme case where two agents alone could

[15] We use the notation $f\left(\succcurlyeq_{I\setminus S}, \succcurlyeq_S'\right)$ to denote the profile where all agents in $I\setminus S$ retain the preferences in the original profile $(\succcurlyeq_1, \ldots, \succcurlyeq_n)$ and those in S change preferences to those $\succcurlyeq_i'$ specified in the definition.

manipulate rules that are (individually) strategy-proof occurs in several instances (Barberà (1979), Barberà, Sonnenschein and Zhou (1991)). Serizawa (2006) has studied the issue specifically, and has explored the consequences of impossing the requirement of pairwise strategy-proofness, which explicitly avoids this extreme form of group manipulability. The requirement of group strategy-proofness, even if very strong, turns out to be compatible with other interesting properties in different domains, motivated by a variety of economic interpretations. Although we shall mention them later, let us announce some: house allocation (Ehlers(2002), Takamiya (2001)), allocation of other indivisible goods (Ehlers and Klaus (2003)), excludable public goods (Olszewski (2004)), matching models (Martínez, Massó, Neme and Oviedo (2004)). A similar requirement is that of bribe-proofness, introduced by Schummer (2000), where again two agents can manipulate through mutually beneficial strategies. This same author, as well as Shenker (1993) discussed different variants which represent additional demands above that of strategy-proofness.

Another interesting fact, that was first remarked by Blair and Muller (1983), is that in different domains the (non-trivial) satisfaction of individual strategy-proofness precipitates that of group strategy-proofness as well. This is the case, for example, when preferences are single-peaked (see Moulin's (1980) characterization in Section 5 below), or single-dipped (Pereman and Storcken (1999)), among other cases. I have already mentioned that this connection does not always hold. Le Breton and Zaporezets (2009), and Barberà, Berga and Moreno (2009) establish conditions on the domains that actually do guarantee that rules satisfying the weaker (individual) version on them will also meet the stronger (group) requirement.

Pattanaik (1976a, 1976b) also considered interesting weakenings of the notion of strategy-proofness, both in the Arrowian context of group decision rules and in that of social choice functions. These weakenings were based on the possibility that some threats of manipulation (either by a single voter or by a coalition) might be diffused by the existence of counterthreats. Then, one might only be concerned with (relevant) threats which are not met by (adequate) counterthreats[16]. This idea is very natural, and has its roots in cooperative game theory, where it is incorporated to different solution

[16] Notice that this vague expression leaves room for many different definitions in a similar spirit, depending of what threats are defined to be relevant, and what counterthreats are considered adequate to difuse them.

concepts, like that of the bargaining set (be it in Maschler', Mas-Colell' and Zhou's version). Unfortunately, Pattanaik's results did prove that weakenings of stability or strategy-proofness along these lines do not significantly improve upon the negative result of the Gibbard-Satterthwaite theorem, for most cases (see also Barberà (1980b)). Maximin behavior of agents has also been analyzed, in different versions, within the context of strategic behavior. See, for example, Moulin (1981), Thomson (1979), Barberà and Dutta (1982)

The general theory of implementation is interested in achieving desirable outcomes (as expressed by a given social choice function, or by a correspondence), through the agent's interplay within a mechanism (or game form). A more specific formulation would not allow for any game form to implement a given function, but rather examine the equilibria of the game that is implicitly defined by some given voting rule, and see whether, in spite of violating strategy-proofness, the rule may be expected to give rise to interesting outcomes. This led several authors to analyze different issues regarding the equilibria associated to certain types of voting methods. Following Farquharson (1969), Moulin (1980b, 1981b, 1983) developed an elaborate theory of sophisticated voting, showing that some families of voting rules based on sequential voting would lead to attractive outcomes. Dutta and Pattanaik (1978), see also Dutta (1980), developed an idea of consistency, which was later followed up by Moulin and Peleg (1982). See also Peleg (2002).

4.3 Other forms of strategic behavior in voting

The notion of manipulation is quite general. Under appropriate interpretations, it formalizes many form of strategic behavior that will not literally consist in misrepresenting preferences. Yet, voting methods are subject to other possible types of strategic behavior, which exceed the limited framework within which the notion of strategy-proofness is formulated. We should just mention some of these aspects, as a sample. Blin and Satterthwaite (1977) pointed out the possibility of an agent manipulating a voting procedure by inducing false beliefs on other agents about his/her true preferences. Dutta, Jackson and Le Breton (2001, 2002) have analyzed the influence of strategic candidacy over the outcomes of single-valued voting procedures. Their work has been extended by Carmelo Rodríguez-Álvarez (2004, 2006) to the case of correspondences and probabilistic rules.

One important strategic question arises in connection with the choice of rules that society will use. Different authors have studied the possibility of guaranteeing the stability in the choice of rules. Koray (2000) proved its impossibility under a very demanding definition. His proof builds on the notion of strategy-proofness in an interesting way. Barberà and Jackson (2006), Messner and Polborn (2004) and others have also investigated the issue for restricted situations and a less demanding definition of stability.

Berga, Bergantiños, Massó and Neme (2006) have studied the strategies of entry and exit to and from a voting body whose decisions affect the voters' satisfaction. Cantalà (2004) considers the case of voluntary consumption of public strategic goods, which also induces additional strategic considerations. This is an incomplete list of issues which I mention because they are close to the standard social choice literature, but there are many others, and with a long tradition. Luce and Raiffa's (1957) wonderful book already listed a number.

Other subjects are so important in political economy and political science, that I will just mention them without references. A very important issue in practical terms is that of agenda manipulation. Another is logrolling. A third one is participation and abstention. A fourth is the issue of candidacy: who participates in an election is an extremely relevant strategic decision. So is the issue of platforms, and whether agents want to look like others, or rather differentiate. Unfortunately, these issues exceed the present essay.

4.4 The analysis of specific social choice rules

Since no voting rule is strategy-proof in the universal domain, authors interested in the strategic performance of specific voting rules must look into their somewhat more limited properties. Their analysis often suggests properties that may be specific to the rule in question, or sometimes extend to others. Brams and Fishburn (2002) contains a complete account of voting procedures analyzed from different points of view. As an example, consider the study of approval voting, a method proposed by Brams and Fishburn (1978). Approval voting is part of a more general class of procedures, called nonranked voting systems. Fishburn (1978) did also study issues of preference revelation in that context. While obviously not strategy-proof in large enough domains, the method satisfies some weaker properties of interest (Brams and Fishburn (1993)). Moreover, under appropriate domain restrictions, it can not only attain strategy-proofness but in fact be the only symmetric neutral and efficient rules to do so (Vorsatz (2007). Another example is given by plurality rule. Early work by Pazner and Wesley (1977, 1978) emphasized that this and other rules, which are obviously manipulable, become less so when the number of voters grows large. Notions of "asymptotic" strategy-proofness, or others related to the size of society have emerged in other contexts (see section on Strategy-Proof exchange). More recent work by Slinko (2002) addresses the asymptotic manipulability of other rules. Another approach to evaluate rules which fail to satisfy strategy-proofness consists in defining some degree of approximate satisfaction of the property. An example is provided by Schummer's (2004) notion of Almost-Dominant strategy implementation. Another example is given by the notion of threshold strategy-proofness, which bounds the gains from manipulation (Ehlers, Peters and Storcken (2004)). Other authors are more interested in comparing different rules among themselves, rather than highlighting any specific one. Then, again, it is useful for them to have a standard different than manipulability, for which all nontrivial rules would fail to meet. One of the concepts that have been used is that of susceptibility to manipulation. In fact, the idea that there one can measure the degree of manipulability of different rules was developed by Kelly (1993) and further studied by Aleskerov and Kurbanov (1999). These works concentrate on the relative size of the subdomains where the rules do satisfy the exact version of strategy-proofness. Another way to measure departures is in Campbell and Kelly (2002, 2003a, 2003b). One can also investigate the minimal size of profiles where manipulation possibilities will arise: Mausa, Peters and Storcken (2007) have established lower bounds on this size.

Work by Campbell and Kelly (2003a) and Merrill (2007) has also explored the limits of the intuition that choosing Condorcet winners would be a strategy-proof rule if it was always well defined. Saari (1990, 2001) has argued that the Borda count is the one member of the general class of scoring rules which has the best relative performance in terms of incentives. The Borda count can be strategy-proof for adequately restricted domains, as shown in Vorsatz (2007) and more generally in Puppe and Tasnádi (2008) and Barbie, Puppe and Tasnádi (2006). Another measure related to manipulation is the degree of complexity required in order to compute a manipulative misrepresentation, which varies from rule to rule. Bartholdi et al (1989) started a literature that has now, years later, become extremely popular among computer scientists. I will not attempt to survey this recent and important literature here.

5 The search for strategy-proofness: an outline

In spite of the limited scope of the issue as compared with the broader question of implementation, this survey concentrates on strategy-proofness.

We shall see that there is much to say on the subject, beyond the most abstract results, when one concentrates on specific concepts, models and interpretations.

It is important to ask under what circumstances it would be possible to design non-trivial strategy-proof decision rules, because strategy-proofness, when attainable, is an extremely robust and attractive property.

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. The other is to assume that there more than two alternatives to choose from, and that at least three of them will eventually be chosen for some admissible state of opinion.

In many cases, 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.

