Muller, Eitan; Satterthwaite, Mark

Working Paper

Strategy-Proofness: The Existence of Dominant Strategy Mechanisms

Discussion Paper, No. 568

Provided in Cooperation with:
Kellogg School of Management - Center for Mathematical Studies in Economics and Management Science, Northwestern University

Suggested Citation: Muller, Eitan; Satterthwaite, Mark (1983) : Strategy-Proofness: The Existence of Dominant Strategy Mechanisms, Discussion Paper, No. 568, Northwestern University, Kellogg School of Management, Center for Mathematical Studies in Economics and Management Science, Evanston, IL

This Version is available at:
https://hdl.handle.net/10419/220928

Discussion Paper No. 568

STRATEGY-PROOFNESS: THE EXISTENCE OF DOMINANT STRATEGY MECHANISMS

by
Eitan Muller
and
Mark Satterthwaite
Department of Managerial Economics and Decision Sciences
Northwestern University
July 1983

to appear in
Social Goals and Social Organization
Essays in Memory of Elisha A. Pazner,
Edited by Hurwicz, Schmeidler and Sonnenschein

STRATEGY-PROOFNESS: THE EXISTENCE OF DOMINANT STRATEGY MECHANISMS

by

Eitan Muller

and

Mark A. Satterthwaite

1. INTRODUCTION

Economic theory takes as axiomatic that individuals have preferences over possible allocations and that they seek their most preferred allocation. Except in unusual and happy circumstances the result is conflict: the several agents disagree over which outcomes are preferable and they resolve their conflict within the rules of whatever allocation mechanism under which they happen to be operating. Since the outcome is important each agent devises a strategy that he believes will be effective in securing, as nearly as possible, an outcome that is highly preferred by his own lights.

This penchant that individuals have for strategizing causes economic theorists trouble because the essence of an individual's strategic choice is to correctly guess the actions of other individuals and to then choose the action that results in the best attainable outcome. This means that the properties of a particular allocation mechanism can not be determined in any

simple way. Specifically, an allocation mechanism might be thought to operate by asking agents to state their preferences and then calculating from this information an outcome that meets an appropriate optimality criterion. Strategic behavior confounds this process because an individual may calculate, given the probable strategies of other agents, that misrepresenting his preferences may result in a more preferred outcome than stating them truthfully. Therefore in studying the properties of a particular allocation mechanism the theorist must not only understand how the mechanism aggregates the information individuals input into it, he must also model the information each agent has about every other agent and how each agent uses this information to decide what information to input into the mechanism. This is difficult.

Strategy-proof mechanisms represent the most direct and elegant means conceivable for cutting through the problems that strategic behavior poses for our understanding of allocation mechanisms' performance. An allocation mechanism is defined to be strategy-proof if and only if telling the truth is always a dominant strategy for every agent. A strategy is dominant for an agent if, irrespective of what strategies the other agents play, no other strategy results in an outcome that the agent prefers. An agent who has a dominant strategy need not guess what other agents are likely to do because that guessing has no utility; the agent's dominant strategy is best no matter what other agents do. Therefore for strategy-proof mechanisms the question of strategy never arises because every agent has no reason not to follow the dominant strategy of truth telling. This makes the analysis of strategy-proof mechanisms trivial in comparison to the analysis of mechanisms that are not strategy-proof because questions of the information that agents possess about other agents can be ignored.

This paper's purpose is to explore the current state of our knowledge concerning the possibilities for constructing strategy-proof mechanisms. We focus on strategy-proof mechanisms rather than dominant strategy mechanisms because every dominant strategy mechanism is equivalent to some strategy-proof mechanism. Consequently no generality is lost by our focus on strategy-proofness rather than dominant strategies. Section 2 of the paper presents an important, base line result: the Gibbard-Satterthwaite Theorem. It states that reasonable strategy-proof allocation mechanisms, while exceedingly attractive in the abstract, simply do not exist when agents' admissible preferences over the set of feasible alternatives are not a priori restricted to some subset of the set of all possible transitive orderings of the feasible alternatives. Thus, in the most general case, strategizing can not be taken out of economic behavior by cleverly designing the allocation mechanism.

The remainder of the paper explores the degree to which the general case must be specialized in order to make the construction of a reasonable strategy-proof mechanism feasible. We pursue two approaches to this problem. In Section 3 we specify with increasing precision what we mean by a reasonable strategy-proof mechanism and then investigate how tightly the set of admissible preference orderings must be restricted in order to make construction of the specified mechanism possible. In Section 4 we reverse the procedure. There we specify restrictions on the set of a priori admissible preference orderings in ways that have economic relevance and then ask what reasonable strategy-proof mechanisms can be constructed given those particular restrictions on domains. Conceptually these two approaches are dual to each other; in practice, however, no one has succeeded in making an adequate formal connection between them. Therefore we present them seperately.

No completely unambiguous conclusion can be drawn from the work discussed

in this paper. As reported in Section 4, for several specific domains of admissible preferences the results are negative in that no reasonable strategy-proof mechanism can be constructed. In Section 3 we report results that show the existence of domains that (i) are large relative to the size of the unrestricted domain and (ii) do permit construction of reasonable strategy-proof mechanisms. Nevertheless no examples have as yet been constructed that succeed in showing that these relatively large restricted domains have relevance to the types of restrictions on admissible preferences that naturally occur in economics.

This paper is not a survey. We only report on a small fraction of the interesting work that has been done on the existence of strategy-proof mechanisms. We have tried to present some essential ideas from this body of research in a manner that contributes to the reader's intuition and understanding.

2. PROBLEM FORMULATION AND A BASIC THEOREM

Basic Model. Most of the work on strategy-proof mechanisms has been conducted in a very simple framework that focuses on agents' preferences and the incentives they have may have to follow dominant strategies in revealing those preferences.[1] A group $I = \{1, 2, \ldots, n\}$ is a fixed set of n individuals who must select an alternative from a feasible set of alternatives. The set $A = \{x, y, z, \ldots, w\}$ is the set of all conceivable resource allocations; it has cardinality of $|A|$. Each individual $i \in I$ has a transitive binary preference relation P_i over the set A. Thus, for all pairs of alternatives $x,y \in A$ and for every individual $i \in I$, one of three cases is true: xP_iy denoting strict preference for x over y, yP_ix denoting strict preference for y over x, or neither xP_iy nor yP_ix denoting indifference between x and y. Indifference

between x and y is alternatively denoted by $x\tilde{P}_i y$.

Not every preference ordering is necessarily admissible. Let Ω be the set of all complete and transitive preference orderings P_i that any individual i might rationally hold. In other words, if $P_i \notin \Omega$, then P_i is a preference ordering that, while being transitive, violates some principle of rationality that clearly applies to the situation in question. For example, in economic contexts if the two-dimensional vector x represents a commodity bundle and that bundle dominates both components of a second bundle y, then the principle of nonsatiation implies that an ordering P_i for which xP_iy may be admissible (and thus be an element of Ω) while an ordering P_i for which yP_ix can not be admissible. The set Ω^n is the n-fold cartesian product of Ω. The group's preference profile is the n-tuple, $(P_1,\ldots,P_n) \in \Omega^n$, of the individual orderings.

The set of feasible allocations, B, may be either A in its entirety or a subset of it. The group's task is to select a single allocation from B. They do this, in effect, by voting. Each individual i reports a preference ordering $Q_i \in \Omega$ for input into the allocation mechanism F that aggregates the profile of reported preferences down to a single element of B. Formally, let $\underline{A}$ be the set of subsets of A. An allocation mechanism is a function F: $\Omega^N \times \underline{A} \to A$. Thus F(Q, B) is the group's choice when the profile of reported preferences is Q and the feasible set is B. The preference ordering Q_i an individual reports may or may not be identical to his preferences P_i; the choice of what to report is his since preferences are private and impossible for outsiders to ascertain.

That individuals can not be forced to report their preferences P_i sincerely for input into the allocation mechanism is the crux of the problem this paper considers. Each individual agent may calculate whether it is in

his or her interest to report honestly. An agent i with preferences P_i has an incentive to manipulate the mechanism F at profile $P/P_i \in \Omega^n$ and feasible set $B \in \underline{A}$ if

$$F(P/Q_i, B)\ P_i\ F(P/P_i, B) \tag{2.01}$$

where $Q_i \in \Omega$, $(P/Q_i) = (P_1,\ldots,Q_i,\ldots,P_n) \in \Omega^n$, and $(P/P_i) = P = (P_1,\ldots, P_i,\ldots, P_n)$. The content of (2.01) is that if agent i is to be able to manipulate the outcome at profile $P = P/P_i$, then he must have available an admissible ordering Q_i that, when played as a substitute for his true preferences P_i, results in an outcome he strictly prefers.

Dominance and Strategy-proofness. A mechanism F is strategy-proof if no admissible profile $P \in \Omega^n$, no feasible set $B \in \underline{A}$, and no agent i exists such that at profile P agent i can manipulate mechanism F. Individuals never have an interest in not reporting their preferences accurately when the mechanism is strategy-proof. An implication is that if a mechanism is strategy-proof, then every agent always has a dominant strategy. Formally, a strategy $Q_i \in \Omega$ is dominant at feasible set $B \in \underline{A}$ for agent i with preferences P_i if no profile $P/Q_i' \in \Omega^n$ exists such that

$$F(P/Q_i', B)\ P_i\ F(P/Q_i, B). \tag{2.02}$$

In other words, the ordering Q_i is dominant for agent i if and only if no profile exists for which playing another ordering Q_i' would result in the realization of a strictly preferred outcome for agent i. A mechanism, F, is a dominant strategy mechanism if, at every $P \in \Omega^n$ and $B \in \underline{A}$, every agent has a dominant strategy.

The great attraction of dominant strategy mechanisms is that agents need no information about other agents' preferences in order to play optimally. Suppose F is not a mechanism for which agents have dominant strategies. Inspection of (2.02) shows that if agent i is to successfully manipulate

mechanism F at profile P, then he or she must know that profile P/Q_i is being realized rather than, for example, profile P'/Q_i. To know this requires good information on i's part about other agents' preferences and strategizing. None of this information is needed if F always gives individual i a dominant strategy. No matter what profile is realized he plays that ordering Q_i that is dominant for his true preference ordering P_i.

