JOURNAL OF ECONOMIC THEORY **21**, 473–482 (1979)

Manipulation and the Pareto Rule*

ALLAN FELDMAN

Department of Economics, Brown University, Providence, Rhode Island 02912

Received September 1, 1978; revised March 5, 1979

This paper examines rules that map preference profiles into choice sets. There are no agendas other than the entire set of alternatives. A rule is said to be "manipulable" if there is a person i, and a preference profile, such that i prefers the choice set obtained when he is dishonest to the one obtained when he is honest. It is "nonmanipulable" if this can never happen. The paper indicates how preferences over choice sets might be sensibly derived from preferences over alternatives, and discusses seven different notions of manipulability associated with seven different assumptions about preferences over sets of alternatives. The paper has two sections of results. In the first I show that the Pareto rule, that is, the rule that maps preference profiles into corresponding sets of Pareto optima, is nonmanipulable in four of the seven senses of manipulability, and manipulable in three of them. In the second section, I examine this conjecture: If an arbitrary rule is nonmanipulable and nonimposed, and if indifference is disallowed, then every choice set must be contained in the set of Pareto optima. The conjecture is true under the strongest definition of nonmanipulability.

I. INTRODUCTION

A voting procedure takes the preferences of voters as its inputs and produces a winning alternative as its output. If some individual can secure a preferred winner by falsifying his preferences, the procedure is manipulable. The question of manipulability has been around for some time [5, 6, 23, 24], and interest in the question has been especially strong recently because of the impossibility theorem of Gibbard and Satterthwaite: The only nonrandom, nondegenerate, nonmanipulable, and single-valued voting procedure is a dictatorship [14, 20, 21].

The question of manipulability also arises for multivalued collective choice rules [1, 2, 12, 17]. While a voting procedure takes individuals' preferences and produces a single winner, a multivalued collective choice

* P. Gärdenfors and D. Schmeidler made helpful comments on an earlier draft of this paper.

0022-0531/79/060473-10$02.00/0

rule takes individuals' preferences and produces a set of alternatives, the choice set, as its output. A simple example is a voting procedure which allows ties for first place. A more complex example is the Pareto rule: it takes the preferences of individuals as inputs, and produces a set of alternatives (namely, the Pareto optima) as its output. It turns out that some useful multivalued collective choice rules are *not* liable to manipulation. In fact, it has already been observed in [17] that the Pareto rule is in one sense nonmanipulable, and that its nonmanipulability is a consequence of the fact that it is multivalued.

This paper is about the senses in which the Pareto rate is manipulable, and the senses in which it is nonmanipulable. It is also about whether or not there is a necessary connection between nonmanipulability and Pareto optimality: Must a nonmanipulable rule always produce Pareto optimal choice sets?

In Section II below, I present the model and examine some of the possible definitions of manipulation for multivalued choice rules. A person manipulates a single-valued rule if by misrepresenting his preferences he secures a single outcome that he prefers to the single outcome when he is honest. A person manipulates a multivalued choice rule if through misrepresentation he secures a set of alternatives that he prefers to the set of alternatives chosen when he is honest. When, therefore, does he prefer one set of alternatives to another set? The question has been considered before [1, 2, 4, 11, 12, 17–19], most thoroughly in [13]. Some of the previous approaches are incorporated in the seven definitions of set preferences used in this paper.

Section III examines the susceptibility to manipulation of the Pareto rule. It turns out that this most important multivalued rule is immune to manipulation in four of the seven senses of manipulation used here. For example, it is immune to manipulation in "maximin" and "maximax" senses: An individual cannot secure a preferred Pareto set if he is only interested in, or attempts to maximize, the worst alternatives for him in the set (maximin behavior), or if he is only interested in, or attempts to maximize, the best alternatives for him in the set (maximax behavior).

In Section IV I look at the question of whether or not there is a *necessary* connection between nonmanipulability, on the one hand, and the Pareto rule, on the other hand. In particular, if an arbitrary collective choice rule is nonimposed and nonmanipulable, must it be contained in the Pareto rule? The answer to this question is a qualified "yes" under the strongest definition of nonmanipulability used here.