Much of the research on strategy-proofness can be seen as an investigation on the structure of alternatives, the existence of domain restrictions on such alternatives which allow for nontrivial strategy-proof social choice functions, and the characterization of such functions, when possible. There are many domain restrictions that had been studied for purposes other than the analysis of strategy-proofness, especially in the context of public decision making, and that also turn out to be interesting for our purpose. For example, domains where preferences are single-peaked, single-plateau, single-dipped, single-crossing, or others where individual preferences are expected to exhibit indifferences between certain alternatives. Let me mention some other types of natural restrictions: that of representability in terms of Von Neumann-Morgenstern utility functions become natural when alternatives are lotteries; or strict convexity that may be appropriate as a preference restriction when dealing with exchange economies, etc. In these and many other cases the restriction of preferences to such classical domains, or to others arising from different models and their interpretations may allow for the existence of nontrivial strategy-proof social choice functions.

Much of what we have learnt about strategy-proofness responds to the following pattern.

First, consider some wide class of social decision problems, and formulate a model that formalizes them. We survey work that addresses, among others, the following issues: how to choose the levels of provision of one or several public goods, elect candidates or locations characterized by a variety of characteristics, how to ration the usage of different production factors, how to match students to colleges, or how to allocate private goods through markets.

Second, discuss under what domain restrictions, defined within each specific model, would it be possible to design non-trivial strategy-proof social decision rules. Notice that there are several levels at which such a discussion may be set and resolved. Impossibility results may be attained, even in specific contexts, if meaningful and stringent enough domain restrictions cannot be found. Partial possibility results may be attained, describing rules that are strategy-proof for specific domains. But the ideal result would be one that jointly characterizes domains that admit strategy-proof rules, and the family of rules that have this property for such domains. When such knowledge can be attained, it is very informative, and it allows for further inquiries. Take, for example, the question of efficiency. If we know that strategy-proof rules exist under some restrictions, and how they look like, then one may ask whether some of these rules can guarantee efficiency. Even if none can achieve it, one can study how far from efficiency they are, and thus measure the efficiency costs of strategy-proofness under specific circumstances. Eventually, such costs should be compared with those arising from incorrect preference revelation under alternative rules not satisfying strategy-proofness.

In addition to strategy-proofness, there are several other requirements of interest that one may impose on social choice rules. One is coverage: rules are interesting only to the extent that they operate on rich enough preference domains. Another is nondictatorship: although within the scope of conceivable mechanisms, dictatorial procedures are formally trivial and normatively unattractive. A third requirement is efficiency, which I have already mentioned. It is worth referring to two related conditions that may help understand why strategy-proof rules may fail to be efficient. One is the extent of their range. Social choice functions may or may not have the whole set of alternatives as their images. Some functions may become strategy-proof at the expense of never choosing some a priori feasible alternatives, regardless of individual preferences. An extreme case is that of functions whose range has two alternatives only, but for restricted domains there may be others. Hence, the flexibility of a social choice function, as measured by the extension of its range, will be an important property to check[17].

Feasibility of the social choices is another concern. Some rules may be focused on partial decisions, and one should then worry about the compatibility between these decisions and the overall resources held by society. For example, the traditional problem of choosing an optimal level for a public decision, along with some transfers of a private good, as studied by Clarke (1971), Groves (1973), and so many other authors, does not a priori require that such transfers should balance. In this survey I will briefly comment on this important line of work, but will mostly concentrate in studying rules that always guarantee that the recommended outcome is feasible.

In what follows, I will describe different groups of articles. All of them deal, in some way, with the trade-offs between strategy-proofness, domain extension, nondictatorship, efficiency, range dimensionality or similar other properties of social choice rules. In order to classify them, I adopt a double criterion.

On the one hand, I distinguish between the case of "common preferences", and that of "personalized preferences". In some social choice situations, it is natural to assume that if a ranking of alternatives is admissible for one agent, then it is also admissible for all others. This is the case, for example, when voters have to rank a number of candidates for office, or if they all have to indicate their desired level for a single variable (the amount of a public good, say). I classify problems where all agents are entitled to have the same preferences into the "common preferences" case. In other cases, the nature of the alternatives and of agent's preferences are such that what is admissible for an agent is not for others. For example, alternatives may be feasible allocations of private goods, and under the assumption of selfishness each

[17] For example, when the domain of preferences contains all those that are single-dipped relative to a given order, the range of strategy-proof rules can only consist of two alternatives, regardless of the number of those on which agents can express their preferences. Other bounds apply for any subdomain of single-dipped preferences (see Peremans and Storcken (1999), Barberà, Berga and Moreno (2009), Manjunath (2009)). This is in sharp contrast with the case where the range consists of single-peaked preferences, where the functions can be onto the set of alternatives under very mild assumptions on the domains.

of the agents will contemplate these allocations from their own perspective. What is an admissible preference for one agent on the alternatives is not admissible for others. But the same happens when alternatives consist of allocations involving some public and some private goods. Even the classical model where decisions consist in determining one level for a public good and a transfer of a private good for each agent belong to this second class, because each agent evaluates transfers to her differently than transfers to others. I call these the "personalized preferences" cases[18].

Probably the most important limitation of this survey is that it does not include a systematic treatment of the very important family of models where the objects of choice are combinations of allocation decisions involving the allocation of goods and of money transfers, and where typically (though not always) the preferences of agents are quasilinear in money. I had to establish some bounds for this survey on Social Choice, and I decided to concentrate on models where money transfers among agents do not play a fundamental, explicit role (although the definition of alternatives may include them implicitly). There will still be some mention to them in specific contexts where the boundaries are hard to establish, as in the analysis of assignment and cost sharing models (see Sections 9.3 and 9.4. But I exclude very important and extensive work on public goods allocation, in the line of Vickrey (1960), Groves (1973), Groves and Loeb (1975), Green and Laffont (1979), or Roberts (1979), and also ignore the very relevant subject of auctions, where economists and, more recently, computer scientists as well, have examined a host of questions related to incentives, in general, and to strategy-proofness in particular.

The second criterion I use to organize the rest of the survey has to do with the formal models, and also with the leading interpretation that each model is given.

I will start (Section 6) with an exhaustive description of models involving sets of alternatives defined by a finite combination of attributes, on which agents hold strict preferences satisfying some form of generalized single-peakedness. These models can be viewed as formalizations situations where alternatives can be described in terms of a finite set of characteristics, measured in units on which individual preferences are not necessarily monotonic. Although these models were mostly developed in the nineties,

[18]These cases are often referred to as the public good case and the private good case, but it should be clear from my discussion that I consider this a misnamer.

and were preceded by many other studies on strategy-proofness, their sequence will provide me with a good example of how a family of questions relating to strategy-proofness can be thoroughly analyzed. These models cover the case where agents must choose among public projects, or establish the level of one or several public goods, but no money transfers are envisaged.

Sections 7 and 8 turn attention to models where the outcomes of the social choice process adopt special structures. In Section 7, I consider rules whose outcomes are lotteries. In Section 8, I consider outcomes to be sets of alternatives. This formulation takes us away from the realm of social choice functions to that of social choice correspondences, and will need a detour regarding the framework and the interpretation of these models, as well as the definition of strategy-proofness within them.

Then, in Section 9, I will describe different models where preferences are definitely personalized and which arise in economic contexts: division problems, exchange, matching and cost sharing. These are examples of setups where the question of strategy-proofness gives rise to interesting analysis and conclusions.

In Section 10 I turn to more abstract domain restrictions, which were explored quite soon after the Gibbard-Satterthwaite theorem were formulated, in order to tie the new result with the traditional Arrowian framework in which many of the developments of social choice theory had taken place. Some of these restrictions applied to the common preference case, while others examined similar questions for personalized preferences. I also refer to the technical but important question of maximality.

6 Common domains. Strategy-proof rules for the choice of multiattribute alternatives

In this section I will thoroughly review the progress in understanding the domain restrictions allowing for non-trivial strategy-proof rules, and the type of rules which preserve this property in such domains, for a specific class of models. The models arise naturally as one considers the problem of deciding among the possible levels for k public goods, the location of facilities on the nodes of a grid, the choice of candidates who can be described by their performance regarding k different criteria. As we shall see, a natural twist in the model does allow to see it as a canonical way to formalize any collective decision problem involving a finite set of alternatives. Because of this, and since many of the ideas and techniques which have been developed in analyzing the issue of strategy-proofness within this model are also extensible to other cases, I will describe the results in this section at length.

6.1 Two alternatives

Although rather special, the case of choice between two alternatives 0 and 1 *is* important and, well studied[19]. Strategy-proof rules can be described as choosing 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[20].

6.2 Linearly ordered sets of alternatives

We now consider situations where a finite set of 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 $\mathcal{T}(\succcurlyeq_i)$, and (2) if alternative z is between x and $\mathcal{T}(\succcurlyeq_i)$, then z is preferred to x (intuitively, this is because z can be considered closer than x to the ideal $\mathcal{T}(\succcurlyeq_1)$). single-peaked preferences were first discussed by Black (1948a, 1948b) and they arise naturally in many contexts. As we shall see later, the same basic idea can be extended to more complex cases, where the betweenness relation does not necessarily arise from a linear order. But we start from this simple case[21].

---

[19] Part 1 of Fishburn's book (1972) is devoted to "Social choice with two alternatives"; Murakami (1968) devoted a chapter to study "Democracy in a world of two alternatives"; many recent works on voting start from the analysis of choices between two alternatives, in order to avoid some of the problems we deal with here, and to concentrate on others.