The set of all possible strategy-proof mechanisms is clearly a subset of the set of mechanisms that always give every agent a dominant strategy. We restrict ourselves to considering only strategy-proof mechanisms because, as Gibbard (1973) showed, every dominant strategy mechanism that is not strategy-proof is equivalent to a strategy-proof mechanism. No generality is gained by looking at the broader class. This equivalence is seen as follows. Suppose F is a dominant strategy mechanism. Therefore for each agent i a function σ_i: $\Omega \to \Omega$ exists that associates his true preference ordering P_i with his dominant strategy Q_i for that particular ordering, i.e., $\sigma_i(P_i) = Q_i$. Define a new mechanism, F^σ, as the composition of the F and σ functions:

$$F^\sigma(P_1,\ldots,P_n, B) = F[\sigma_1(P_1),\ldots, \sigma_n(P_n), B]. \tag{2.03}$$

The mechanism F^σ is strategy-proof. If, contrary to the assertion, it were not strategy-proof, then an agent i, a profile $P \in \Omega^n$, a feasible set $B \in \underline{A}$, and an ordering $Q_i' \in \Omega$ would exist such that i would have an incentive to manipulate F^σ:

$$F^\sigma(P/Q_i', B)\ P_i\ F^\sigma(P/P_i,B). \tag{2.04}$$

Relation (2.04) may be rewritten in terms of the original mechanism F:

$$F(Q/Q_i'', B)\ P_i\ F(Q/Q_i, B) \tag{2.05}$$

where $Q = [\sigma_1(P_1),\ldots, \sigma_n(P_n)]$ is the vector of the agents' dominant strategies, $Q_i = \sigma_i(P_i)$ is agent i's dominant strategy when his preferences

are P_i, and $Q_i'' = \sigma_i(Q_i')$ is agent i's dominant strategy when his preferences are Q_i'. But (2.05) contradicts the hypothesis that $Q_i = \sigma_i(P_i)$ is a dominant strategy for agent i because he does better playing Q_i''. Therefore F^σ is a strategy-proof mechanism because if it were not, then F would not be a dominant strategy mechanism as initially assumed. Finally, in addition to being strategy-proof, F^σ is equivalent to F for agent i because if, in utilizing each mechanism, every agent always plays his dominant strategy, then, for any preference profile, F and F^σ give identical payoffs.[2]

Impossibility Theorem. Can strategy-proof mechanisms be constructed? Certainly, inasmuch as we can easily identify four general types:

1. Let, for all admissible profiles $P \in \Omega^n$, $F(P, B) = x$ where x is a fixed element of A. This is an imposed mechanism. It is strategy-proof because, since agents' preferences do not influence the outcome, each agent has nothing to gain from misrepresenting his preferences.

2. Let, for some i, all admissible sets $B \in \underline{A}$, and all admissible profiles $P \in \Omega^n$, $F(P, B) = \max_B(P_i)$ where $\max_B(\cdot)$ picks the element of B that is maximal according to the ordering P_i.[3] This is a dictatorial mechanism where agent i is the dictator. It is strategy-proof because agent i gets his most preferred alternative if he reports his preferences truthfully and no other agent has any influence on the decision.

3. Let A consist of only two elements, {x, y}, and define F to be majority rule: select y if the number of agents i for whom yP_ix exceeds the number for whom xP_iy and x otherwise (including ties). This is strategy-proof because, with only two alternatives, voting against one's preferred alternative can lead to it losing and can

not lead to it winning.

4. Let A = {x, y, z} and let the set of admissible preference orderings consist of two orderings: Ω = {(xzy), (yzx)}. The notation (xzy) stands for the ordering xP_iz, xP_iy, and zP_iy. If the feasible set is the full set A, define F to be majority rule as before except that z is selected in the case of a tie between x and y. If the feasible set is just two elements, define F to be majority rule as in the previous example. This, too, is a strategy-proof mechanism because, with Ω restricted to two elements, the addition of the third alternative z changes nothing essential.

The first two of these mechanisms are unsatisfactory because they do not give sufficient scope for each agent's preferences to affect the choice. The second two mechanisms are unsatisfactory because they only apply to restricted situations: two alternatives in the case of (3) and a severely restricted set of admissible preferences in the case of (4).

Therefore the real question is: Do strategy-proof mechanisms exist that can accomodate any size feasible set, gives agents' preferences an opportunity to affect the group's choice, and apply to a broad class of preference profiles? These three requirements are easily formalized. The first is simple: feasible sets of three or more elements should be admissible. Second, a mechanism should give agents influence over the outcome at least to the extent of satisfying the unanimity requirement of the Pareto principle and being nondictatorial. A mechanism F satisfies the Pareto criterion if, for any set $B \in \underline{A}$, for any profile $P \in \Omega^n$, and for any $x, y \in B$, xP_iy for all $i \in I$ implies $F(P, B) \neq y$. It is strongly nondictatorial if no agent i exists such that, for at least one feasible set $B \in \underline{A}$ ($|B| \geq 2$), $F(P, B) = \max_B(P_i)$ for all $P \in \Omega^n$. Finally, let $\Sigma_{\sim}$ be the set of all possible complete and

transitive orderings that are defined on the conceivable set A. A somewhat narrower, but still very broad set is the set of all possible complete and transitive orderings that are strict, i.e., indifference is excluded. We denote this set by Σ. Therefore, for a mechanism F to be maximally flexible and applicable, setting Ω equal to either $\Sigma_{\sim}$ or Σ is desirable.

This set of requirements is impossible to meet. Gibbard (1973) and Satterthwaite (1973, 1975) showed this basic impossibility result.

THEOREM 2.1 (Gibbard-Satterthwaite Theorem). If $|A| \geq 3$ and preferences are unrestricted ($\Omega = \Sigma_{\sim}$ or Σ), then an allocation mechanism F can not simultaneously be strategy-proof and satisfy both the Pareto criterion and strong nondictatorship.

Feldman (1979) has devised a simple proof of the Theorem for the special case of three alternatives, two agents, and domain Σ^2. We present his proof here because its construction yields insight into how the conditions of Theorem 2.1 may be modified in order to obtain possibility results.

The proof is this. The mechanism F is defined for the set A $= \{x, y, z\}$ and on the domain Σ^2. Table 2.1 shows the restrictions that the the Pareto criterion imposes on F when the feasible set is A. For example, if agent one has preferences (xyz) and agent two has preferences (zxy), then F can not select y because to do so would violate the Pareto criterion. Note that, because F is strategy-proof and thus induces truthful revelation, we need not make a distinction between reported preferences and true preferences. If both report (xyz), then the Pareto criterion requires selection of x. An entry that is a "?" indicates that the Pareto criterion places no restrictions on which alternative is selected.

The mechanism F is single-valued. Therefore a single element of A must be assigned to each cell that does not have a determinate element. Suppose

element x is assigned to the cell labeled 1 (as indicated by the superscript 1). This violates neither the Proposition nor the Pareto criterion. This assignment, however, implies that agent one is a dictator when the feasible set is A. We see this as follows.

Assigning x to cell 1 implies that x must be assigned to cell 2. Suppose to the contrary that the only other possibility, z, were assigned to cell 2. Agent one would then have an incentive to manipulate profile [(xyz), (zxy)] by reporting (xzy) instead of (xyz). That would give him the preferred outcome of x rather than z. Therefore x must be assigned to cell 1 because to assign z to it would be to violate strategy-proofness. This same logic can be used to fill every indeterminate cell on Table 1.

Table 2.2 reports this logic for all cells above the diagonal. Consider as an example the assignment of y to cell 11. Since the Proposition rules x out as a possibility for cell 11, the only alternative outcome that could have been assigned to it is z. If, however, z were assigned, then agent two could manipulate F at the profile [(yzx), (zyx)] ≡ (4,6) by playing the manipulative strategy (zxy):

$$\{F[(yzx), (zxy), A]=z\} \; P_2 \; \{F[(yzx), (zyx), A]=y\}. \tag{2.06}$$

In the notation of Table 2.2 where each of the six orderings of A are assigned an integer label, (2.06) becomes

$$\{F(4,5) = z\} \; P_2 \; \{F(4,6) = y\}. \tag{2.07}$$

The assignment of outcome y to F[(yzx), (zyx), A] was made on the previous line of Table 2.2. Therefore z can not be assigned to cell 11, which leaves y as the sole possibility.

Filling in each indeterminate cell in this manner, both above and below the diagonal, results in agent one being a dictator for F(•, A) and therefore completes Feldman's proof. If, at the beginning, for cell 1 we had assigned

alternative z instead of alternative x, then agent two would have ended up as F's dictator.

Comment. Theorem 2.1 is a negative result. The remainder of this paper is concerned almost exclusively with how Theorem 2.1's conditions can be relaxed in order to obtain existence of a strategy-proof mechanism rather than nonexistence. Examination of the theorem's conditions shows immediately that only one condition--the assumption of unrestricted preferences--can sensibly be relaxed. Nondictatorship and the Pareto criterion are minimal conditions on how power should be distributed among the agents. If anything, they should be strengthened, not weakened. The definition of a voting mechanism can not be relaxed in any obvious way.[4] The number of alternatives that the mechanism can handle certainly must be maintained at three or more.

Theorem 2.1 applies only when preferences are unrestricted, i.e., $\Omega = \Sigma$. Within Feldman's proof if admissible preferences are restricted by, for example, excluding the ordering (zyx) from Ω, then the rightmost column and the bottom row are struck from Table 2.1 because the mechanism would not have to be defined for profiles that involve the ordering (zyx). But striking column six affects the construction of Table 2.2. Our demonstration that cell 11 must be filled with alternative y depended on cell 10, which is in column six, being filled with alternative y in the proof's previous step. Column six's presence is essential for this argument. If enough rows and columns are struck, then the chain of inference that we constructed in Table 2.2 may break causing existence rather than nonexistence.

Relationship with Arrow's Impossibility Theorem. Strategy-proof allocation mechanisms are intimately related to the social welfare functions about which Arrow (1963) proved his famous impossibility theorem. In order to understand the conditions under which reasonable strategy-proof mechanisms

exist one must understand the basics of this relationship. A social welfare function for A is a singlevalued function f that maps the set Ω^n of admissible preference profiles into the set Σ (or $\Sigma_\sim$) of transitive orderings of A. Thus f: $\Omega^n \to \Sigma$. In other words, a social welfare function orders the set A, presumably from best to worse. Associated with every social welfare function f is an allocation mechanism: $F_f(P, B) = \max_B[f(P)]$. If an allocation mechanism F has associated with it a social welfare function, then F is a rational allocation mechanism. Such a mechanism F_f earns this title because it selects that element of B that the social welfare function f ranks highest. Clearly not every allocation mechanism is rational.