This paper is in certain respects similar to other papers that have extended the Gibbard–Satterthwaite impossibility result to multivalued [2, 17] and random [1, 3, 15] choice rules. There is a slightly different bias, however, since I have several *positive* results—e.g., the Pareto rule is in some important ways nonmanipulable. There is also a significant difference between the models of [2, 17] and the model below: In this paper a collective choice rule maps preferences into choice sets, rather than into complete, reflexive, and acyclic binary relations as in [2], and there is only one agenda—the whole set of alternatives—rather than a large class of agendas as in [17]. No consistency-over-agendas assumptions are made here. This paper also differs from [1], which uses the preferences into choice sets framework: Reference [1] assumes strong properties for the choice rule (unanimity and positive responsiveness) that are not assumed here, and it uses a definition of manipulation logically independent of those used here.

II. The Model and Notation

The object of study is a group of n persons, indexed by i or j, who make choices within a finite set X of alternatives. Each person has a preference relation R_i, defined on X, which is assumed to be reflexive, complete, and transitive (that is, an order). P_i is i's associated strict preference relation, and I_i is i's indifference relation. Let $R = (R_1, ..., R_n)$ represent a *preference profile*, that is, a specification of the preferences of all individuals.

A *collective choice rule* $C(\cdot)$ maps preference profiles into nonempty subsets of X. $C(R)$ represents the choice set corresponding to the preference profile R. The most important choice rule in this paper is the Pareto rule.

An alternative x is said to be *Pareto optimal* if there exists no alternative y such that yR_jx for all j and yP_jx for some j. The collective choice rule which maps preference profits into corresponding sets of Pareto optima is called the *Pareto rule.* I will let $P(\cdot)$ represent the Pareto rule, and $P(R)$ the Pareto optimal set corresponding to the profile R.

In this paper attention is focused on the possible actions of a single individual, say person i, and their consequences for him. Basically i can do two things. He can reveal his "true" preferences, or he can conceal them. An unprimed R_i represents i's "true" or "sincere" preferences, while a primed R_i' represents "false" or "insincere" preferences. Whether i is sincere or insincere will affect the choice set. For notational simplicity, I will adopt these conventions:

$$C(R_i) \equiv C(R) = C(R_1, ..., R_{i-1}, R_i, R_{i+1}, ..., R_n)$$

is the "true" choice set; and

$$C(R_i') \equiv C(R_1, ..., R_{i-1}, R_i', R_{i+1}, ..., R_n)$$

is the "false" choice set.

For the Pareto rule, $P(R_i) \equiv P(R)$ is the "true" set of Pareto optima; the optima contingent on i's reporting R_i as his preference relation. $P(R_i')$ is the "false" set of Pareto optima; the optima contingent on i's reporting R_i' as his preference relation.

The question raised below is this: Can i ever *manipulate* the collective choice rule; can he ever ensure a better choice set by being insincere? That is, can i find an R_i' so that he prefers $C(R_i')$ to $C(R_i)$?

In order to make this question unambiguous, it is necessary to be precise about when i prefers one set of alternatives to another. When the sets are singletons, there is no ambiguity for one need only refer to i's strict preference relation P_i. When the sets are not singletons, there are many definitions of "better than" which are consistent with R_i. Let A and B be distinct non-empty subsets of X. I will write AP_i^*B for "i prefers the set of alternatives A to the set of alternatives B." Here are the definitions of P_i^* used in this paper: (The i subscripts have been dropped.)

DEFINITION 1. AP^*B if for all $x \in A$ and $y \in B$, xRy, with strict preference holding for at least one pair.

DEFINITION 2. AP^*B if for all $x \in A - B$, $y \in A \cap B$ and $z \in B - A$, $xPyPz$.

DEFINITION 3. AP^*B if for all $x \in A - B$, $y \in A \cap B$, and $z \in B - A$, $xRyRz$, with strict preference holding for at least one pair.

DEFINITION 4. AP^*B if there exists a $y \in B$ such that xPy for every $x \in A$.

DEFINITION 5. AP^*B if there exists an $x \in A$ such that xPy for every $y \in B$.

The next definitions are in terms of expected utilities from even-chance lotteries. A utility function $u_i(\cdot)$ *represents* a preference relation R_i if $u_i(x) \geqslant u_i(y) \Leftrightarrow xR_iy$, for all x and y in X. A *lottery* $\{p_x\}$ over a set of alternatives A is a probability distribution over A. The *even-chance* lottery assigns all elements of A probability $1/|A|$. Given a representation $u_i(\cdot)$ of R_i, the *expected utility from an even-chance lottery over* A is $(1/|A|) \sum_{x \in A} u_i(x)$.