[20] References on simple games and their use in social choice theory are Peleg (1984, 2002), Abdou and Keiding (1991).

[21] Another important domain restriction in many types of analysis is that of single-crossing. An analysis of its consequences for the existence of strategy-proof rules is found

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 (1980a), 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, $\mathcal{T}_1$ and $\mathcal{T}_2$ 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

in Saporiti and Tohmé (2003) and Saporiti (2009).

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

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

We shall state immediately that this formula generalizes. There are also other ways to write the same rule. These are described in the next pages.

Moulin (1980a) characterized the class of all strategy-proof social choice functions on single-peaked domains. 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. An adaptation of Moulin's characterization is as follows.

Construction. For each coalition $S \in 2^N \backslash \emptyset$, fix an alternative a_s. Define a social choice function in a such a way that, for each preference profile $(\succsim_1, \ldots, \succsim_n)$,

$$f(\succsim_1, \ldots, \succsim_n) = \inf_{S \subset N} \left[\sup_{i \in S} (a_S, \mathcal{T}(\succsim_i)) \right]$$

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 Definition 2 below.

Theorem 3 *(Moulin, 1980a) 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 4 *(Moulin, 1980a) 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,*

$$f(\succsim_1, \ldots, \succsim_n) = med(p_1, \ldots, p_{n+1}; \mathcal{T}(\succsim_1), \ldots, \mathcal{T}(\succsim_n))$$

(A similar statement, with f defined with only $n-1$ phantoms, characterizes strategy-proof and efficient social choice functions)[22].

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

[22] The statement of this result by Moulin without his tops only requirement must be carefully qualified in the case of a continuum of preferences. The version we provide for expository purposes should be finessed. Two ways to do it would be either assuming a further condition that the range is connected, or a unanimity requirement on f. See Barberà and Jackson (1994) and Le Breton and Weymark (1999).

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 remember the case when we must choose among only two alternatives, 0 and 1. Strategy-proof rules can be described as choosing 1 unless there is enough support for the opposite, in which case 0 will be selected, where "enough support" is given by the list of those coalitions that will get 0 if all their members prefer it to 1. If that list defines a monotonic simple game, then the rule is strategy-proof.

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 the 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 (Section 5.3.1) 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* $\mathcal{C}$ *assigning to every* $\alpha \in B$ *a collection of non-empty coalitions* $\mathcal{C}(\alpha)$, *satisfying the following requirements:*

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

We'll denote left coalition systems by $\mathcal{L}$, and right coalition systems by $\mathscr{R}$. Elements of $\mathcal{L}$ 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* $\mathcal{L}$ *(resp.* $\mathscr{R}$*) on* $B = [a, b]$, *its associated generalized median voter scheme is defined so that, for all profiles* $(\succcurlyeq_1, \ldots, \succcurlyeq_n)$

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

and

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

(respectively,

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

and

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

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 they operate.

The description of generalized median voter schemes was first proposed by Barberà, Gul and Stacchetti (1993). It is easily extended to the case where the choices must be made not on a finite ordered set, but on the real line (Barberà, Massó and Serizawa (1998)).

These two expression of the form of strategy-poof rules on the real line (or on integer intervals) are not the only ones. Others are due to Kim and Roush (1984), and to Ching (1992, 1994, 1996, 1997), who provided another, representation of the same object. He maintained Moulin's idea of using medians and defined "augmented median rules", which allow for variable phantoms and can thus relax the anonymity that is implicit in the original version.

For clarification, I propose a couple of simple examples

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

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

$$\begin{aligned} f(1,2,3) &= 2 \\ f(3,2,3) &= 3 \\ f(1,3,1) &= 1 \end{aligned}$$

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*

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

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:

$$
\begin{aligned}
g(1,4,4) &= 1 \\
g(3,3,1) &= 3 \\
g(3,2,2) &= 2
\end{aligned}
$$

Other, earlier authors, had also introduced alternative descriptions of the rules leading to strategy-proof choices on the line. These were less constructive, and relied on the properties that actually characterize the rules. Chichilnisky and Heal (1981, 1997) proved that these rules must be "locally simple", that is, they must be locally constant or locally dictatorial. Border and Jordan (1981, 1983), and then Peters, van der Stel and Storken (1991) identified another property that is common and exclusive to such rules. They named it uncompromissingness: it means that no changes in the peak of any agent has any impact on the value of the function unless that agent's peak changes from the right of this value to its left (or vice-versa).

A few authors have considered how to extend the results to the case where alternatives can be assumed to be located on a graph, and preferences on the voters of this graph satisfy an extended notion of single-peakedness due to Demange (1982). Results are positive when the graph is a tree, and become negative when loops are allowed. See Schummer and Vohra (2000) and Danilov (1994).

Notice that although single-peaked preferences do not preclude indifferences among other alternatives, they are defined so that there is a unique maximal element for each agent.

As it turns out, allowing for indifferences in this and in other models tends to complicate the analysis of strategy-proof rules. In the present case, the basic result of Moulin is essentially preserved if one allows for "single-plateaud" preferences, having several contiguous maximal elements, in addition to single-peaked ones. Berga's (1998) careful discussion of this extension clearly illustrates how indifferences complicate the essential picture, even in such a simple model.

An interesting extension of the model we have discussed arises when agents must choose more than one point in the line. This problem was first described by Miyagawa (1998, 2001) and it accommodates, among other possible interpretations, the idea that the location (and modes of use) of several public facilities have to be jointly decided upon. Miyagawa (2001), characterizes a class of rules that are coalitionally strategy-proof, under the assumption that preferences over simple facilities are single-peaked, no congestion effects and the use of only one facility. Bogomolnaia and Nicolò (1999), characterize rules that are strategy-proof efficient and stable rules for the case where facilities congestion affects individual preferences.

6.3 n-dimensional grids

6.3.1 Strategy-proofness for generalized single-peaked domains

The assumption that social alternatives can be represented by a set of linearly ordered values is a very fruitful one. But 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 same formalism applies to the more classical problem of choosing simultaneously the level of provision of k different public goods. And many other interpretations are possible, including location decisions. Yet, each particular interpretation may suggest what are "natural" or relevant restrictions on the agents' preferences of these k-dimensional objects. At some levels, generality prevails while, at others the particulars derived from interpretation of the model do matter.

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 start with a specification that assumes a finite set of alternatives.

Similar results can be expressed in a continuous setting, and will be discussed in parallel. But the continuos setting also allows for new questions, regarding the connection of the model with the standard economic treatment of preferences on public goods. We shall consider these additional questions in the next section.

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 $\mathcal{T}(\succcurlyeq_i)$, then z is preferred to x. But in order to make the "betweenness" relationship precise, we must take a stand. Following Barberà, Gul, and Stacchetti (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 $\mathcal{T}(\succcurlyeq_i)$, if $z \in MB(x, \mathcal{T}(\succcurlyeq_i))$. 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 $\beta, \gamma \in B$, $\beta \in MB(\mathcal{T}(\succcurlyeq_i), \gamma)$ 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 these 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 $\mathcal{L}$ (resp. $\mathscr{R}$) be a family of K left (resp. right) coalition systems, where each $\mathcal{L}_k$ (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

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

and

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

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

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

and

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

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* $\mathcal{L}_1$ *be as* $\mathcal{L}$ *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,*

$$\begin{aligned} h((1,1),(2,4),(3,4)) &= (2,1) \\ h((3,3),(2,3),(3,1)) &= (3,3) \\ h((1,3),(3,2),(1,2)) &= (1,2) \end{aligned}$$

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 5 *(Barberà, Gul, and Stacchetti (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.*

Results in the same vein had been obtained by previous authors for a variety of contexts and under different assumptions regarding individual preferences.

Border and Jordan (1981, 1983)[23] did characterize strategy-proof rules for the k-dimensional problem with a continuum of choices on each dimension, under different assumptions regarding the preferences of agents. They got positive results, close in spirit to that of Theorem 5, for narrow and symmetric enough classes of preferences, which they call separable star-shaped, and include the quadratic case. Their characterization is in terms of the properties they had discovered to be required for one-dimensional rules to be strategy-proof under single-peakedness. It is based on the fact that the projections of these highly symmetric preferences on each of the axis does in fact induce a subclass of (symmetric) single-peaked preferences[24].

A similar extension to k-dimensional spaces, but expressed in terms of the local simplicity of the rules to be used on each of the axis, was obtained by Chichilnisky and Heal (1981, 1997)[25].

In that same context, a very special case arises if one restricts attention to euclidean preferences, whose indifference classes are hyperspheres. In that case, the choice of axis on which to project the different preferences becomes an additional issue. Laffond (1980) gave an early treatment of this case, which

[23] I mention the working paper, as well as its published version, because the latter is quite incomplete and requires constant reference to the original version.

[24] Notice that in k-dimensional settings with continous preferences, these will systematically contain indifference classes. I have already remarked that indifferences may introduce complications in the analysis. Such complications do not arise in these contexts because only the projections of the general preference on the axis matters, and these projections are single-peaked for restricted enough domains.

[25] Their 1981 paper remained umpublished until 1997, when an improved version appeared in *Social Choice and Welfare*.

was also, tackled by Kim and Roush (1984), van der Stel (2000), Peters, van der Stel and Storcken (1991). Additional results on preferences generated by strictly convex norms are contained in van der Stel (2000), and Peters, van der Stel and Storcken (1992, 1993)[26].