Arrow investigated the existence of social welfare functions f whose associated rational allocation mechanisms F_f satisfy the Pareto criterion, weak nondictatorship, and two additional conditions, independence of irrelevant alternatives and monotonicity. A mechanism F satisfies weak nondictatorship if no agent $i \in I$ exists such that, for all feasible sets $B \in \underline{A}$, $F(P, B) = \max_B(P_i)$ for all $P \in \Omega^n$. Contrast this with strong dictatorship where an individual is classified a dictator if he is dictator over even a single feasible set B ($|B| \geq 2$) while here he is classified a dictator only if he is dictator over every feasible set.

A mechanism satisfies independence of irrelevant alternatives (IIA) if whenever any two profiles, $P, Q \in \Omega^n$, agree on the feasible set $B \in \underline{A}$, then $F_f(P, B) = F_f(Q, B)$. Profiles P and Q agree on B if, for all agents i and for all pairs of allocations $(x, y) \in B \times B$, xP_iy if and only if xQ_iy. Independence means that agents' preferences over the feasible set should be the only determinant of the group's choice; preferences over the feasible set's complement should be irrelevent.

To define monotonicity, let $B \in \underline{A}$ be a feasible set, let $x \in B$ be an

allocation within the feasible set, and let $C = B-x$ be the feasible set less the element x. The mechanism satisfies monotonicity if whenever (i) two profiles $P, Q \in \Omega^n$ agree on C and (ii) xP_iy implies xQ_iy for all $y \in C$, then $F_f(P, B) = x$ implies $F_f(Q, B) = x$. Monotonicity means that if one or more agents move a feasible allocation x up in their preference orderings relative to other feasible allocations, then that can not cause x to be dropped as the group's choice. Rational choice on the part of individuals obeys both of these conditions and as such they are reasonable requirments to place on group choice.[5]

Exactly as in Theorem 1.2, Arrow's conditions are impossible to meet when A contains at least three elements and preferences are unrestricted.

THEOREM 1.2 (Arrow's Theorem). If $|A| \geq 3$ and preferences are unrestricted ($\Omega = \Sigma_{\sim}$ or Σ), then a social welfare function f and its associated allocation mechanism F_f can not simultaneously satisfy the Pareto criterion, weak nondictatorship, independence of irrelevant alternatives, and monotonicity.

Social welfare functions that satisfy Arrow's requirements are inextricably intertwined with strategy-proof allocations mechanisms. If preferences are unrestricted and a social welfare function with its associated allocation mechanism satisfy IIA and monotonicity, then the mechanism is strategy-proof.[6] This permits Theorem 2.2 (Arrow) to be proved directly from Theorem 2.1 (Gibbard-Satterthwaite). Specifically, for the case of $|A| \geq 3$ and unrestricted preferences, suppose that--contrary to Arrow's theorem--a social welfare function exists that satisfies the Pareto criterion, nondictatorship, independence of irrelevant alternatives, and monotonicty. Then the associated rational allocation mechanism is strategy-proof. This, however, is impossible because no strategy-proof allocation mechanism (whether

rational or not) exists that satisfies the Pareto criterion and non-dictatorship for the case of $|A| \geq 3$ and unrestricted preferences. Therefore Arrow's Theorem is true.

In the opposite direction, if preferences are unrestricted and an allocation mechanism is rational and strategy-proof, then it also satisfies independence of irrelevant alternatives and positive association.[7] This result together with Arrow's Theorem can be used to show directly that, for the case of $|A| \geq 3$ and unrestricted preferences, no rational, strategy-proof allocation mechanism exists that satisfies the Pareto criterion and weak nondictatorship. This nonexistence result concerning rational, weakly nondictatorial, strategy-proof allocation mechanisms generalizes with some effort to Theorem 2.1, which applies to both rational and nonrational mechanisms and to strong nondictatorship as well as weak nondictatorship.[8]

3. SUFFICIENTLY RESTRICTED DOMAINS AND STRATEGY-PROOFNESS

Within the general theme of restricting the domain of admissible preferences, three approaches have been followed in trying to resolve the fundumental problem that Theorem 2.1 poses for the construction of strategy-proof mechanisms. The first approach begins with a specific allocation mechanism (e.g., majority rule) and searches for domain restrictions that are sufficient to make the mechanism strategy-proof.[9] We do not explore this approach in this paper since it is the least general of the three approaches. The second approach begins with a fixed restricted domain, expressed in terms of economic restrictions such as convexity, continuity and the like, and then looks for nondictatorial strategy-proof mechanisms. We discuss this approach in Section 4. The third approach, which is the most general, fixes neither the domain nor the aggregation rule. It looks for

necessary and sufficient conditions on preferences such that the resulting domain, Ω, permits construction of a strategy-proof mechanism that satisfies the Pareto criterion and nondictatorship plus, in some cases, additional criteria on the distribution of power. This is the approach we explore in this section.

In the previous section we discussed the relationship between rational, strategy-proof allocation mechanisms and social welfare functions that satisfy the conditions of monotonicity and IIA. This relationship is intensively exploited in this Section; with one exception all the results presented apply exclusively to rational allocation mechanisms. Thus the typical result for this section is: if Ω satisfies the following conditions, then Ω admits the construction of a weakly nondictatorial and rational strategy-proof allocation mechanism. This technique, however, is not costless. We discuss the rationality condition in greater depth at the end of this section and show an example of a domain that (i) permits construction of a nonrational, strongly nondictatorial, strategy-proof allocation mechanism and (ii) does not permit construction of a rational, weakly nondictatorial social welfare function. Thus requiring rationality, as this section does, creates a binding constraint. To what extent the results of this section can be generalized if the rationality constraint were dropped is an open question.

Characterization of the domains that admit rational strategy-proof mechanisms requires some notation whose purpose is to allow the structure of a given domain, Ω, to be examined. The set of <u>ordered triples</u> within a domain Ω is defined as $t(\Omega) = \{(xyz) \mid P \in \Omega$ exists such that $xPyPz\}$. Two domains Ω_1 and Ω_2 are <u>equivalent</u> if they share the same set of ordered triples, i.e., $t(\Omega_1) = t(\Omega_2)$. Two domains may be disjoint and equivalent. For example, if $\Omega_1 = \{(xyzw),(yxwz)\}$, $\Omega_2 = \{(xywz), (yxzw)\}$, the $t(\Omega_1) = t(\Omega_2) = \{(xyz),$

(yxz), (xyw), (yxw), (xzw), (xwz), (yzw), (ywz)}. The importance of the equivalence relation that $t(\Omega)$ defines on the set of posssible domains is that if two domains are equivalent, then the first domain permits construction of a strategy-proof, nondictatorial, rational mechanism if and only if the second domain permits construction of such a mechanism.

The set of ordered pairs within Ω is $T(\Omega) = \{(xy) \in A \times A \mid x \neq y\}$. The set of trivial ordered pairs within Ω is $TR(\Omega) = \{(xy) \mid$ a $P \in \Omega$ exists such that xPy and no $Q \in \Omega$ exists such that $yQx\}$. A trivial pair is a pair of alternatives over which no controversy exists because every agent, no matter what element of Ω describes his preferences, agrees on how those two alternatives should be ranked.

Decisiveness Implications. The concept of decisiveness implications is of great importance because it constitutes the technology that has made the statements and proofs of the theorems presented in this section possible. This technology is inextricably bound up with the rationality requirement that we have imposed for this entire section; decisiveness implications do not work for nonrational mechanisms. The thrust behind this technology may be summarized as follows.

Given a rational mechanism F, the members of J are said to be decisive for a over b if a is selected when the feasible set is {a, b} and the members of J report preferences that rank a over b. Formally, J is decisive over the ordered pair (ab) if $F(P, \{a, b\}) = a$ for all $P \in \Omega^n$ such that, for all $i \in J$, $aP_i b$. In terms of social welfare functions, decisiveness means that coalition J can force a social preference for a over b. A dictator is decisive over all pairs in A because, no matter what other agents vote, he secures the outcome he desires.

Suppose a rational, strategy-proof mechanism F that satisfies the Pareto

criterion is defined on a domain Ω^n. Because F is rational, a social welfare function f_F that satisfies the Pareto criterion underlies F. Because the mechanism is strategy-proof, both F and f_F satisfy monotonicity and IIA.[10] Now suppose that a coalition $J \subset I$ is decisive for an alternative $a \in A$ against another alternative $b \in A$. Suppose additionally that an alternative $c \in A$ exists such that the ordered triples (abc) and (bca) are in $t(\Omega)$, i.e., (abc), (bca) $\in t(\Omega)$. Let the coalition J vote (abc) while its complement votes (bca). In other words, the reported profile P has the property that aP_ibP_ic for all $i \in J$ and bP_icP_ia for all $i \notin J$. Since J is decisive over (ab), a is socially preferred to b. Application of the Pareto criterion implies that b is socially preferred to c and, as a consequence of transitivity, a is socially preferred to c. Coalition J is therefore decisive for a over c because (i) $a = \max_{\{a,c\}} f_F(P) = F(P, \{a, c\})$ and (ii) f_F's monotonicity and IIA together imply that no matter how members of J's complement change their votes, the outcome is fixed at a. Therefore, we can conclude, if (abc) and (bca) are in $t(\Omega)$, then any individual or coalition that is decisive over (ab) is necessarily decisive over (ac) as well. This is decisiveness implication number one. Note the central role that transitivity (i.e., rationality) played in its derivation. Our use of the labels a, b, and c here for the elements of A is to emphasize that the implication applies to any ordered triple, e.g., in a particular application a may be assigned the value y, b the value z, and c the value x.

Parallel arguments lead to decisiveness implications numbers two through four: (ii) if (abc), (bca) $\in t(\Omega)$ and a coalition J is decisive over (ca), then J is necessarily decisive over (ba); (iii) if (abc) $\in t(\Omega)$, (bca) $\notin t(\Omega)$, and coalition J is decisive over (ab) and (bc), then J is also decisive over (ac); (iv) if (abc) $\in t(\Omega)$, (bca) $\notin t(\Omega)$, and a coalition J is

decisive over (ca), then J is decisive over either (ba) or (cb).

Consider a domain Ω of admissible preferences and a set of ordered pairs $R \subset T$. The question is: does a rational, strategy-proof mechanism, F, satisfying the Pareto criterion exist such that some coalition J is decisive over exactly the set of pairs contained in R. The answer is: for such a mechanism, the set R can be the collection of pairs over which coalition J is decisive only if R is closed with respect to the four decisiveness implications. The set R is closed with respect to the decisiveness implications if for every $(ab) \notin R$, then, given Ω and R, none of the decisiveness implications implies that J must be decisive over (ab). The idea underlying the definition is that if J can be shown to be decisive over the pair (ab), then by definition (ab) belongs to R already. The domain Ω is decomposable if such a closed set R exists that is a strict subset of the set of all pairs and is a strict superset of the set of trivial pairs. Thus if R is decomposable, then $TR(\Omega) \subsetneq R \subsetneq T$.