DEFINITION 6. AP^*B if for every $u(\cdot)$ that represents R, the expected utility from an even-chance lottery over A is greater than the expected utility from an even-chance lottery over B.

DEFINITION 7. AB^*B if there exists a $u(\cdot)$ that represents R, for which the expected utility from an even-chance lottery over A is greater than the expected utility from an even-chance lottery over B.

Definition 1 parallels Kelly's [17] definition of manipulation, and provides what he calls a "clear" case of when one set of alternatives ought to be preferred to another. Definitions 2 and 3 are largely drawn from Brams and Fishburn [4] and Fishburn [11]. When $A = \{x\}$ and $B = \{x, y\}$, or $A = \{x, y\}$ and $B = \{y\}$, both definitions are compelling, since both have $\{x\}P^* \{x, y\}$ if xPy, and $\{x, y\} P^* \{y\}$ if xPy. These connections between P and P^*, which Fishburn calls axioms, are in fact crucial in Kelly [17] and Barberá [1, 2].

Definition 4 is a maximin approach to set comparisons, and is analogous to Pattanaik's [18, 19] definition of preferences over sets. However, when the worsts in the sets A and B are equally bad, Pattanaik's apprach is to compare the second worsts, which Definition 4 does not do. Definition 5 provides a maximax approach to set comparisons. Other definitions akin to 1–5 are clearly possible; one notable one is that of Gärdenfors in [12]. Gärdenfors [13] has a survey of such definitions.

Definitions 6 and 7 are superficially of different character than 1–5, since they make explicit references to expected utilities. Definition 6 is drawn from Gärdenfors [13] and indirectly from Fishburn [9]. In Gibbard [15], preference profiles are mapped into probability distributions over X, and then an approach akin to Definition 7 is made. In Barberá and Sonnenshein [3], preference profiles are mapped into lotteries over social preference relations.

With Definitions 1–7 in hand, I define *manipulation* of a collective choice rule $C(\cdot)$ as follows: $C(\cdot)$ can be *manipulated by i in the sense of Definition* 1, 2, 3, 4, 5, 6, *or* 7, respectively, if there exist a preference profile R and a false preference relation R_i' such that $C(R_i') P_i^* C(R_i)$, where P_i^* is as in Definition 1, 2, 3, 4, 5, 6, or 7, respectively. If, under Definition 1, 2, 3, 4, 5, 6, or 7, respectively, this is impossible for all i and all preference profiles, then $C(\cdot)$ is said to be *nonmanipulable* or *cheatproof*, in the sense of Definition 1, 2, 3, 4, 5, 6, or 7, respectively.

The following connections among definitions of nonmanipulability are easy to establish:

PROPOSITION 1. *$C(\cdot)$ is nonmanipulable in the sense of Definition* 7 $\Rightarrow$ *$C(\cdot)$ is nonmanipulable in the sense of Definition* 6 $\Rightarrow$ *$C(\cdot)$ is nonmanipulable in the sense of Definition* 3 $\Rightarrow$ *$C(\cdot)$ is nonmanipulable in the sense of Definition* 2 $\Rightarrow$ *$C(\cdot)$ is nonmanipulable in the sense of Definition* 1.

PROPOSITION 2. *$C(\cdot)$ is nonmanipulable in the senses of Definitions* 4 *and* 5 $\Rightarrow$ *$C(\cdot)$ is nonmanipulable in the sense of Definition* 2.

III. MANIPULATION OF THE PARETO RULE

This section examines the Pareto rule, to determine in what senses it is immune to manipulation, and in what senses it is not. I will show that the Pareto rule is nonmanipulable in the sense of Definition 1, a result alluded to by Kelly [17]. Also, the Pareto rule is nonmanipulable in the sense of Definition 2. That is, it is nonmanipulable under the first extension of the Fishburn axioms. However, the Pareto rule *is* manipulable in the sense of Definition 3; that is, it is manipulable under the second extension of the Fishburn axioms. By Proposition 1, the Pareto rule must also be manipulable in the senses of expected utility comparisons over even-chance lotteries, i.e., Definitions 6 and 7.