6.3.2 A special case: voting by committees

Theorem 5 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, Sonnenschein, and Zhou (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,

[26] For a more complete treatment of this point, see Sprumont (1995).

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, Sonnenschein, and Zhou (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 5 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 Section 6.1, 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).

Taking up from this particular model, where each proposal takes two possible values in each dimension, Ju (2002a, 2002b) considered the case where three possible valuations of an object, as being “good”, “bad” or “null” notice that this is a way to introduce indifferences in that simple model. He provides characterization of strategy-proof rules satisfying some additional properties. Larson and Svensson (2006) have also elaborated on the consequences of indifference in this and other contexts.

6.3.3 The issue of strategy-proofness for broader domains

As I already pointed out, Moulin’s (1980a) seminal paper was written for a general framework where agents could have preferences on, and choose from a one dimensional continuum of possible values. These can be interpreted in many ways. A leading interpretation is that agents must choose the level of a public good. Because of this, different authors soon tried to extend Moulin’s analysis to k dimensional spaces.

I have started to describe these extensions in Section 6.3.2, with the analysis of results on generalized single-peakedness. By doing that, I have given precedence to positive results that are possible in these restricted domains, and not in larger ones. But I have not respected the temporal sequence of the literature on the subject. It is now time to refer to earlier attempts to explore the issue of strategy-proofness on broader domains, which led to different negative results.

An early paper on strategy-proof choice in environments where preferences satisfy standard economic assumptions was due to Satterthwaite and Sonnenschein (1981). They examined the issue for different types of goods, and in the case of public goods they came up with an analog of the Gibbard-Satterthwaite theorem. Their analysis was based on some additional restrictions regarding the smoothness of allocation rules under consideration, and it only established that these rules should be locally dictatorial, a property that does not imply global dictatorship.

Several of the papers I will refer to did also contain some positive results for particular subclasses of preferences. To that extent, they have already been mentioned in the preceding section, and I will only emphasize their impossibility results here.

Border and Jordan was a major article in this direction (1981, 1983).

They explored different classes of saturated preferences, and proved that an impossibility arises as soon as the domain includes some which are not separable. Another early paper by Chichilnisky and Heal (1981, 1997) also proved negative results for large domains of preferences.

These papers had assumed that preferences satisfied some k-dimensional version of single-peakedness, and showed that the route to avoid impossibilities was to use preferences satisfying, in addition, some strong form of separability. This additional property, whose essence we have already imposed in the preceding section, was later proven to be essential in establishing the borders between possibility and impossibility results. This is addressed in the following section.

More in general, one could wonder about the possibility to extend Gibbard and Satterthwaite's result to environments where preferences satisfied some of the assumptions that are standard in economics. A paper by Barberà and Peleg (1990) tackled this issue in a larger setting that contains our k-dimensional spaces as a particular case.

Letting the set of alternatives be any metric space, and the admissible preferences be the set of all continuous utility functions on such alternatives, they proved that all nondictatorial rules whose range has more than three alternatives will be manipulable. This result opened the door to many others, since it had to develop new techniques of proof: indeed, prior proofs of the Gibbard-Satterthwaite result were based on arguments requiring the use of discontinuous utilities, and were not applicable. As we have already mentioned in Section 2.3.5, these new techniques are quite useful in proving that classical result as well. Yet, Barberà and Peleg's result on continuous preferences is neither implied nor implies Gibbard and Satterthwaite's. As soon as we have continuous preferences and an infinity of alternatives on a multidimensional space, more work is needed to attain results.

A paper by Zhou (1991a) continued the same line of research. He considered the set of alternatives to be any convex (non-empty) subset of the k-dimensional real space. The admissible preferences were those satisfying continuity and convexity. The latter is a strong requirement missing in Barberà and Peleg. Hence, Zhou's result is stronger to that extent, though it needs of some additional requirements. He established that the only strategy-proof nondictatorial rules on such domain had to operate on an extremely rigid range, whose dimension had to be less than two. This dimensionality condition is closely related to the condition that the range contains at least three alternatives: here, the only way in which three or more alternatives belong to a space of dimension less than two is by all lying on the same line.

Bordes, Laffond and Le Breton (1990) also obtained impossibility results in the k-dimensional setting.

Interesting variants of Zhou's results were attained by Moreno (1999) and Moscoso (2001), who used different domains, more amenable to economic interpretations.

Barberà and Jackson (1994) looked for a more constructive approach. They built on Zhou's remark on the special form of ranges that allow for strategy-proof rules in k-dimensional spaces and did provide their characterization for convex preferences. Essentially, they are rules that only use the restriction of these convex preferences on the linear range. Because these restrictions are typically single-peaked, Moulin's characterization for one- dimensional ranges and single-peaked preferences applies, with some technical qualifications.

6.4 Constraints. A first approach

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 $B \setminus Z$ 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. For example, 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, Massó, and Neme (1997); also Barberà, Massó, and Serizawa (1998) 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, Massó, and Neme (1997))

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 $\mathcal{T}(\succcurlyeq_i) \in Z$.*

Definition 5 *Let $Z \subset B$ and let f be a generalized median voter scheme on B, defined by the left coalition system $\mathcal{L}$ or, alternatively by the right coalition system $\mathcal{R}$. 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 $\mathcal{R}(\alpha_k)$ and $\mathcal{L}(\alpha_k)$, respectively, we have*

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

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 $(\alpha, S) \in \left(B - Z, 2^K\right)$.

Theorem 6 *(Barberà, Massó, and Neme (1997)) 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 $\mathcal{S}_Z$ the set of all single-peaked preferences with top on $Z \subset B$. Let f be an onto social choice function with domain $\mathcal{S}_Z^n$ and range Z.

Theorem 7 *(Barberà, Massó, and Neme (1997)) If $f : \mathcal{S}_Z^n \to Z$ is strategy-proof, then f is a generalized median voter scheme.*

Theorem 8 *(Barberà, Massó, and Neme (1997)) 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.*

6.5 The structure of strategy-proof rules

6.5.1 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.

This application of the general results in Barberà, Massó and Neme (1997) does appear in the same paper.

6.5.2 Embedding alternatives in a grid

In the preceding subsection we have seen how an appropriate identification of any abstract set of k alternatives as k points in a k dimensional grid could precipitate the Gibbard-Satterthwaite theorem. The theorem can be obtained as a corollary of results we already know regarding the structure of strategy-proof rules on grids when preferences are single-peaked and feasibility constraints may be required. This is just an example of a larger set of questions that one may ask regarding strategy-proof rules with the aid of our previous knowledge.

To see what is the more general nature of these questions, let us recapitulate. Essentially, we have learned that, whenever the domain of preferences for a strategy-proof social choice rule includes all the single-peaked preferences on the range of this rule having their top on that range, then the rule in question must be a generalized median voting scheme satisfying the intersection property.

One could boldly state the following converse of the statement. "Take any strategy-proof social choice function. There will always exist a method that identifies the alternatives in its range with some points in a grid, in such a way that: (a) the rule is a generalized voter scheme, and (b) the preferences in the domain of the rule are single-peaked for that embedding". As far as I know Faruk Gul stated this conjecture for the first time in the late eighties, and attempts to make it precise have been quite productive even if, as I shall comment, it cannot be exact. But I like to comment on it, even if only because it introduces a new point of view, connecting results on "abstract" alternatives with others regarding rather structured environments. In the context of grids, each alternative can be viewed as the conjunction of certain characteristics (one for each dimension of the grid) satisfied at some level of performance (determined by the position of the alternative on the scale that refers to that characteristic in particular).

Hence, in some way, each strategy-proof rule would be associated with an appropriate embedding in a grid, and that embedding would reveal the structural characteristics around which one should organize the choice of alternatives.

In fact, as we shall see, the conjecture is not exact, but inspiring. Even getting close to an appropriate statement needs many qualifications, each of which provides some insights. Let us consider them in turn. To do that, assume we are given a strategy-proof social choice function. First of all, can we always embed into a grid the set of abstract alternatives belonging to the range of this function? The answer is obviously yes, unless we impose any further restrictions on the embedding. So, let us formulate the question in a way closer to Gul's conjecture. Can we embed the range of the function in such a way that all the preferences in the domain are single-peaked? Again, the answer is trivial: yes we can. Just identify each alternative, as suggested in the previous section, with the unit vectors in a grid where each alternative accounts for one dimension. Since any preferences are single-peaked in that space, then all preferences in the domain, however small it is, should be single-peaked. Of course, this identification would be a dead end, because our characterization results require not only that some single-peaked preferences should belong to the domain, but that all single-peaked preferences should be admissible. So, let us start again. Can we embed the range of the function in such a way that all the preferences in the domain are single-peaked, and so that all single-peaked preferences with top on one of the alternatives (relative to the embedding) are in the range? Now, we are closer. Because if the answer is yes, we know that the rule will be a generalized median voter scheme. Moreover, we know that it will have to satisfy the intersection property.

Gul's conjecture cannot be completely true. We have no guarantee that the set of preferences in the range of strategy-proof social choice function will coincide with a rich enough set of single-peaked preferences, relative to any embedding, let alone with the whole set of them. And yet, some richness of domain is needed to guarantee that only median voting rules can satisfy strategy-proofnessness.

Even so, it has driven research by different authors, some of whose partial results are indeed enlightening.