Nondictatorial Strategy-proof Mechanisms. Kalai and Muller (1977) used the concept of decomposibility to characterize the domains on which nondictatorial strategy-proof mechanisms can be constructed.

THEOREM 3.1. For $n \geq 2$ an n-person, weakly nondictatorial, rational, strategy-proof mechanism on $\Omega \subset \Sigma$ exists if and only if a two-person, weakly nondictatoral, rational, strategy-proof mechanism on Ω exists.

THEOREM 3.2. For $n \geq 2$, the following three statements are equivalent for every Ω.

a. $\Omega \in \Sigma$ is decomposable.

b. The equivalence class of Ω permits construction of an n-person weakly nondictatorial, rational, strategy-proof

mechanism that satisfies the Pareto criterion.

c. The equivalence class of Ω permits construction of an n-person weakly nondictatorial social welfare function that satisfies the Pareto criterion and IIA.

Consider, for simplicity, the two agent case (n=2). The necessity that Ω be decomposable for construction of a nondictatorial strategy-proof mechanism follows directly from the observation that if the only set of pairs R that is closed and nontrivial is the set, T, of all ordered pairs, then one agent must be a dictator. Suppose, contrary to the observation, neither agent is a dictator and their orderings in the profile P disagree on the nontrivial pair $\{x, y\}$. Since $F(P, \{x, y\})$ is singlevalued, one agent or the other must get his way and is thus decisive on the pair. But if he is decisive over one pair, he is decisive on all pairs because the only closed nontrivial R is identical to T. Therefore the agent is a dictator.

As for the sufficiency of decomposability, if a closed set of ordered pairs $R_1 \subsetneq T$ exists, define R_2 as the set of ordered pairs whose inverses are not in R_1. With this we eliminate the risk of an agent being decisive over a pair (ab) while another agent is decisive over the inverse pair (ba). Define the following social welfare function: Let agent one be decisive over the pairs in R_1. Let agent two be decisive over the pairs in R_2. Let the coalition of agents one and two be decisive over all pairs. If there are more agents, let them be dummies who have no effect on the outcome. It can be shown that this function is a weakly nondictatorial social welfare function satisfying the Pareto criterion and IIA and that it underlies a nondictatorial, strategy-proof mechanism. That this mechanism may be only weakly nondictatorial follows from the fact that--in the sense of strong dictatorship--agent one is a dictator whenever the feasible set is a pair of

alternatives, {a, b}, ε $\underline{A}$, for which (ab), (ba) ε R_1. Agent one is then decisive over both (ab) and (ba); consequently $F(P, \{a, b\}) = \max_{\{a,b\}} P_1$.

The statement of parts (b) and (c) of the equivalence in Theorem 3.2 has a surprising feature. It would have been more intuitive, simpler to prove, and more consistent with the other theorems reported in this section to state in (c) that the social welfare function, f, satisfies the Pareto criterion, IIA, and monotonicity. In the theorem, however, monotonicity is not assumed and thus cannot be used. Instead when existence of a strategy-proof mechanism is to be established given that a nondictatorial n-person social welfare function exists, Theorem 3.1 implies the existence of a two-person nondictatorial social welfare function. Observe that a two-person social welfare function satisfying the Pareto criterion is necessarily monotonic and thus the two-person allocation mechanism that it underlies is strategy-proof. To construct the n-agent, nondictatorial, strategy-proof mechanism, which is the goal of the exercise, add the remaining n-2 agents as dummies.

The generalization of Theorem 3.1 for allocation mechanisms that are not rational has been proven by Kim and Rousch (1981). To our knowledge this is the only result of this section that has been generalized from the rational case to the nonrational case.

Graphical Representation. The graph that the decisiveness implications creates among the elements of the set of all pairs helps understand the decomposability condition. As a first step, consider Feldman's proof of Theorem 2.1, which we presented in Section 2, for the special case of two agents and three alternatives. Here we use decisiveness implications to create an analogous proof for a somewhat weaker result: for three alternatives and two agents every strategy-proof, rational mechanism that satisfies the Pareto criterion is dictatorial. The result is weaker because

this section's technique only applies to rational mechanisms; Feldman's technique applies to both rational and nonrational mechanisms.

The domain, $\Omega_1 \equiv \Sigma$, consists of all six strict orderings that are possible when the number of alternatives is three. In Figure 3.1 the nodes of the graph consists of all the ordered pairs T. The directed branches represent application of decisiveness implications (i) and (ii) to each of the six orderings. For example, if agent one is decisive over a pair (xy) and, as is the case, (xyz), (yzx) ε Ω_1, then decisiveness implication (ii) implies that he has to be decisive over (xz) as well. Thus a directed branch connects (xy) to (xz) because (xyz), (yzx) ε Ω_1. This follows from the first decisiveness implication if x is assigned to a, y to b, and z to c. Similarly, decisiveness implication (i) implies that a directed branch connects (xy) to (zy) because (yzx), (zxy) ε Ω_1. If all branches are filled in, the graph in Figure 3.1 results. It is evident that the direct branches generated by decisiveness implications (i) and (ii) span the whole set of pairs. No sinks (i.e. closed sets of ordered pairs that are strict subsets of T) exist there. A set R is identified graphically to be a sink if branches only go into it while none come out of it. Agent one is therefore necessarily decisive over all six pairs and is a dictator.

A slightly different way to see that one agent must be a dictator is to replicate for rational mechinisms the several steps of Feldman's proof (see Section 2) that led to the conclusion that if alternative x is assigned to cell 1, then agent one is necessarily a dictator. Assignment of x to cell 1 means that we resolve in favor of agent one the conflict over the pair {x, z} that occurs when agent one votes (xzy) and agent two votes (zxy) The Pareto criterion eliminates alternative y as a possible outcome. Thus that assignment makes voter one decisive over (xz). It also implies that x

must be assigned to cells 2 and 3 because agent one is decisive over (xz) and the Pareto criterion eliminates y from consideration. In cell 4 agent one's decisiveness over (xz) eliminates z as a possible outcome. If x is assigned to a, z to b, and y to c, then decisiveness implication (ii) states that agent one is decisive over (xy) because he is decisive over (xz). This process may be continued until all cells are assigned agent one's preferred choice.

If we delete even one ordering out of Σ, then a number of sinks result, which means that the resulting Ω is decomposable and, according to Theorem 3.2, a rational, nondictatorial, strategy-proof allocation mechanism can be constructed on Ω. To be specific, let (zyx) be deleted from Σ and call the resulting domain Ω_2. Figure 3.2 shows its graph. It differs from Figure 3.1 as follows. The two direct branches from (yx) to (zx) and from (xz) to (xy) that decisiveness implications (i) and (ii) respectively would generate if (zyx) were an element of Ω_2 are deleted. But because (xzy) $\in \Omega_2$ and (zyx) $\notin \Omega_2$ both a joining branch and a splitting branch are eligible to be added. Decisiveness implication (iii) generates the joining branch; it connects both (xz) and (zy) to (xy) and means that if R contains both (xz) and (zy), then it must also contain (xy). It is not drawn because the decisiveness implication (zy) $\rightarrow$ (xy) makes this joining branch redundant. Decisiveness implication (iv) generates the splitting branch; it connects (yx) to both (zx) and (yz) and means that if R contains (yx), then it must also contain either (zx) or (yz). It, too, is not shown because it is redundant. Note, however, that if in addition (yzx) were dropped as an element of Ω_2, then neither the joining nor the splitting branch would be redundant.

These changes cause Figure 3.2 to have four sinks: $R_1 = \{(xz)\}$, $R_2 = \{(zy), (xz), (xy)\}$, $R_3 = \{(xz), (yz), (yx)\}$, and $R_4 = R_2 \cup R_3 = T - (zx)$. Associated with each sink is a distinct, weakly nondictatorial, strategy-proof

mechanism. Therefore four distinct, strategy-proof, nondictatorial, rational mechanisms can be constructed on the domain Ω_2. Note that R_4 includes (xy) and its inverse (yx) and (yz) and its inverse (zy); this means the mechanism that is associated with it is not strongly nondictatorial because whenever B = {x, y} or {z, y} the agent who is decisive over the elements of R_4 dictates the choice. A similar argument applies for R_1, but not for R_2 or R_3. Thus if $|A| = 3$, the deletion of a single ordering from Σ is sufficient to reverse the Theorem 2.1's impossibility result.

Essentiality and Symmetry. In the discussion that followed Theorems 3.1 and 3.2 we described how to construct a nondictatorial, strategy-proof mechanism when Ω is decomposable. That mechanism, however, distributes power with unacceptable uneveness: n-2 of the individuals are dummies. As this particular mechanism illustrates, requiring that a mechanism satisfy nondictatorship is a toothless requirement that comes nowhere near describing the criteria by which we judge a distribution of power acceptable or not acceptable. Nondictatorship (strong or weak) is a necessary, but not sufficient, condition for a mechanism to be acceptable and as such is useful within the context of impossibility theorems. Possibility theorems need additional conditions that capture what we mean when we judge a particular power distribution acceptable.

Two such conditions are essentiality and symmetry. For a mechanism F an agent i is essential if a preference profile $P \in \Omega^n$ and ordering $Q_i \in \Omega$ exist such that if agent i changes his ordering from P_i to Q_i, then the outcome changes from $F(P, B)$ to $F(P/Q_i, B) \neq F(P, B)$. A mechanism is essential if all agents are essential. In essential mechanisms each individual has some, though not necessarily equal, power. Symmetry, on the other hand, mandates equal power without specifying the magnitude of the power. A mechanism F is

symmetric (sometimes called anonymous) if any permutation of the individuals leaves the outcome unchanged: for all $P \in \Omega^n$, $B \in \underline{A}$, and permutations ρ: $\{1,\ldots, n\} \to \{1,\ldots, n\}$, $F(\ldots,P_i,\ldots, B) = F(\ldots, P_{\rho(i)},\ldots, B)$. Neither of these conditions completely supplants the strong nondictatorship conditions or the Pareto criterion. For example, an imposed mechaninism is symmetric because under such a mechanism every individual is identical in having no influence over the outcome.