I will also show that the Pareto rule is nonmanipulable under the maximin and maximax definitions of preferences over sets, Definitions 4 and 5. Therefore, if an individual is concerned only with the *worst* alternatives in the choice set, he cannot manipulate the Pareto rule. Or, if he is concerned only with the *best* alternatives in the choice set, he cannot manipulate the Pareto rule. So $P(\cdot)$ stands up well under either of these extreme definitions of set preferences.

THEOREM 1. *Under Definition* 1, *the Pareto rule is nonmanipulable.*

Proof. This will follow from Proposition 1 and Theorem 2, below.

THEOREM 2. *Under Definition* 2, *the Pareto rule is nonmanipulable.*

Proof. This will follow from Proposition 2, and Theorems 3 and 4 below.

THEOREM 3. *Under Definition* 4, *the Pareto rule is nonmanipulable.*

Proof. Suppose to the contrary that there is an $R = (R_1, R_2, \ldots, R_n)$ and an R_i' such that $P(R_i')\ P_i^*\ P(R_i)$, under Definition 4. Then there is an alternative y_0 in $P(R_i)$ such that xP_iy_0 for any x in $P(R_i')$.

Obviously, y_0 cannot also be in $P(R_i')$. Consequently, there is an $x_0 \in P(R_i')$ such that $x_0R_jy_0$ for all $j \neq i$, $x_0R_i'y_0$, and the strict preference holds for someone. Also, since $x_0 \in P(R_i')$, $x_0P_iy_0$.

Since $y_0 \in P(R_i)$, $x_0R_jy_0$ for all $j \neq i$ and $x_0P_iy_0$ is impossible. This is a contradiction. Q.E.D.

THEOREM 4. *Under Definition* 5, *the Pareto rule is nonmanipulable.*

Proof. It is clear that $P(R_i)$ must contain some of i's favorite alternatives. For if y is one of i's best, and if y is not Pareto optimal, there exists an x in $P(R_i)$ for which xR_iy; so x is also one of i's best. Consequently, $P(R_i')\ P_i^*\ P(R_i)$ under Definition 5, the maximax definition, is impossible. Q.E.D.

The following examples show the Pareto rule is manipulable in the Definition 3, 6, and 7 senses. Example 1, which involves ties in the individuals' rank orders (or indifference), establishes manipulability in the three senses. Example 2 establishes manipulability in the Definition 6 and 7 senses without recourse to indifference.

EXAMPLE 1. There are two individuals with the following preferences:

$$1: x_1(x_2x_3) \qquad 1': x_1x_2x_3$$
$$2: (x_2x_3)\, x_1 .$$

(That is, according to 1's true preference relation, shown by 1: etc., he prefers x_1 to x_2 and x_3, and is indifferent between x_2 and x_3. According to his false preference relation, shown by 1′: etc., he prefers x_1 to x_2 to x_3. Person 2 is indifferent between x_2 and x_3, but prefers both to x_1.)

Now $P(R_1) = \{x_1, x_2, x_3\}$, while $P(R_i') = \{x_1, x_2\}$. Consequently, $P(R_1') - P(R_1) = \varnothing$, $P(R_1') \cap P(R_1) = \{x_1, x_2\}$, and $P(R_1) - P(R_1') = \{x_3\}$, so $P(R_1')\, P_1^*\, P(R_1)$ according to Definition 3.

It is also necessarily the case that $P(R_1')\, P_1^*\, P(R_1)$ according to the expected utility Definitions 6 and 7. Consequently, individual 1 can manipulate in the senses of Definitions 3, 6, and 7.

EXAMPLE 2. There are two individuals with the following preferences:

$$1: x_1x_2x_3x_4 \qquad 1': x_2x_1x_4x_3$$
$$2: x_4x_3x_1x_2 .$$

Now $P(R_1) = \{x_1, x_3, x_4\}$, while $P(R_1') = \{x_1, x_2, x_4\}$. If $u_1(\cdot)$ is any utility function that represents R_1,

$$\tfrac{1}{3}u_1(x_1) + \tfrac{1}{3}u_1(x_2) + \tfrac{1}{3}u_1(x_4) > \tfrac{1}{3}u_1(x_1) + \tfrac{1}{3}u_1(x_3) + \tfrac{1}{3}u_1(x_4).$$

Consequently, $P(R_1')\, P_1^*\, P(R_1)$ by Definitions 6 and 7, and individual 1 can manipulate in the senses of Definitions 6 and 7.