For example, even when an embedding guarantees that preferences in the domain are single-peaked, it will seldom be the case that the set of alternatives consists of a full box. (Not, for example, if the number of alternatives is a prime, except for the most favorable case when they can be embedded on a line). Therefore, the discovery of the intersection property by Barberà, Massó and Neme (1997) was essential in allowing the very statement of the conjecture to have some meaning. Bogomolnaia (1998) studied carefully conditions under which one could properly speak of median voting (after an appropriate embedding of alternatives) when the initial setup lacks structure. The interest of the approach was indicated in Barberà (1996), along with several examples. Two recent papers by Nehring and Puppe (2005, 2007) have made a systematic search of several aspects related to the conjecture. In Nehring and Puppe (2007b) they extend the notion of generalized single-peakedness to cover a variety of structures, based on abstract notions of "betwenness". They also provide useful procedures that one could try when attempting to actually "construct" an appropriate space where to embed the alternatives in the range of a function. In Nehring and Puppe (2005) they exploit the requirements imposed by the intersection property in order to characterize the shapes of ranges that allow for "nice" strategy-proof rules, and to set them apart from other range forms that precipitate dictatorship.

6.5.3 The general structure of strategy-proof rules

Some essential features of strategy-proof rules emerge again and again in different contexts. Although they are not necessarily held by all such rules in all possible contexts, I find it convenient to discuss them here, because the results we just discussed regarding rules on K-dimensional grids are paradigmatic.

One first characteristic of strategy-proof rules in several contexts is that they are only responsive to a limited amount of the information contained in the preferences of agents. Specifically, rules that only use the information regarding what is the preferred alternative of each one of the agents emerge as the only candidates to respect strategy-proofness in many different cases. Remember that proving this "tops only" requirement (Weymark (1999) calls it nontop-insensitivity) is the first step in one of the proofs of the Gibbard-Satterthwaite theorem (Barberà and Peleg (1990)). The same requirement is proved to be necessary for strategy-proof rules defined over single-peaked preferences in k-dimensional grids. Sprumont (1991) and Barberà and Jackson (1994) showed that it was necessary in the one-dimensional case (Moulin (1980) had a priori restricted attention to the class of rules satisfying the property). The different papers involving the use of generalized median voter schemes start by showing some version of the "tops only property", even if the same basic idea may require slight additional qualifications depending on the context. For example, in those contexts where the best alternatives of agents might not be feasible (See Barberà, Massó and Neme (1997), or Weymark (1999)), then the condition still holds but now requires that the rules should process information about those feasible alternatives that are best. In other terms, the rules should only use information about those alternatives that are top on their range. Weymark (2006) explores the possibility of finding general conditions under which the "tops only" condition is a necessary condition for strategy-proofness.

A second important property of strategy-proof rules operating on k-dimensional alternative spaces is that they must be decomposable: that is, they should be possible to express as the combination of k rules, each one operating on one of these dimensions, each one being itself strategy-proof. This decomposability is quite independent of the particular set of preferences that are admissible in each of these dimensions. When preferences on each dimension are single-peaked, then the rules can be decomposed in k rules, each one of them being of the type described by Moulin (1980). But if the set of admissible preferences on each dimension is broad enough, then only dictatorships on each dimension may be strategy-proof, and only compositions of dictatorial rules may emerge. The bottom line is that, although decomposability comes along with single-peakedness in positive results, it is a more general requirement for strategy-proof rules operating on k-dimensional spaces.

Decomposable rules must be such that the combination of characteristics chosen in each dimension generates a feasible outcome. This is no problem when the ranges of functions are full k-dimensional boxes: any combination of choices, dimension by dimension, generate an element in the box. No coordination is needed between what are the choices in one dimension and what emerges in another. However, in contexts where the ranges are not full boxes (due to constraints, for example), then decomposability is not the final requirement. The rules in each one of the dimensions must be "coordinated" enough to ensure that they will never recommend the choice of unfeasible combinations of characteristics. This is the role played by the intersection property.

These are basic features that appear in all their neatness under special conditions. Specifically, they require the existence of a single top alternative in each of the relevant dimensions of the problem, and the possibility to identify these "partial" tops from just knowing what the global top alternative is for an agent. Thus, complications arise in contexts when agents may be indifferent among several possible top alternatives. Even then, and sometimes using additional restrictions (like non-bossiness), results regarding the simplicity of inputs and the decomposability of strategy-proof rules may still hold for these more complex environments. Another aspect to watch for, because it is determinant for these general features to arise, is that the domains of definition of the functions should be rich enough.

The first results on decomposability were due to Border and Jordan (1983) and Chichilnisky and Heal (1981, 1977). Different papers developed their own separability results as they worked along to get specific results. A systematic application of the principles evoked here appear in Barberà, Sonnenschein and Zhou (1991) and in Barberà, Massó and Neme (1997).

More specifically, different authors have carefully studied the general conditions under which each of these properties become necessary for strategy-proofnessness. Le Breton and Sen (1995, 1999) and Le Breton and Weymark (1999) studied separately the cases when preferences are strict and the additional complications induced by the presence of indifferences. Weymark (1999) concentrated on decomposability and obtained additional results based on a variety of domain conditions.

A superb account of these structural features and their discovery is found in Sprumont (1995).

A major issue we have already hinted at in section 3.2 is that of the difficult compatibility between efficiency and strategy-proofness. Two well known domains where these two conditions can be jointly satisfied by non trivial rules involve extensions of the notion of single-peakedness to appropriate sets of alternatives. One is the case when one candidate or one location has to be elected, and preferences are single peaked (Moulin, 1980). Another case arises when one good must be rationed and the agent's preferences over shares are single peaked (Sprumont, 1991). See Sections 6 and 9.1. A third case, this time in a two-dimensional space, is considered in Kim and Roush (1984). Nehring and Puppe (2007a) discuss the trade-off between efficiency and strategy-proofness in a very constructive way, by concentrating on a large class of preference domains and characterizing the rules that can meet both requirements. The analysis of these rules proves that compatibility requires either a low dimensional space of alternatives (as in the above mentioned references) or the rule to be "near" dictatorial.

6.6 Constraints revisited

Until now, the papers I have described on the issue of constrained ranges were based on the assumption that the domain of preferences only included those satisfying two conditions:

1. That the preferences on the range were the restriction on the alternatives in the range of some single-peaked preferences on the minimal box containing them.
2. And that the preferences on the range had their top in the range.

I will now discuss how the consequences of dropping this second condition. Before I do, let me briefly argue that it is sometimes appropriate to use it, while in other cases it seems unnecessarily restrictive.

We have mentioned two scenarios under which it seems natural to concentrate in social choice functions whose range is not a box. One such scenario comes from considering any social choice function on abstract sets of alternatives and then embedding them in a k-dimensional space. In that case, it is perfectly natural to think of the preferences on alternatives in the range as the primitives, and assumption 2 just assumes that there is one best feasible alternative. We can then see assumption 1 as a restriction on the domain that is sufficiently rich to provide characterization results, and sufficiently restrictive to allow for non-trivial strategy-proof rules.

The second scenario is one where the structure of the set of alternatives comes naturally with the k-dimensional model, so that in principle preferences can be defined on all alternatives, feasible or not. In that context, we may study the consequences of additional constraints, precluding some alternatives from being chosen. Under this interpretation, assumption 1 is still perfectly acceptable, but it seems unnatural to exclude the possibility that agents might prefer most some alternative that happens to be unfeasible.

Barberà, Massó and Neme (2004) studied the impact of feasibility constraints when the top of agent's preferences may be out of the range. They provide a characterization of all possible strategy-proof rules for all conceivable constraints. Since this is a complex task, they do it for the simplified world that was first described in Barberà, Sonnenschein and Zhou, (1991), as described in Section 6.3.2 above. This is the world where each alternative can be described by a collection of characteristics, each of which is binary.

In that context, rules that satisfy strategy-proofness are still voting by committees, with ballots indicating the best feasible alternative for each agent. Yet, the committees for different objects (or combinations of characteristics, depending on the interpretation that we give to vectors of values) must now be interrelated, in precise ways which depend on what alternatives are feasible. Specifically, each family of feasible subsets (in one of the interpretations) will admit a unique decomposition , which in turn dictates the exact form of the strategy-proof and onto social choice functions that can be defined on it. One important feature arising in that context is that results are more sensitive to the domain of admissible preferences than they are in all the contexts we described till now. Specifically, the class of rules that can be strategy-proof when preferences are additively representable is substantially larger than the set of rules having that property when all separable single-peaked preferences are allowed. This is in contrast with the results in the preceding literature on k-dimensional grids, where the same characterizations obtain for both sets of restricted preferences.

The characterization of strategy-proof rules for the case where all separable preferences are allowed is quite negative. The decomposition result implies that, except for very special cases, non-dictatorial rules will be manipulable. In contrast, the result for additively representable preferences leaves a wider slack. Depending on the shapes of feasible alternatives, results with a positive flavor may arise, and strategy-proof non-dictatorial rules may exist.

The contrast between these results, allowing for restricted preferences whose best element need not be feasible, and the preceding results, is quite striking. The result in Barberà, Massó and Neme is complex, even if restricted to the case where each dimension admits two values only. Not much could be gained by considering the more general case where multiple levels are allowed for in each of the relevant dimensions of an alternative. An exploration in this direction is to be found in Svensson and Tortensson (2008).