For the case of essential mechanisms Blair and Muller (1983) have generalized the concept of decisiveness implications and proved the natural extension of Theorem 3.2. The example, which follows, of an essential mechanism and the domain on which it is constructed shows that essential mechanisms, while an improvement over weakly nondictatorial mechanisms, only incompletely captures the considerations that enter our evaluations of whether a particular distribution of power is acceptable. Let $A = [X^1,\ldots,X^K]$ where each X^i is a vector of three alternatives. The domain Ω consists of all orderings in which the elements of X^k appears always above X^ℓ for all $k < \ell$ and, within X^k, the three alternatives form a free triple, i.e., all six orderings of the three elements of X^k are permissible. Let each voter be a dictator on at least one of the free triples. The result is an essential monotonic SWF for K or fewer voters. Any additional voters must be dummies and are therefore not essential.

Symmetry is a more stringent condition than that of essentiality. It too can be approached using the technology of decisiveness implications. A domain Ω is transitively decomposable if a nontrivial set R exists that is (i) closed under decisiveness implications one through four and (ii) transitive. Transitivity in this context means that R must satisfy: (i) if $(xy), (yz) \in R$, then $(xz) \in R$ and (ii) $(xy) \in R$ if and only if $(yx) \notin R$. The following two

theorems summarize the symmetric case, and are adapted from Muller (1982). The equivalence between the second and third parts of Theorem 3.4 are extensions that are not proven in the original paper, but can straightforwardly be shown by means similar to those used in Blair and Muller (1983).

THEOREM 3.3. A symmetric, n-person, rational strategy-proof mechanism on Ω exists for all $n \geq 3$ if and only if a symmetric, three-person rational strategy-proof mechanism on Ω exists.

THEOREM 3.4. The following three statements are equivalent for every $\Omega \in \Sigma$:

a. Ω is transitively decomposable.

b. For all $n \geq 3$ the equivalence class of Ω permits construction of a symmetric, monotonic, social welfare function that satisfies the Pareto criterion and IIA.

c. For all $n \geq 3$ the equivalence class of Ω permits construction of a symmetric, rational, strategy-proof mechanism that satisfies the Pareto criterion.

The social welfare function in part (b) is required to be monotonic because, unlike in the case of Theorems 3.1 and 3.2, no theorem exists that reduces the n-person case to the two-person case. Indeed, with respect to Theorem 3.3, an Ω exists for which a a symmetric two-person social wnlfare function may be constructed, but not a symmetric three-person social welfare function. In parts (b) and (c) no reference is made to nondictatorship because any symmetric mechanism that satisfies the Pareto criterion also satisfies strong nondictatorship.

Group Strategy-proofness. A mechanism is strategy-proof if no single individual ever has an incentive to misrepresent his preferences. Blair and

Muller (1983) have shown the suprising result that, for rational mechanisms, strategy-proofness for individuals is equivalent to strategy-proofness for coalitions of individuals. A coalition J, $|J| = k < n$, has an incentive to manipulate the mechanism F at profile P and feasible set $B \in \underline{A}$ if orderings $Q_i \in \Omega$ exist such that, for all $i \in J$,

$$F[P/Q, B] \; P_i \; F(P, B) \tag{3.01}$$

where $Q = \{Q_i\}_{i \in J}$. A mechanism F is <u>group strategy-proof</u> if no admissible profile $P \in \Omega^n$, no set B, and no coalition J exists such that at profile P coalition J can manipulate mechanism F.

The driving force behind this equivalence of group strategy-proofness and individual strategy-proofness is the rationality condition. To show this we first observe that if a mechanism is group strategy-proof, then by definition it is individually strategy-proof. We then show that a rational mechanism that is individually strategy-proof is also group strategy-proof by demonstrating that if a mechanism is manipulable by some coalition, then it is also manipulable by some individual within the coalition. Therefore an individually strategy-proof mechanism must also be group strategy-proof.

Suppose, in order to see that group manipulability implies individual manipulability, that a group $J = \{1,\ldots,k\}$, $k < n$, can manipulate F at profile P and feasible set B:

$$F(Q_1,\ldots, Q_k, P_{k+1},\ldots, P_n, B) \; P_i \; F(P, B). \tag{3.02}$$

Let $F(Q_1,\ldots, Q_k, P_{k+1},\ldots, P_n, B) = x$ and $F(P, B) = y$. Note that xP_iy for all $i \in J$. The rationality of F implies that $F(Q_1,\ldots, Q_k, P_{k+1},\ldots, P_n, \{x, y\}) = x$ and $F(P, \{x, y\}) = y$. Therefore (3.02) continues to be true when B is replaced as the feasible set by $\{x, y\}$. Moreover, because $F(P, \{x, y\}) \in \{x, y\}$ for all $P \in \Omega^n$, a $j \in J$ must exist such that $F(Q_1,\ldots, Q_{j-1}, P_j, P_{j+1},\ldots, P_n, \{x, y\}) = x$ and $F(Q_1,\ldots, Q_{j-1}, Q_j, P_{j+1},\ldots,$

P_n, {x, y}) = y, i.e. j ε J is the critical voter who switches the outcome. Voter j, as a member of J, has preferences xP_iy; therefore he can individually manipulate F at feasible set {x, y} and profile ($Q_1,\ldots, Q_{j-1}, P_j, P_{k+1},\ldots, P_n$). Note that the rationality requirement was what allowed us to reduce the problem to that of selection between two alternatives.

The Private Goods Case. The discussion to this point has considered only a perfectly general conceivable set, A, that has no a priori structure imposed on it. Suppose, however, that each alternative within A is a vector of n distinct private goods' bundles, each one of which is to be allocated to one of the n agents. To accomodate this change, let A represent each individual's consumption set, x_i ε A be the bundle of private goods agent i consumes, P_i be his preferences over A, and let Ω ⊂ $\Sigma_{\sim}$ be the set of orderings over A to which P_i is a priori restricted. Note that indifference is permitted as indicated by Ω being contained in $\Sigma_{\sim}$ rather than Σ. Also, note that every individual i is selfish in that he is concerned only with his own component of the alternative $x = (x_1,\ldots,x_n)$.

An allocation $x = (x_1,\ldots, x_n)$ ε A^n is the n-vector of the agents' private goods' bundles. Redefine, for this subsection, an allocation mechanism to be a function F: $\Omega^n \times \underline{A} \to A^n$. Thus $F(P, B) = [F_1(P, B),\ldots, F_i(P, B),\ldots, F_n(P, B)]$ is a vector of n functions where the ith function, $F_i(P, B)$ ε B, specifies the allocation of private goods agent i receives.

Two new definitions must be introduced in order to characterize the domains on which weakly nondictatorial, strategy-proof mechanisms can be constructed. First is a strengthening of the Pareto criterion. A mechanism F satisfies the strong Pareto criterion if, for any pair x, y ε B, xP_iy for at least one agent and yP_ix for no agent, then F(P, B) ≠ y. This strong version differs from the weak version in that the strong version does not require

unanimity and permits some agents to be indifferent between x and y, i.e. it may apply even if $x\tilde{P}_i y$ for some agents.

The second new condition is Ritz's (1981, 1983) noncorruptibility condition. A mechanism is noncorruptible if for all sets $B \in \underline{A}$, all profiles $P \in \Omega^n$, all agents $i \in I$, and all orderings $Q_i \in \Omega$, $F_i(P, Q) \tilde{P}_i F_i(P/Q_i, B)$ implies $F_j(P, B) = F_j(P/Q_i, B)$ for all agents (j $j \neq i$). Recall that $\tilde{P}_i$ signifies indifference. Thus, for a noncorruptible mechanism, an agent must change the utility value of the outcome to himself in order to affect the physical outcome of other agents. Informally, if a mechanism is corruptible, then agent i, who may be thought of as a potential corruptor or boss, does not directly improve his own outcome as is the case in manipulation. Rather, he changes the value of the outcome to others. He thus creates a possibility of indirectly improving his position by threatening other agents and demanding side payments. Thus corruptibility sets the stage for indirect manipulation as opposed to the direct manipulation with which strategy-proofness is concerned.

Kalai and Ritz (1980) and Ritz (1981, 1983) have used the technology of decisiveness implications and decomposability to make substantial progress on the private goods case. The private goods decisiveness implications to which the following theorem of Ritz (1981, 1983) makes reference are not reproduced here in the interests of brevity. They may be found in Ritz's original papers.[12]

THEOREM 3.5. For the private goods case, when $n \geq 2$, the following three statements are equivalent for every Ω:

a. Ω is decomposable over private alternatives.

b. Ω permits construction of an n-person, weakly nondictatorial social welfare function that satifies the

strong Pareto criterion and IIA.

c. Ω permits construction of an n-person, weakly nondictatorial, rational, noncorruptible strategy-proof mechanism that satisfies the strong Pareto criterion.

This theorem parallels Theorem 3.2 in not requiring the social welfare function in part (b) of the theorem to be monotonic. The reason is that Ritz, like Kalai and Muller, exploited the permissiveness of the nondictatorship condition to construct through the use of dummies n-person mechanisms from two person mechanisms.

Restrictiveness of the Decomposability Conditions. The results presented in this section succeed in characterizing for several contexts the domains on which construction of strategy-proof mechanisms is possible. The question that remains is: How restrictive are these conditions? The ideal way to answer this question would be to determine, for a variety of different economic environments, if the a priori restrictions on agents' preferences that those environments naturally induce satisfy the characterizations for strategy-proof domains that have been presented. This approach has not been successfully carried out. A second approach, which has met with some success, is to calculate how close to unity the ratio $|\Omega|/|\Sigma|$ can be made to come when Ω is restricted to admit the construction of a nondictatorial, strategy-proof mechanism. If examples exist in which, even with a large number of alternatives, the size of the restriced domain is still "respectable" relative to the size of the full domain, then that is an indication that these characterizations are not very restrictive.

Kim and Roush (1981) have shown that if $|A| = m$, then

$$\frac{m!/2 + (m-1)!}{m!} = \frac{1}{2} + \frac{1}{m} \tag{3.03}$$

is the upper bound on $|\Omega|/|\Sigma|$ for weakly nondictatorial, rational, strategy-proof mechanisms satisfying the Pareto criterion. Because essential and symmetric mechanisms are also weakly nondictatorial, (3.03) is also the upper bound for domains that permit construction of essential or symmetric) rational strategy-proof mechanisms. Blair and Muller (1983), based on the work of Kalai and Ritz (1979), have constructed an example of an essential mechanism that achieves this bound. Expression (3.03) is thus a least lower bound for essential and weakly nondictatorial mechanisms.