These theorems and examples establish some limits to the manipulability, or nonmanipulability, of the Pareto rule. For maximiners and for maximaxers, the Pareto rule is not manipulable. The Pareto rule is not "clearly" manipulable in Kelly's sense. It is nonmanipulable in one Fishburn sense, and manipulable in another. In terms of expected utilities, it *is* manipulable.

IV. ARBITRARY NONMANIPULABLE COLLECTIVE CHOICE RULES

In the section above I asked the question: If $C(\cdot)$ is the Pareto rule, must it be nonmanipulable? In this section I turn the question on its head. If an arbitrary rule $C(\cdot)$ is nonmanipulable, must it be related to the Pareto rule? This question parallels the theorems of Gibbard [14] and Satterthwaite [20], which show single-valued nonmanipulable rules must be dictatorial, and it also closely parallels the theorems of Barberá [2] and Kelly [17], which show multivalued nonmanipulable rules satisfying certain regularity conditions must be weakly dictatorial.

To ease the analysis in the rest of the paper, I will assume the domain of $C(\cdot)$ is restricted to "strict" or "linear" preferences. That is, all preference relations are *antisymmetric*: For all x, y and R_i, xR_iy and $yR_ix \Rightarrow x = y$. Also, to avoid trivial collective choice rules, I will make a modest nonimposition assumption. The collective choice rule $C(\cdot)$ is *nonimposed* if, for any $x \in X$, there exists a strict preference profile R such that $C(R) = \{x\}$.

The question is: Does nonmanipulability imply something about Pareto optimality? To be more specific, I consider the following:

Conjecture. Suppose an arbitrary collective choice rule $C(\cdot)$ is nonmanipulable (in some sense), nonimposed, and its domain is restricted to strict preferences. Then $C(\cdot) \subset P(\cdot)$; that is, $C(R) \subset P(R)$ for all R.

This is an attractive conjecture, for it would mean that, with little baggage in terms of buttressing assumptions, nonmanipulability for a choice rule implies Pareto optimality for all chosen alternatives. Unfortunately, it is not right, at least for most definitions of nonmanipulability. The following example shows why.

EXAMPLE 3. Suppose there are two people and three alternatives $\{x_1, x_2, x_3\}$. Under the strict preferences domain restriction, there are 36 possible preference profiles, which the reader can chart in a 6×6 matrix. (See also [21].) Let the collective choice rule be given by:

$$\begin{aligned} C(\cdot) &= \text{The first choice of both, if they agree about first} \\ &= \text{The first and second of both, if they agree only about third} \\ &= \{x_1, x_2, x_3\} \text{ otherwise.} \end{aligned}$$

I call this the *semioptimal rule*.

The reader can easily convince himself that, in the 2-person, 3-alternative strict preferences case, the semioptimal rule is *nonmanipulable in the Definition* 1, 2, 5, *and* 6 *senses*. (Note that Definition 3 collapses into Definition 2 under strict preferences.)

But is the semioptimal rule contained in the Pareto rule? In fact, it is not. For example, if the preference profile R is given by

$$\begin{aligned} 1&: x_1x_2x_3 \\ 2&: x_3x_1x_2, \end{aligned}$$

then $P(R) = \{x_1, x_3\}$ while $C(R) = \{x_1, x_2, x_3\}$. Therefore, the *conjecture is wrong for Definitions* 1, 2, 5, *and* 6 of nonmanipulability.

Under the Definition 4 (maximin) version of nonmanipulability, the conjecture is true for $n \leqslant 3$ and $|X| \leqslant 3$, but false for arbitrary n and $|X|$.

However, under the strong expected utility definition, that is, *under Definition 7, the conjecture is correct*. In other words, if no individual can ever increase his expected utility from an even-chance lottery over the choice set by misrepresenting his preferences, then the choice set must always be contained in the Pareto optimal set.

THEOREM 5. *Suppose an arbitrary collective choice rule is nonmanipulable in the Definition 7 sense, nonimposed, and its domain is restricted to strict preferences. Then $C(\cdot) \subset P(\cdot)$; that is, $C(R) \subset P(R)$ for all R.*

This theorem follows easily from Gibbard's "weak version" theorem in [15], or from his Corollary 1 in [16].