7 Common domains. Probabilistic voting schemes

Voting and chance have been combined as collective decision devices since ancient times. In one extreme, certain public officials were in the past and are even now chosen at random among those eligible for office. In the other, voting determines the outcome of the election to the last detail. But there are many variants in between, where agents vote and chance also plays a role to determine the final choice. In this section I will review work that explicitly models the outcomes of voting as lotteries. In the next section I will review social choice correspondences: while these admit different interpretations, many authors have actually analyzed them under the implicit or explicit assumption that the final choices among the different alternatives pre-selected by the correspondence will eventually be made by resorting to chance[27].

The study of methods which combine voting with chance has a long tradition, but I shall restrict my account relating to incentive theory and start with the pioneering work of Zeckhauser (1973). The author provides a framework where agents are endowed with Von Neumann-Morgenstern preferences over lotteries, and rules determine a lottery over alternatives as a function of preference profiles. He proposes a good definition of strategy-proofness and discusses some of the characteristics of his proposed methods. His results on strategy-proofnessness are partial and not fully correct, (as shown years later by other authors), but the paper is inspiring and generated a number of follow-ups, in different directions.

In fact, when modelling the interaction between voting and chance, one must be specific about the range of objects that will result from the interac-

[27] There is a substantial literature where social preferences are modeled as fuzzy relations. We shall not survey it here, especially because not much of it addresses the issue of strategy-proofness. An exception is Perote and Piggins (2007).

tion among voters, and also on the domain of preferences that the rule will elicit.

In my opinion, the neatest choice, for a range is to describe the outcomes of the voting process as lotteries (probability distributions) over outcomes. And the most natural choice for a domain, in that case, is to allow voters to express their preferences on such lotteries. This was Zeckhauser's initial proposal, and one that we shall explore.

Of course, not much mileage is to be gained from such a model unless some domain restrictions are predicated. Notice that, in the case of lotteries, it is natural to assume that their range contains a continuum of alternatives, even if they are based on a finite number of prizes, and this case is not fully covered by the Gibbard–Satterthwaite theorem. Yet, other results in the same spirit suggest that models that do not restrict the agent's preferences over lotteries would simply fall into the same kind of impossibility that is announced by the theorem.

Yet, a very natural restriction arises in that context, and it is to assume that agents´ preferences over lotteries satisfy the Von Neumann-Morgenstern axioms (or some variant of them) and are therefore representable by utility functions satisfying the expected utility property.

To make things simpler, we'll deal with *lotteries over a finite set of alternatives* A, and denote their set by $L(A)$.

Utility representations of preferences over $L(A)$ will be denoted by $u, v, \ldots,$ and their set by U. We'll assume that they are normalized, in such a way that each preference is represented by one and only one function. In order to fix ideas, we'll assume that the normalization is such that the most preferred alternative is assigned utility 1, the worse alternative has utility 0, and that not all alternatives can be indifferent to all others.

Then, our first object of study will be functions of the form $f : U^n \rightarrow L(A)$, where n stands for the number of voters.

I shall call them probabilistic rules. As for a definition of strategy-proofness in that context, it is just a re-writing of the same notion we have discussed until now. Namely, a probabilistic rule f is strategy-proof iff, for all utility profiles all $(u_1, ..., u_i, ..., u_n) \in U^n$, all $i \in \{1, ..., n\}$ and all $u_i^{'} \in U$ we have that

$$u_i{\cdot}f\,(u_1, ..., u_i, ..., u_n) \geqslant u_i{\cdot}f(u_1, ..., u_i^{'}, ..., u_n).$$

Notice that this is, in a different notation, the same definition we have been using all along: the lottery resulting from any agent declaring his/her true preferences is at least as preferred (i.e. provides as much expected utility) as any one attainable by declaring any other preference (i.e., any other admissible utility function).

Even if this is a very natural setup, some early literature departed from this general model and considered an alternative scenario (which can in fact be re-interpreted as a particular case), where the outcomes are lotteries over outcomes and preferences are rankings of these outcomes.

According to the logic underlying this survey, these models appear to be poorly specified, since the preferences expressed by the voters do not refer directly to the objects to be chosen. Yet, the models are attractive from a less formal point of view, since they can be considered the natural expression of voting rules which rely on the ordinal preferences of voters over outcomes, but introduce a chance element by determining that the result of the process is a lottery over these outcomes. Formally, they take the form

$$g : \mathcal{R}^n \to L(A)$$

Where $\mathcal{R}$ stands for the set of preorders over A (alternatively, in some cases, for the set of strict orders).

This was the setup of a pioneering contribution by Gibbard (1977), who gave these functions the name of decision schemes. He also proposed the following definition of strategy-proofness in that context.

A decision scheme $g : \mathcal{R}^n \to L(A)$ is strategy–proof iff, for all preferences profiles $(R_1, ..., R_i, ..., R_n)$, all $i \in \{1, ..., n\}$, all $R_i^{'} \in \mathcal{R}$, and all "$u_i \in U$ *fitting* R_i",

$$u_i{\cdot}g\left(R_1, ..., R_i, ..., R_n\right) \geqslant u_i{\cdot}g\left(R_1, ..., R_i^{'}, ..., R_n\right)$$

This definition, due to Gibbard, may seem to incorporate an additional and somewhat extraneous element in the definition of strategy–proofness. Indeed, the expression "u_i *fitting* R_i" refers to any $VN-M$ utility function that respects the ranking of alternatives established by R_i and it appears as an added element in the valuation of outcomes.

However, as I already mentioned, there is a way to re-interpret voting schemes as a particular subclass of probabilistic rules. And Gibbard's definition will become equivalent to the standard one under this re-interpretation, which I now offer.

Notice that, given preferences over lotteries ($v \in U$), they determine uniquely a preorder over outcomes ($R \in \mathcal{R}$). Conversely, each order over outcomes can be identified with an equivalence class in U, naturally defined as the set of utility functions that give rise to the some order or even with a single one, if we accept the normalization suggested for the preceding class of models. In fact, given a v, the elements in its class are all the (normalized) utility functions which result from a monotone transformation of it. With this in view, we can consider that a voting scheme g is a particular probabilistic rule satisfying the following invariance property:

Invariance:

For all n-tuples of monotonic transforms

$$\varphi i : U \rightarrow U, \text{and all } (v_i, ..., v_n) \in U^n,$$
$$g(v_i, ..., v_n) = g\left(\varphi_i\left(v_i\right), ..., \varphi_n\left(v_n\right)\right)$$

This is equivalent to say that such functions do not process any cardinal features of the $VN-M$ utility functions, but just the order of final outcomes.

Under this identification, Gibbard's definition becomes standard, if we think that the image of a scheme g will be the same for all of the utilities belonging to a given class, and yet it will suffice that one of the preferences over lotteries within the class recommends manipulation in order to violate the condition of strategy-proofness class function.

We now turn to the known results about decision schemes. Even if the more general framework is that of probabilistic rules, it is good to start with Gibbard's (1977) results, since his characterization of strategy–proof decision schemes is particularly elegant, and its knowledge will facilitate the presentation of other facts.

Gibbard provides the following definitions, which I present somewhat informally.

- A decision scheme is *unilateral* if its image is exclusively determined by the preferences of one agent alone. Notice that a dictatorial rule, where the dictator's preferred outcome gets probability one, is unilateral. But another example is the rule where an agent's favorite alternative gets $\frac{2}{3}$, her second best gets $\frac{1}{6}$, and her third best another $\frac{1}{6}$.

- A decision scheme is *duple* if its image always consists of a lottery which attaches positive probability to at most two alternatives, which are always the same. An example of a duple scheme is one that pre-selects two alternatives x and y, never attaches any probability to any other, and determines the choice probability of each of these two alternatives as a function of the support that agents give to one of them over the other.

- A decision scheme is *non-perverse* if switching the place of an alternative upward in an agent's ranking never decreases the weight attached to that alternative.

- A set X of alternatives *heads a ranking* P_k if all alternatives in X are preferred to those in $V - X$ according to P_k. A decision scheme d is *localized* iff for any agent k and any pair of preference profiles $(P_i, ..., P_k, ..., P_n)$ and $\left(P_i, ..., P_k^{'}, ..., P_n\right)$, which only differ in k´s preference, if X heads both P_k and $P_k^{'}$, the total weight attached by the scheme to alternatives in X is the same for both profiles.

We can now state Gibbard´s elegant characterization of strategy–proof decision schemes.

Theorem 9 *(Gibbard 1977). A decision scheme is strategy proof if and only if it is a probability mixture of decision schemes, each of which is non-perverse, localized and either unilateral or duple.*

The above statement uses the term "probability mixture" to describe the combination of rules attaching to each profile a convex combination of the lotteries obtained by each of the decision scheme that would be "mixed". Since the outcomes of each of this decision are lotteries over the same set of alternatives, convex combinations of these lotteries are well defined.

An important corollary of the above characterization is due to Sonnenschein (cited in Gibbard).

Corollary 1 *A decision scheme is strategy-proof and ex-post efficient if and only if it is a random dictatorship* (*for* $\#A \succcurlyeq 3$).

Ex-post efficiency requires that no alternative which is Pareto dominated ever gets a chance of being chosen. Clearly, ex-post efficiency rules out the possibility of duple schemes being part of the decomposition of a strategy-proof scheme. It also rules out the possibility of unilateral schemes allowing for a positive probability to dominated alternatives. What is left is random dictatorship: the probability mixture of dictatorial schemes.