The domain, $\Omega \subset \Sigma$, for Blair and Muller's example is defined by a single restriction: it contains an ordered pair, $(xy) \in T(\Omega)$, with the property that no alternative $z \in A$ and ordering $P_i \in \Omega$ exist such that xP_izP_iy, i.e. no $z \in A$ exists such that $(xzy) \in t(\Omega)$. The pair (xy) is thus inseparable in the sense that an admissible ordering can rank x immediately above y or someplace below y. Given this domain, an essential rational mechanism is this. Let voter one be decisive over all ordered pairs $(ab) \in T(\Omega)$ except (xy) and let each other individual have veto power over (xy). Thus $x = F(P, \{x, y\})$ if and only if xP_iy for all $i \in \{2, 3, \ldots, n\}$. This defines a social welfare function that, with one exception, makes agent one the dictator in the sense that his ordering becomes the social ordering. The exception occurs when agent one ranks x just above y and some other agent (the vetoer) objects by ranking y above x. In that event the social ordering is modified by placing y immediately above x. It is straightforward to check that this defines an essential, monotonic, weakly nondictatorial social welfare function that satisfies the Pareto critierion and IIA. It thus also defines an essential, weakly nondictatorial, strategy-proof, rational allocation mechanism that satisfies the Pareto criterion.

The size of this domain is $m!/2 + (m - 1)!$. This formula is easily derived by considering that subset of Ω for which $xP_i y$ seperately from that subset for which $yP_i x$. Since m alternatives may be strongly ordered in m! different ways, the $|\Omega|/|\Sigma|$ ratio equals the value of (3.03). The weakness of this example is the already emphasized fact that essential mechanisms may incorporate an unacceptable distribution of power among the participating individuals. In this particular example that is surely the case because agent one is nearly a dictator. Therefore this particular example is not convincing as evidence that the decomposability conditions are relatively unrestrictive.

Examples for the symmetric case, which might be more convincing, have not been constructed yet. Kim and Roush (1981) showed that if mechanisms that give agents veto power are excluded from consideration, then the $|\Omega|/|\Sigma|$ ratio goes to zero as m goes to infinity. This, however, is not convincing evidence in the opposite direction because, since symmetric mechanisms with veto power may be constructed, no compelling reason to prohibit the use of the veto is apparent.

The Rationality Requirement. Throughout this section we have only considered rational allocation mechanisms. This is not a benign requirement. If the rationality condition is dropped, then the opportunities for constructing strategy-proof mechanisms increase. This point is made most concretely by an example due to Maskin (1976). His example identifies a domain, Ω, that has two important properties: (i) a strategy-proof, strongly nondictatorial mechanism satisfying the Pareto criterion exists on it and (ii) no weakly nondictatorial social welfare function satisfying the Pareto criterion and IIA exists on it.

We present in this subsection a corrected and much simplified version of

Maskin's example. Let $\Omega = \{(xzyw), (yzwx), (yxwz), (wxzy), (zwxy), (wzyx)\}$. Denote these six admissible orderings by P_i, i=1,...,6, according to the order they appear in Ω. It is straightforward to check that the mechanism defined in Table 3.2 is strategy-proof when the feasible set consists of all four alternatives in A = {w, x, y, z}.

The nonrationality of F means that, for each of the four feasible sets B containing three of the four alternatives in A ({x, y, z}, {w, x, y}, etc.), the mechanism F(•, B) may be defined without reference to the way Table 3.2 defines it for the case where A is the feasible set. Inspection of the orderings contained within Ω shows that if one alternative is eliminated from each of its constituent orderings, then five (out of the six possible) distinct orderings of the three remaining alternatives are left. For example, if B = {x, y, z}, then the domain that results by striking w from each ordering in Ω is

$$\Omega_{\{x,y,z\}} = \{(xzy), (yxz), (yzx), (zxy), (zyx)\} = \Sigma - (xyz). \qquad (3.04)$$

Earlier in this section we showed that, when $|A| = 3$, elimination of one ordering from Σ is sufficient to permit construction of a strongly nondictatorial, rational mechanism on that feasible set. Define F(•, B) to be one of those mechanisms whenever $|B| = 3$ or $|B| = 2$. The result is a nonrational, strongly nondictatorial strategy-proof mechanism that satisfies the Pareto criterion.

To complete the example we have to show that this domain does not permit construction of a weakly nondictatorial social welfare function that satisfies the Pareto criterion and IIA. This is easily done by constructing the graph of Ω to arrive at Figure 3.3. Since that graph does not contain any sink Theorem 3.2 implies that a weakly nondictatorial social welfare function does not exist on Ω.

This example shows that the upper bounds on the $|\Omega|/|\Sigma|$ ratio that Kim and Roush (1981) derived for rational mechanisms do not necessarily hold for nonrational mechanisms. This emphasizes that our knowledge is quite imperfect concerning the degree to which the admissible domain must be restricted in order to permit construction of a reasonable strategy-proof mechanism.

4. STRATEGY-PROOFNESS ON SPECIFIC RESTRICTED DOMAINS

The last section reported on work that has been done to characterize those domains of preferences that are restrictive enough to permit the construction of strategy-proof allocation mechanisms that share power among a group's members in some acceptably democratic way. Considerable progress has been made on this approach, but as yet no researcher has succeeded in relating those characterizations to the domains of admissible preferences that occur in economic situations. This section reports on work that has taken the less general approach of beginning with a domain where preferences are restricted to belong to a class that naturally arise in economic environments and then characterizing the strategy-proof allocation mechanisms that can be constructed on that domain. In other words, the methodology of the last section is turned on its head here: instead of beginning with properties that a strategy-proof mechanism should possess and deriving those domains that are consistent with those properties, we begin with a domain and derive the properties of the mechanisms that are consistent with that domain.

Economists often represent bundles of commodities as points in Euclidean space. Therefore, in this section where we are concerned exclusively with economic environments, A is no longer a set of discrete points without structure. Instead an alternative $x = (x_1, \ldots, x_\ell)$ is a point within a consumption set A that is itself a subset of ℓ-dimensional Euclidean space.

The interpretation of x_k, the k-th component of x, is that the bundle x contains x_k units of good k. Imposition of this Euclidean structure on A enables us to utilize the concepts of continuity and differentiability. Specifically, given this structure on A, a natural restriction to place on the admissible preferences of agents is that they be representable by twice differentiable, strictly concave utility functions, $u_i(x)$, that are increasing with respect to each of their arguments.

Clearly such a restriction on Ω, the set of admissible preferences, is strong. It strength can be seen by letting $\ell = 2$ and considering a sequence of ten points $\{x^1,\ldots, x^{10}\}$ that are randomly selected from a convex consumption set A that has a nonempty interior. The probability that the ordering $(x^1\ x^2 \ldots x^{10})$ is consistent with Ω is minuscule--certainly less than 0.5. It therefore is in some sense a stronger restriction than some of the restrictions on preferences identified in Section 3. Recall, in particular, Blair and Muller's (1983) example of a domain Ω that (i) contains more than half the possible orderings that can be defined on A and (ii) admits the construction of a rational, weakly nondictatorial, strategy-proof mechanism satisfying both essentiality and the Pareto criterion.

In addition to restricting ourselves in this discussion to preferences that are sufficiently smooth, we also restrict ourselves to mechanisms that have continuous derivatives. A sensible allocation mechanism in an economic environment can not be everywhere nondifferentiable. To be nondifferentiable everywhere would mean that whenever an individual agent perturbed his preferences, then the outcome would jump in a new direction. Clearly, however, an allocation mechanism need not be smooth everywhere; it is quite acceptable for the allocation to jump at some points. This means that the results reported in this section should be considered to be local

characterizations of the possible strategy-proof mechanisms. As a consequence we do not discuss the results, for example, of Border and Jordan (1983) who for a very restrictive domain of admissible preferences consider strategy-proof mechanisms that are nondifferentiable at isolated points.

These ideas are easily formalized provided that we change the manner in which the agent i reports his preferences from being P_i, a binary relation on A, to u_i, a real valued utility function on A. A utility function $u_i(\cdot)$ represents the preference ordering P_i if: xP_iy if and only if $u_i(x) \geq u_i(y)$.[13] Let A, the set of admissible alternatives, be a compact, convex subset of $\mathbf{R}^\ell$ with nonempty interior. Redefine Ω to be the set of admissible utility functions on A. We assume that every $u_i \in \Omega$ is twice continuously differentiable and that Ω itself is a convex subset of a linear function space that is endowed with the C^2 topology.[14]

Rationality plays no role in this section. Therefore the feasible set, $B \in \underline{A}$, can be fixed equal to A because, with rationality no longer an issue, permitting B to vary serves no function. An allocation mechanism within economic environments is therefore a function $F: \Omega^n \to A$. Note that, since the feasible set is fixed, B is dropped as an argument. A mechanism F is <u>strategy-proof</u> if, for all profiles $u \in \Omega^n$, all utility functions $u_i' \in \Omega$, and all agents i, $u_i[F(u)] \geq u_i[F(u/u_i')]$.

Let $C^2(A)$ be the set of all twice continuously differentiable functions on A. The mechanism F is continuously differentiable at $u \in \Omega^n$ if for all $v \in [C^2(A)]^n$,

$$D_vF(u) = \lim_{\lambda \to 0} \frac{F(u + \lambda v) - F(u)}{\lambda} \tag{4.01}$$

exists, is continuous in both u and v, and has the standard property that $D_{cv+dw}F(u) = cD_vF(u) + dD_wF(u)$ for all scalars $c,d \in \mathbf{R}$ and all functions v, $w \in [C^2(A)]^n$. Note that v and w are vectors of n distinct C^2 functions. This

means that D_v is defined in terms of all n of the agents' utility functions being perturbed simultaneously. To represent the more restrictive case where only one agent's utility function is perturbed, let (v/i) be the element of $[C^2(A)]^n$ that has as its ith component the function $v_i \in C^2(A)$ and has as its other n-1 components the constant function with value zero. The derivative $D_{(v/i)}F$ is therefore the direction within A in which the allocation F(u) moves as the function v_i perturbs agent i's utility function u_i. Agent i affects the allocation F(u) at $u \in \Omega^n$ if a $v_i \in C^2(A)$ exists such that $D_{(v/i)}F(u) \neq 0$. Agent i affects agent j's utility at $u \in \Omega^n$ if a $v_i \in C^2(A)$ exists such that $D_{(v/i)}u_i[F(u)] \neq 0$ where $D_{(v/i)}u_i[F(u)]$ represents the derivative of agent i's utility when his utility function is perturbed by v_i.