Unfortunately, the powerful notion of nonmanipulability of Definition 7 does *too* much. For it turns out that any $C(\cdot)$ satisfying the assumptions of Theorem 5 must be either a dictatorship or a duumvirate. (See [8].) In the context of the model of this paper, then, if the assumption of nonmanipulability is strengthened enough to ensure that nonmanipulability implies optimality, nonmanipulability will imply dictatorial or duumviral rule as well.

A different model, in which the agendas can be any subsets of the whole set of alternatives, in which manipulation is defined only in terms of 1- and 2-element agendas, and in which a contraction consistency property for the choice rule is met, does generate a theorem like Theorem 5, without the unfortunate concomitant dictatorship or duumvirate. (See Theorem 8 of [7].)

REFERENCES

1. S. BARBERÁ, The manipulation of social choice mechanisms that do not leave "to much" to chance, *Econometrica* **45** (1977), 1573–1588.
2. S. BARBERÁ, Manipulation of social decision functions, *J. Econ. Theory* **15** (1977), 266–278.
3. S. BARBERÁ AND H. SONNENSCHEIN, "Mixture Social Welfare Functions," Princeton University Econometric Research Program Memorandum No. 215, 1977.
4. S. J. BRAMS AND P. C. FISHBURN, Approval voting, *Amer. Pol. Sci. Rev.*, forthcoming.
5. A. DUMMETT AND R. FARQUHARSON, Stability in voting, *Econometrica* **29** (1961), 34–44.
6. R. FARQUHARSON, "Theory of Voting," Yale Univ. Press, New Haven, Conn., 1969.
7. A. M. FELDMAN, Nonmanipulable multi-valued social decision functions, *Public Choice* **34** (1979), 39–50.
8. A. M. FELDMAN, Strongly nonmanipulable multi-valued collective choice rules, *Public Choice*, forthcoming.
9. P. C. FISHBURN, Even-chance lotteries in social choice theory, *Theory and Decision* **3** (1972), 18–40.
10. P. C. FISHBURN, Lotteries and social choices, *J. Econ. Theory* **5** (1972), 189–207.

11. P. C. Fishburn, A strategic analysis of non-ranked voting systems, mimeo, 1977.
12. P. Gärdenfors, Manipulation of social choice functions, *J. Econ. Theory* **13** (1976), 217–228.
13. P. Gärdenfors, On definitions of manipulation of social choice functions, *in* "Aggregation and Revelation of preferences" (J.-J. Laffont, Ed.), North-Holland, Amsterdam, 1979.
14. A. Gibbard, Manipulation of voting schemes: A general result, *Econometrica* **41** (1973), 587–601.
15. A. Gibbard, Manipulation of schemes that mix voting with chance, *Econometrica* **45** (1977), 665–681.
16. A. Gibbard, Straightforwardness of game forms with lotteries as outcomes, *Econometrica* **46** (1978), 595–614.
17. J. S. Kelly, Strategy-proofness and social choice functions without singlevaluedness, *Econometrica* **45** (1977), 439–446.
18. P. K. Pattanaik, On the stability of sincere voting situations, *J. Econ. Theory* **6** (1973), 558–574.
19. P. K. Pattanaik, Stability of sincere voting under some classes of non-binary group decision procedures, *J. Econ. Theory* **8** (1974), 206–224.
20. M. A. Satterthwaite, Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions, *J. Econ. Theory* **10** (1975), 187–217.
21. D. Schmeidler and H. Sonnenschein, Two proofs of the Gibbard-Satterthwaite theorem on the possibility of a strategy-proof social choice function, *in* "Decision Theory and Social Ethics" (H. W. Gottinger and W. Leinfellner, Eds.), Reidel, Dordrecht, 1978.
22. A. K. Sen, "Collective Choice and Social Welfare," Holden–Day, San Francisco, 1970.
23. W. Vickrey, Utility, strategy, and social decision rules, *Quart. J. Econ.* **74** (1960), 507–535.
24. R. Zeckhauser, Voting systems, honest, preferences and Pareto optimality, *Amer. Pol. Sci. Rev.* **67** (1973), 934–946.