Gibbard´s characterization, and its corollary, have been usually interpreted as rather negative, but this involves some fallacy of composition. While unilateral and duple schemes are certainly unattractive, their probability mixtures need not be. In fact, Barberà (1979b) provides a number of results proving that Gibbard´s class contains rather attractive methods, which result from the natural extensions to a probabilistic framework of the two basic properties used in making deterministic choices: majority and positional (or point) voting. These positive results, in turn, must be taken with a grain of salt, since all these schemes could be manipulated to the benefit of two-agent coalitions (Barberà (1979a), Dutta (1980b)).

A necessary qualification to Gibbard´s characterization result is that it only applies when individual preferences over alternatives are strict. When indifferences are allowed, the same general idea still applies, but hierarchies of agents enter the picture in a complex manner (see Gibbard (1978) and Benoît (2002)).

Gibbard´s proof is constructive and elegant, but complicated. Nandeibam (1998) and Duggan (1996) provided simpler proofs of the main corollary of the result, for the special case where efficiency is also required.

I have already noted in Sections 1 and 2 that the connections between Arrow's and Gibbard and Satterthwaite's results were the subject of a lot of attention. An analog parallel was established by Pattanaik and Peleg (1986) and by McLennan (1980) among Gibbard's strategy-proof decision schemes and the probabilistic analogues of Arrow's social welfare functions studied in Barberà and Sonnenschein (1978).

A few papers have investigated the properties of rules that choose lotteries over more structured alternatives. Ehlers, Peters and Storcken (2002) studied rules which choose lotteries over the real line, when the ordinal preferences of agents are single-peaked. They proved that an extension of Moulin's results for social choice functions does characterize the decision schemes satisfying strategy-proofness. In this extension, fixed probability distributions play the role that phantom voters (or fixed ballots) play in the deterministic case.

A further extension is due to Dutta, Peters, and Sen (2002). In that case, lotteries are defined over a convex set of Euclidean spaces, and agents are endowed with strictly convex, continuous single-peaked preferences on that set. Their main result is that all unanimous mechanisms satisfying strategy-proofness in the sense of Gibbard (1977) must be random dictatorships. Ehlers (2002) analyzes probablilistic methods when preferences are single-dipped. In this work, and also in Ehlers and Klaus (2003b) applications of probabilistic rules to assignment problems are discussed.

I now turn to the study of probabilistic rules, which assign a lottery to any profile of preferences over lotteries satisfying the axioms of expected utility.

To my knowledge, a full characterization of strategy-proof probabilistic rules is not available, but it is possible to go a long way in the understanding of such rules by following the path of the previous result by Gibbard (1977), and adapting its definitions to the present (and larger) framework.

Let us first extend the notion of a unilateral rule.

Select any subset of lotteries, and consider the function obtained by letting a single fixed, agent to select the best lottery from that set, (given any profile). Such a function (if well defined) will be a unilateral probabilistic rule, and it will obviously be strategy-proof. Depending on the set's characteristics, the chosen agent's best set may contain several lotteries. In that case, we may extend the notion to that of *hierarchically unilateral* rules. These are defined by some fixed set of alternatives and a given order of agents, in such a way that, for any profile, the first agent in the list selects her best lotteries, the second chooses her best among those, the third agent picks her best among those still left, etc. Eventually, a fixed tie-breaking rule may be appended to guarantee that a unique choice is finally made.

Gibbard's notion of a duple scheme can also be extended. Now, one can fix any two lotteries[28] and declare that a probabilistic rule is *duple* if it always selects a convex combination of the two. Clearly, if the choice is made in such a way that the probabilities attached to each of the basic lotteries responds "adequately" to the preferences of agents over them, this gives rise to strategy-proof rules. For the purpose of the following statement, we incorporate into the definition of a duple the notion that it properly responds to agent's preferences.

Proposition 1 *(Barberà, Bogomolnaia, van der Stel (1998)). All probabilistic mixtures of (hierarchically) unilateral and duple probabilistic (social choice) functions are strategy-proof.*

Notice that this falls short of a characterization à la Gibbard, because the converse is not present. Proving that all strategy-proof rules are in that shape is a hard task. To begin with, the set of alternatives in this context is a continuum, and the type of constructive proof provided by Gibbard is a non-starter. Moreover, hierarchical versions of duples are also strategy-proof and hard to describe. At any rate, the proposition above allows for a number of interesting qualifications. One is that, in this context, it makes sense to extend the notion of probabilistic mixtures to even consider integrals of probabilistic rules. Another one is that some rules may now be expressed

[28] Notice that duples, in Gibbard's sense, can only be based on degenerate lotteries.

as combinations of duples and also as combinations of unilaterals (an added difficulty for a full characterization). Most importantly, it allows to prove a partial result which explains the difficulties to design well-behaved strategy-proof functions, even when they exist.

Proposition 2 *(Barberà, Bogomolnaia, van der Stel (1998)). If a strategy-proof probabilistic social choice function is twice continuously differentiable, then it is a convex combination of unilaterals.*

This result was improved by Bogomolnaia (1998), who relaxed the continuity requirement. It shows that strategy-proofness requires, in a deep sense, some form of discontinuity, as also shown in other contexts. In fact, early results by Freixas (1984) had already bumped into this kind of difficulty, when applying the standard techniques developed by Laffont and Maskin (1979).

Again, if one requires efficiency, the only possible rules to satisfy strategy-proofness collapse to random dictatorship, as in the case of decision schemes. This was proven by Hylland (1980). For a new proof and some additional results in this context, see Dutta, Peters and Sen (2007).

8 Common domains. Social choice correspondences

The Gibbard-Satterthwaite theorem and many of the results that ensued, apply to social choice functions. That is, to rules which select one alternative, and only one, for each preference profile. Yet, in many cases, collective decision processes are formalized by means of correspondences. These allow for nonempty sets of alternatives, not necessarily singletons, to be associated with each preference profile. That formalization may be natural under different interpretations. If we view social choices as the maximal elements of some social preference relation, then sets will be chosen if more than one alternative is best. If we seek some symmetry in the treatment of voters and/or alternatives, sets should be chosen under preference profiles that are highly symmetric. But this formalization also requires an appropriate interpretation. Since alternatives are mutually exclusive, the images of a social choice correspondence cannot be interpreted as giving a final social outcome. Some different interpretation must be provided, to connect the choice of a set of alternatives with the choice of a single social outcome, and the most common is as follows. The set is interpreted as the result of a first screening process, after which every alternative in the set, and no other, is still a candidate to be the final choice. Different assumptions about the process to be used for the final selection will result in different methods to evaluate the individual preference for sets. And, as we shall see immediately, these evaluations are crucial to make sense of the notions of manipulation and strategy-proofness, when applied to correspondences.

Many models in economics and in politics formalize collective decision-making by means of multi-valued rules, or correspondences. The Walrasian correspondence assigns to each economy the set of allocations that are sustained as equilibria of this economy, for some price vector. Typically, there will be several such allocations for the same economy. Similarly, the core correspondence in economic or political games assigns to each relevant game the set of imputations in its core. Again, this set is seldom a singleton. Arrowian social welfare functions assign a social preference to each profile of individual preferences, and it is usually understood that the social preference will be used to choose the socially preferred among the feasible alternatives. Once more, there may be several of them, once we accept that the social preference might be a preorder.

As a matter of fact, most of the analysis of incentives in public decision-making does benefit from formalizing such processes by means of correspondences, rather than functions. Implementation theory, with its multiple possibility results, would collapse to little if one restricted attention to social choice functions alone.

Yet, the study of strategy-proofness for social choice correspondences has been quite problematic, especially because different authors interpret these objects in different manners. In social choice theory, alternatives are defined as being mutually exclusive: if an alternative a occurs, then any other alternative b cannot occur. Hence, the meaning of having more than one alternative socially chosen must be clarified. In general terms, there is agreement to interpret the set of alternatives resulting at a given preference profile as the result of some screening process: alternatives not selected do not qualify as adequate. But it is unusual to specify how a final choice is to be made among those that pass the screening.

In many cases, being silent about the resolution of this indeterminacy does not pose any problems. Which competitive allocation or what core allocation will prevail when there are several are not essential issues in economic analysis. But this silence becomes a problem when trying to define strategy-proofness.

Consider social choice correspondences, of the form

$$c : \mathcal{R}^n \rightarrow 2^A \backslash \emptyset$$

Intuitively, this correspondence is manipulable iff for some profile of preferences, some agent can obtain a better result by misrepresenting her preferences than by declaring them truthfully. Yet, the outcomes at the truthful and the nontruthful profile may both be sets of alternatives, and who knows when a set is better than another? Certainly not the analyst who studies the model, unless additional assumptions are made, allowing to extend the voter's preferences from the set of alternatives to its power set.

A possible reaction to this difficulty is to declare social choice correspondences as being poorly specified. If agents choose sets of alternatives, they should be asked about preferences on sets, not on alternatives in isolation. I shall go back to this point of view when describing the work of Barberà, Dutta and Sen (2001). However, the use of social choice correspondences is pervasive, and it is worth pursuing the issue as we have formulated it. One should keep in mind, though, that all the literature on strategy-proof correspondences must include some implicit or explicit assumption about the way in which agents rank sets of alternatives, even if this information is extraneous to the description of the correspondence, when its domain simply contains rankings of alternatives.