The Simplest Case. The constraints that strategy-proofness places on the design of allocation mechanisms within economic environments are most easily seen within the simple one agent, two good implementation problem (i.e., $n = 1$ and $\ell = 2$) that Guesneri and Laffont (1982) have analyzed. Let the function to be implemented be $G: \Omega \to A$ where A is convex subset of $\mathbf{R}^2$. Thus if $u \in \Omega$ represents the agent's true prefences, then the outcome should be $G(u) = x \in A$. The goal, as it always is in dominant strategy implementation problems, is to devise a mechanism F such that (i) the agent has an incentive to report u accurately and (ii) the outcome that reporting u acccurately generates is G(u). The second requirement means that F must be identical to G; otherwise F would not select G(u) when it induces accurate revelation. Consequently G is implementable in dominant strategies if and only if G is a strategy-proof mechanism.

Let $u, u' \in \Omega$ be smooth and strictly concave utility functions defined on A and let, for all $x \in A$ and all $\theta \in [0, 1]$, $u_\theta(x) = \theta u(x) + (1-\theta)u(x)$ be a

linear combination of u and u'. Define the single agent's admissible preferences to be $\Omega = \{u_\theta \mid \theta \in [0, 1]\}$, i.e., Ω is the family of smooth and strictly concave functions u and u' generate. The agent reports his preferences to the mechanism by reporting an $\eta \in [0, 1]$. He then receives allocation $G(u_\eta) \in A$. Therefore as θ (or η) varies between zero and one the image of $G(u_\theta)$ traces out a curve in A. This curve is the choice set for the agent. Depending on his true value of θ he reports the η that maximizes his true preferences.

The solid curves Γ and Γ' that appear in Figures 4.1 and 4.2 respectively represent two possible images of G. In both figures point a corresponds to $G(u_\theta)$ when $\theta = 0.0$, point b to $G(u_\theta)$ when $\theta = 0.3$, and point c to $G(u_\theta)$ when $\theta = 0.5$. The dotted curves represent the indifference curves that u_θ generates: the left pair of dotted curves in each diagram are for $u_{0.3}$ and the right pair are for $u_{0.5}$. Figure 4.1 is consistent with G being strategy-proof because at points b and c respectively the indifference curves of $u_{0.3}$ and $u_{0.5}$ are tangent with Γ. Therefore, subject to the constraint that he must pick a point on Γ, the agent maximizes his utility by reporting his preferences truthfully: $\eta = \theta$. Figure 4.2 is inconsistent with strategy-proofness because the indifference curves of $u_{0.3}$ are not tangent to Γ' at point b and the indifference curves of $u_{0.5}$ are not tangent at point c.

Figure 4.2 is the generic case while Figure 4.1 is the exceptional case. Given a family of utility functions such as u_θ and an arbitrary function G, then typically a not strategy-proof situation like Figure 4.2 occurs. It, in an informal sense, is an exceptional event (occuring with zero probability) that Figure 4.1 with its very special, carefully drawn tangencies occurs. More formally, Guesnerie and Laffont's (1982) result is that generically in the single agent case an arbitrary function G can not be

implemented in dominant strategies.

An Impossibility Theorem for Many Agents. Satterthwaite and Sonnenschein (1981) have shown that the Gibbard-Satterthwaite Theorem carries over to economic environments whenever public goods only are being allocated and the mechanism F is broadly applicable. Public goods only means that agents care about and have preferences over all ℓ dimensions of the consumption set. A mechanism is broadly applicable if Ω, the set of admissible utility functions, is open. The openess of Ω, coupled with its linearity and C^2 topology, has an important implication: if Ω is open, $u_i \in \Omega$, and $v \in C^2$, then a $\delta > 0$ exits such that $(u_i + \lambda v) \in \Omega$ for all $\lambda \in [0, \delta)$. In other words, if a mechanism is broadly applicable, then any admissible utility function remains an admissible utility function when it is perturbed slightly through the addition of another C^2 function, λv. The logic behind the broad applicability requirement is that a perturbed admissible utility function should itself be admissible because "while preferences within an economic environment may have considerable a priori structure such as strict convexity, preferences are not naturally limited to any particular parametric form."[15]

Let $\Gamma_i(u_{-i}) = \{x \in A \mid x = F(u/u_i') \text{ where } u_i' \in \Omega\}$ be the choice set of agent i. Note that Γ_i only varies with $u_{-i} = (u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n)$. A profile $u \in \Omega^n$ is a regular point of a strategy-proof mechanism F if:

a. The mechanism, F, is continuously differentiable in u.

b. For all i and all u_{-i}' in some neighborhood of u, $\Gamma_i(u_{-i})$ is continuously differentiable in u_{-i} and is a k_i-dimensional, $0 \leq k_i \leq \ell-1$, smooth manifold in a neighborhood of the allocation F(u).

c. For all i, F(u) is the unique and well behaved maximizer of u_i on $\Gamma_i(u_{-i})$.

A regular point therefore is a point where each agent faces a smooth choice set that changes position smoothly as the other agents' preferences change.

These definitions and notation allow us to state Satterthwaite and Sonnenschein's (1981, Theorem 3) public goods only result.

THEOREM 4.1 If an allocation mechanism F allocates public goods only, is strategy-proof, and is broadly applicable, then at every regular point $u \in \Omega^n$ an agent i exists who is a dictator at u.

A dictator within this context is an agent who selects his most preferred point from an exogenously given set of achievable points. In other words, an agent i is a dictator if $\Gamma_i(u_{-i})$ is a constant as u_{-i} varies.

Several comments should be made about this result. First, the result is true only for public goods. The private goods analogue is discussed below. Second, the result is local. If agent i is a dicatator in some neighborhood of u, then a second regular point u', which is seperated from u, may exist at which some different agent is the dictator. Satterthwaite and Sonnenschein observe, however, that if the set of regular points is a connected set and the mechanism, for all regular points, is total, then a single agent i is the dictator at all the regular points. In the of public goods only context, a mechanism is total if at every regular point u at least one agent affects the allocation F(u).

Third, the Theorem is stated without the Pareto criterion. Therefore imposed mechanisms are consistent with the Theorem. An imposed mechanism permits no individual to influence the choice of outcome, i.e., F(u) is a constant function as u varies. Thus if a mechanism is imposed at u, then, for all agents i, the manifold $\Gamma_i(u_{-i})$ is a zero dimensional, nonvarying point within A, which means formally that every agent is an (inconsequential) dictator. Fourth, if a mechanism is not imposed and agent i is dicatator at

u, then for all agents j ($j \neq i$), $\Gamma_j(u_{-j})$ is the point within A that agent i, the dictator, selects from his exogenously given choice set Γ_i.

The most interesting step in the proof of Satterthwaite and Sonnenschein's proof of Theorem 4.1 is contained in their Lemma 2. That Lemma in the public goods only case is this: If at a regular point u of a broadly applicable and strategy-proof mechanism an agent i exists who affects the utility of some other agent j, then agent j can not affect the utility of agent i. To begin a simple proof by contradiction, suppose that each of the two agents can affect the other's utility, the mechanism F is both broadly applicable and strategy-proof, and that (without loss of generality) n=2 and ℓ=2.

Figure 4.3 shows what this supposition means. Point a is the base point for the proof and is the allocation $F(u) = F(u_1, u_2)$ where $(u_1, u_2) \in \Omega^2$ is a regular point. At u agent one's choice set is $\Gamma_1(u_2)$ and agent two's is $\Gamma_2(u_1)$. The indifference curves of agents one and two that pass through point a are the dotted lines labeled respectively u_1 and u_2; in conformance with the requirements of strategy-proofness and regularity they are tangent to their respective choice sets. If agent one perturbs his preferences u_1 slightly to become u_1', which is admissible because the mechanism is broadly applicable, then his most preferred point on his choice set, $\Gamma_1(u_2)$, becomes $F(u_1', u_2)$, which is labeled as point b. His indifference curve through point b is labeled u_1'. This changes agent two's achievable set to become $\Gamma_2(u_1')$. Note that agent two prefers point b to point a; therefore the hypothesis that agent one can affect agent two's utility is met.

Figure 4.4 develops the contradiction from the basic situation of Figure 4.3. Because F is broadly applicable agent two can construct a small perturbation of his preferences from u_2 to u_2' so that the following three

specifications hold simultaneously:

a. Point c is $F(u_1, u_2')$. The indifference curve for u_2' is tangent to $\Gamma_2(u_1)$ at point c. Agent 1, when his utility function is u_1, prefers point c to point a, which means that agent two affects agent one's utility as the proof's initial hypothesis requires.

b. Point b, by construction, is $F(u_1', u_2')$ as well as $F(u_1', u_2)$.

c. Again by construction, agent one's choice set becomes $\Gamma_1(u_2')$ when agent two perturbs his utility function from u_2 to u_2'. Note that $\Gamma_1(u_2')$ crosses $\Gamma_1(u_2)$ at point b.

Because point b is $F(u_1', u_2')$, strategy-proofness and regularity require that point b be that point on $\Gamma_1(u_2')$ where agent one's utility is maximized when his preferences are u_1', i.e, the u_1' indifference curve must be tangent to $\Gamma_1(u_2')$ at point b. But this is a contradiction because the u_1' indifference curve through point b is necessarily tangent to $\Gamma_1(u_2)$ and, at b, $\Gamma_1(u_2')$ crosses $\Gamma_1(u_2)$. Therefore the proof is complete: at a regular point of a broadly applicable and strategy-proof mechanism agents one and two can not each affect the other's utility.

Theorem 4.1 generalizes from the public goods only case to settings that include private as well as public goods. To accomodate this change from public to public and private goods, let A be each agent's consumption set and redefine an allocation mechanism to be a function $F: \Omega^n \to A^n$. Thus $F(u) = [F_1(u),\ldots, F_i(u),\ldots, F_n(u)]$ is a vector of n functions where the ith function, $F_i: \Omega^n \to A$, specifies the allocation agent i recieves. The function F_i itself has ℓ components: $F_i = [F_{i1},\ldots, F_{i\ell}]$ where F_{ik} is the amount agent i receives of good k. If some components of each agent's allocation is a public good, then all the functions F_i are constrained to give each agent the same amount of the public good. Thus if good one is a public

good, then $F_{11} = F_{21} = \ldots = F_{n1}$. Agents are assumed to have preferences over only their own consumption set, e.g., agent i's utility is $u_i[F_i(u)]$.