Because of the instant success of the Gibbard-Satterthwaite result, there was an early literature concerning the possibility of extending it to the case of correspondences.

Different authors took different paths in doing it.

The essential difference among them was in regard to their assumptions about preferences over sets. Clearly, there was no need to be specific about the complete extension of preferences from alternatives to their power set. It was enough to allow for some partial comparisons in order to get manipulability of correspondences.

To illustrate the nature of these assumptions, let us consider three of the routes taken by early papers in this field.

Gärdenfors (1976) based his analysis on the following assumption on how these preferences were connected:

Gärdenfors' partial extension

"For all nonempty subsets A and B of X, $A\mathfrak{R}_G B$ iff one of the following conditions is satisfied:

(i) $A \subset B$, and for all $x \in A$ and $y \in B - A$, xRy.

(ii) $B \subset A$, and for all $x \in A - B$ and $y \in B$, xRy.

(iii) Neither $A \subset B$ nor $B \subset A$, and for all $x \in A - B$ and $y \in B - A$, xRy.

From the general definition of the strict relation it follows that $A\mathfrak{P}_G B$ iff either (i) and there exist $x \in A$ and $y \in B - A$ such that xPy or (ii) and there exist $x \in A - B$ and $y \in B$ such that xPy or (iii) and there exist $x \in A - B$ and $y \in B - A$ such that xPy."

Another early proposal was due to Kelly (1977):

Kelly's partial extension

"For all nonempty subsets A and B of X, $A\mathfrak{R}_K B$ iff xRy for all $x \in A$ and $y \in B$.

From this definition it is easy to show that $A\mathfrak{P}_K B$ iff $A\mathfrak{R}_K B$ and there exist $x \in A$ and $y \in B$ such that xPy."

A third proposal was made by Barberà (1977a):

Barberà's partial extension

If xPy, then $< x > \mathfrak{P}_B < xy > \mathfrak{P}_B < y >$.

Notice that these three criteria imply that some comparisons among sets are possible, once we know the agent's preferences, while they do not assume all sets to be comparable. Each one of them was used to obtain an impossibility result that was analogue to the one by Gibbard and Satterthwaite, under some additional conditions

Before turning to the rest of the literature, let me mention some of the difficulties associated with these early attempts, and with many later ones.

First of all, how should one judge the merits of some suggested extension *vis à vis* those of another proposal? From a formal point of view, the weakest the extension, the stronger the impossibility result that one would obtain. In that respect, *ceteris paribus*, an impossibility obtained under Barberà's extension would be stronger than one obtained by using any of the other two, since it would require less assumptions about the voter's way to rank sets. However, the results in any of these three papers are not strictly comparable. This is because each of the authors had to introduce some additional assumption in order to get an impossibility result: Gärdenfors assumed anonymity and neutrality conditions on its rules. Barberà assumed a positive response property. Kelly imposed a consistency requirement across the choices that the rule would make if some alternatives would be dropped.

In balance, the literature on strategy-proof social choice correspondences confirms the robustness of the Gibbard-Satterthwaite impossibility result.

Although some additional requirements are needed, and technical problems tend to marr the sharpness of their statements, most contributions end up proving some analogue of the classic result for functions in the multi-valued case. From that point of view, the different results reinforce each other: any combination of conditions which induces an impossibility is informative, and all of them taken together establish that there isn't any significant room for strategy-proofness that results from relaxing the single-valuedness assumption.

In spite of that, some authors have explored the general issue from a more relativistic viewpoint. Alan Feldman (1979a, 1979b, 1980), in a series of interesting articles, investigated the extent to which different definitions of strategy-proofness for social choice correspondences would precipitate impossibilities or rather allow for some possibility results. In particular, he examined the question for the Pareto rule under a variety of definitions. He concluded that this important rule would be considered manipulable under some definitions, and not for others.

One author who deserves special mention again here is Pattanaik. As I have already mentioned, he was close to the same result that Gibbard and Satterthwaite did attain, but his results were not as neat because he was working in a context similar to that of Arrow: he considered rules whose outcomes were binary relations, and this led him to also deal with the connections between the choices that such relations would induce under different agendas, or subsets of alternatives. One of the questions that arises naturally in that context, if one follows Arrow's tradition to allow for social preorders is that social choices are often multivalued. Hence, Pattanaik (1973, 1974, 1978) usually assumed that, when comparing different sets, agents would use some form of the maximin rule. Another issue in that context is the consistency of choices across agendas, an issue that Gibbard and Satterthwaite did avoid by only dealing with functions that choose an alternative from the complete agenda, and being silent about choices when not all alternatives are available. Part of the early literature on strategy-proofness of social choice correspondences inherited from Arrow and Pattanaik's formulation in requiring some consistency conditions. See Kelly (1977) and Barberà (1977b), for example, as well as many of the results surveyed in Muller and Satterthwaite (1985). We'll go back to that issue in Section 10.1.

Until now I have referred to different proposals on how to make assumptions regarding preferences of sets which are not explicit in the model, but described them as purely formal. A richer point of view is to discuss the merits of each of these extension proposals by referring to specific interpretation of correspondences. In fact, people may rank sets of alternatives in many different ways, depending on the meaning of such sets. Barberà, Bossert and Pattanaik (2004) contains an extensive review of several strands of literature about set rankings based and their connections with underlying rankings of singletons. Parts of that literature are of interest to those attempting to define strategy-proofness for correspondences.

What are the leading interpretations for a correspondence? Not all authors are explicit about it, but one very common assumption is that some random device will eventually be used to "break the ties" and choose one of the alternatives in the image of the correspondence. In an early survey of different criteria to partially rank sets, Gärdenfors (1976) was quite explicit about the fact that many of the proposals had this interpretation in the background. However, other authors insist that this interpretation is only one among several.

After a period when the issue was not much under discussion, the question of manipulation of social choice correspondences made a comeback. As we shall see, the interpretation of sets of alternatives as the basis for some unspecified lottery was quite explicit in that later work.

But there were exceptions. As I already mentioned, some authors reject quite explicitly the interpretation that the images of social choice correspondences are the basis for lotteries. Campbell and Kelly (2002, 2003a, 2003b) are papers which propose other alternative interpretations and define manipulability under set comparisons which are justified by other means. Specifically, leximin orderings of sets are used to establish a number of impossibility results regarding the existence of strategy-proof non-resolute social choice procedures. In a similar spirit, see also Sato (2008). His results are extended by Rodriguez-Alvarez (2009). Nehring (2000): discusses the relationship between the properties of monotonicity and a generalized version of strategy-proofness for correspondences. Ozyurt and Sanver (2009) provide a general impossibility result when preferences over sets are based on lexicographic orderings. Other variants of the topic include the analysis of group strategy-proof correspondences (Umezawa (2009)), or the study of the consequences of imposing restrictions on the size of the chosen sets of alternatives (Oyzurt and Sanver (2008)). The latter is connected with the analysis of multi-dimensional alternatives discussed in Section 6.3.

Duggan and Schwartz (2000)) worked within the classical framework of social choice correspondences. They define a correspondence to be manipulable if there exists an individual, a “true” preference and a “false” ranking such that for every lottery over the set obtained by lying, and for every lottery obtained from telling the truth, there is an expected utility function consistent with the voter’s true ranking of basic alternatives for which the expected utility of the first lottery exceeds the expected utility of the second.

This definition is quite demanding, and it is not easy to prove an impossibility result without, again, some additional conditions. Duggan and Schwartz manage to keep such conditions at a minimum, by showing that there will always be such type of manipulation when there are at least three alternatives, for all correspondences satisfying citizen sovereignty, nondictatorship and what they call residual resoluteness. This last named condition requires the correspondence to choose singletons for some special profiles. Rodríguez-Álvarez (2007) provides an alternative proof and extension of the result.

Ching and Zhou (2002) also consider social choice correspondences. Their definition is also based on interpreting their images as the basis for some lottery, and on the explicit assumption that individuals will be “Bayesian rational”, in the following sense.

“He has a priori a subjective probability measure μ_i over A as well as a von Neumann-Morgenstern utility function u_i over A. Then for any subsets X and Y of A, he can compare X and Y by calculating the expected values of u_i conditional on X and Y: he ranks X over Y if and only if $E_{\mu|X}u_i \geq E_{\mu|Y}u_i$. This actually induces a complete ranking of subsets of A. Of course, this ranking depends on both μ_i and u_i, and information of neither is contained in R_i. Hence, if only R_i is known, then X might be ranked above Y for one particular set of choices of μ_i and u_i, consistent with R_i, but below Y for some different choices of μ_i and u_i also consistent with R_i. Nonetheless, for some pairs of subsets X and Y, such comparisons always lead to the same ranking.”

In fact, they proceed to characterize the cases where the comparisons of sets are unambiguous, in that the ranking of one above the other is independent of the chosen expectation and utility. A subset X of alternatives will always be preferred in that sense to a subset Y if and only if

(i) aR_ib *for all* $a \in X\backslash Y$ *and* $a \in Y$ *if* $X\backslash Y \neq 0$, *and*

(ii) cR_id *for all* $c \in X$ *and* $d \in Y\backslash X$ *if* $Y\backslash X \neq 0$.

After this observation, they define a correspondence to be strategy-proof if it is not unambiguously manipulable by a Bayesian rational voter.

With this definition in hand, they prove that any social choice correspon-