Satterthwaite and Sonnenschein call a mechanism nonbossy if, for all $u \in \Omega^n$, all agents i, and all $u_i' \in \Omega$, $F_i(u) = F_i(u/u_i')$ implies $F_j(u) = F_j(u/u_i')$ for all agents j. The idea of nonbossiness is that if an agent i changes his preferences in a manner that leaves his own allocation unchanged, then the allocations that all other agents receive should also remain unchanged. This condition, which has intuitive appeal, is satisfied at most points by the competitive allocation mechanism.[16] It is closely related to Ritz's noncorruptibility condition; in fact, noncorruptibility implies nonbossiness. Within the private goods setting agent i affects agent j at a regular point $u \in \Omega^n$ if a $(v/i) \in [C^2(A)]^n$ exists such that $D_{(v/i)}F_j(u) \neq 0$. At each regular point the affects relation defines a binary relation among the agents; we write iH(u)j if agent i affects agent j at u.

The private-public goods version of Theorem 4.1 is this. If an allocation mechanism is broadly applicable, nonbossy, and strategy-proof, then at each regular point $u \in \Omega^n$ the affects relation H is acyclic. This means if agent i affects agent j, then no agent k (or sequence of agents) can exist who is affected by agent j and who in turn affects agent i. Thus the theorem states that agents can not mutually accomodate each other's preferences; all accomodation must consist of agents who rank lower on an exogenously given hierarchy adjusting to the preferences of those agents who rank higher on the hierarchy.

Serial dictatorship is an example of a strategy-proof mechanism that is nonbossy, broadly applicable, and--as the result requires--acyclic in the affects relation. The canonical serial dictatorship is the mechanism where agent 1 selects from an exogenously fixed feasible set Γ_1, agent 2 selects

from a feasible set $\Gamma_2(u_1)$ that depends on agent one's choice (or, equivalently, his utility function provided nonbossiness is respected), etc. Serial dictatorship is unattractive because the distribution of power is lopsided and, as Satterthwaite and Sonnenschein showed, the outcome generally violates Pareto optimality whenever the production possibility frontier is not linear.

5. Conclusions

This paper has used two approaches to examine the possibility of constructing stategy-proof (i.e., dominant strategy implementable) mechanisms. The first approach begins with the environment within which the mechanism is to be applied and then characterizes the strategy-proof mechanisms that are possible within it. In Section 2 we applied this approach to the most unstructured of environments: discrete alternatives and all preference orderings admissible. The main result for this environment is negative: if there are at least three alternatives, then all strategy-proof mechanisms that satisfy the Pareto criterion are dictatorial. In Section 4 we applied this approach to the structured environments found in economic models: the alternative set is a subset of Euclidean space and preferences are a priori restricted, for example, to be representable by a twice differentiable utility function. There we reported additional negative results.[17]

The second approach, which we employed in Section 3, is exactly the opposite of the first approach. In it we first specify the properties the mechanism should possess in addition to strategy-proofness and then characterize the environments in which that mechanism can exist. Substantial progress has been made in this area, though it difficult to characterize this

progress as either positive or negative. The positive aspect is that nondicatorial, strategy-proof mechanisms do exist for particular environments in which preferences are restricted only slightly. The negative aspect is that the environments for which these reasonable mechanisms do exist have, as of yet, no known relation to environments of the sort that naturally arise in economic models.

From the results that are presented and developed in this paper, we believe there are three main lessons that can be drawn. First, the theory of strategy-proof mechanisms is not a neatly finished body of knowledge. Numbers of interesting questions are still open. For example, on the technical side, the two approaches we have used in this paper need to be drawn together, i.e., how do the results of Section 3 relate to the results in Section 4? A tantalizing, but unexploited, connection is the parallel that exists between Ritz's noncorruptibility condition and Satterthwaite and Sonnenschein's nonbossiness condition. On the substantive side, very little work has been done on strategy-proofness in repetitive situations. Our intuition is that an important reason why individuals often choose not to misrepresent their desires in group decision situations is that they do not find it in their interest to acquire the reputation of a manipulator.

Second, with only one important exception, economic life is by and large not straightforward in the sense of always giving each agent a dominant strategy. Even though the theory as it currently stands is not absolutely conclusive concerning the impossibility of constructing strategy-proof mechanisms for economic environments, it has clearly established that strategy-proofness can only be achieved in certain environments and then only by using carefully designed mechanisms. Thus an economic agent in his individual optimizing behavior does generally find it in his interest to worry

about other agents' intentions and to play the game of trying to correctly anticipate their actions in planning his own actions even as they try to anticipate his actions in planning their actions. The exception to this generalization is the large number of agents case. For example, in an exchange economy that has a continuum of competitors, every agent is unable to influence prices, becomes a price taker, and finds it a dominant strategy to report his demand function accurately and without consideration of the demands that other agents are reporting. If, however, the number of agents is small, then each agent can affect the price and no longer has a dominant strategy. The demand function an agent wants to report then depends importantly on the demands other agents are expected to report.[18]

Finally, the theory of strategy-proof mechanisms has philosophical implications. Bok (1978, ch. 1) in her book that reviews and expands the ethical arguments extant against lying defines a lie to be an intentionally misleading statement. By this definition, in those situations where a group's decision process can usefully be represented by an allocation mechanism, an agent who misrepresents his preferences may sometimes legitimately be said to be lying. The impossibility results concerning strategy-proof mechanisms suggest that, no matter how well we redesign the social system, agents from time to time have an incentive to lie. This incentive is intrinsic to social mechanisms. It is as much a reflection of the imperfectability of society generally as it is of the imperfectness of society specifically. Therefore an individual's decision to be honest and not to lie is truly an ethical decision because, even in principle, society can not be designed so that honesty is self enforcing. The excuse that a lie is society's fault since its structure gave the liar the incentive to perpetrate his deception is empty because a society that gives no incentive to lie is logically inconceivable.

NOTES

1. An exception is Postlewaite (1979) who wrote about the incentives individuals may have to misrepresent their initial endowments.

2. This is the revelation principle in its original and simplest form.

3. If more than one element of B is maximal, then the max operator uses an arbitrary rule to pick one element from among the set of maximal elements.

4. We have defined an allocation mechanism to be a singlevalued function. This definition may appear to be a candidate for relaxation. For example, an allocation mechanism could be permitted to select as its output probability mixtures of two or more allocations that are contained in B, the set of feasible allocations. This relaxation, however, is an illusion because A, the set of conceivable outcomes, should be defined for such an allocation mechanism as all possible probability mixtures of the conceivable allocations, not simply as the set of conceivable allocations. Once this is done, then the allocation mechanism is again singlevalued and, unless preferences over this set of probability distributions are restricted, Theorem 2.1 continues to apply. For example, the assumption that each agent evaluates probability mixtures in accordance with a von Neumann-Morgenstern utility function is a strong restriction on agents' preferences. Two examples of papers that explore the consequences of permitting probability mixtures to be outcomes are Barbera (1977) and Gibbard (1978).

5. Blin and Satterthwaite (1978) discuss the parallels that exist between an individual's choices and a group's choices.

6. Blin and Satterthwaite (1978, Theorem 2) stated this result for the case of unrestricted preferences.

7. This result is stated in exactly this form in Blin and Satterthwaite (1978, Theorem 4). Its forebears include Satterthwaite (1975, Lemma 8), an intermediate result of Gibbard (1973), and Pattanaik (1973, Theorem 2).

8. See Blin and Satterthwaite (1978, Theorem 5). That particular proof was based on a proof of Schmeidler and Sonnenschein (1978), which in turn had been based on Gibbard's (1973) original proof of Theorem 1.1. All three of

these proofs of Theorem 2.1 have the common feature of using Arrow's theorem to create a contradiction. Satterthwaite's (1973, 1975) original proof of Theorem 2.1 and a second proof of Schmeidler and Sonnenschein (1978) are constructive and do not use Arrow's theorem. Thus the discussion of using Theorem 2.1 to prove Arrow's theorem is not empty.

9. See, for example, Sen and Pattanaik (1969). Their paper does not deal explicitly with strategy-proofness; rather it deals with the transitivity of majority rule. But, as the results of Section 3 show, if a mechanism is transitive and satisfies IIA as majority rule does over appropriately restricted domains, then it is also strategy-proof.

10. For a rational mechanism F defined on some Ω, strategy-proofness is equivalent to monotonicity and IIA. This may be seen in two steps. For the first step, suppose voter i can manipulate F at profile P and feasible set B by playing Q_i. There then exists a pair of alternatives, x and y, such that $x = F(P, B)$, $y = F(P/Q_i, B)$, and yP_ix. This means, because F is rational and has a social welfare function f_F underlying it, $xf_F(P)y$ and $yf_F(P/Q_i)x$. If yQ_ix then f_F violates IIA. If xQ_iy then f is nonmonotonic since, when agent i changes his reported preference from yP_ix to xQ_iy, the social ordering changes perversely for $xf(P)y$ to $yf(P/Q_i)x$. This proves that strategy-proofness implies monotonicity and IIA. For the second step, study the violation of IIA and the violation of monotonicity that are set up in the first step. Inspection shows that if either occurs, then agent i can manipulate F. Therefore, for rational mechanisms, monotonicity and IIA imply strategy-proofness. Blin and Satterthwaite (1978, Theorem 2) and Blair and Muller (1983) stated this result for unrestricted preferences and restricted preferences respectively.

11. Muller (1982) developed this graphical analysis.

12. Ritz's theorem is true as stated here only if Ω permits an agent to have the strict preference ordering (abc) over some three alternatives a, b, and c contained in A. This is a very weak condition that is satisfied by any interesting Ω.

13. For a given P_i many utility representations u_i exist. This indeterminancy has no effect on the results that we present in this section because the results are impossibility theorems.

14. Under the C^2 topology two utility functions, u and u', are close to each other if at every point within A their values are close, their vectors of first derivatives are close, and their matrices of second derivatives are

close.

15. Satterthwaite and Sonnenschein (1981, p. 591).

16. Note that nonbossiness is trivially satisfied in the public goods only case because every agent receives the same allocation. In the private goods only case the competitive mechanism satisfies nonbossiness except, as Mark Walker has privately pointed out, in special circumstances where a continuum of equilibria exist. Satterthwaite and Sonnenschein (1981) incorrectly assert that the competitive mechanism is nonbossy at all regular points.

17. Sections 2 and 4 neglected two well known cases of preference restrictions: single-peaked profiles and transferable utility. For the case of single-peaked profiles majority rule is strategy-proof. See Blin and Satterthwaite (1976). For the case of public goods in the presence of transferable utility Groves schemes are strategy-proof. A large literature exists on Groves schemes, e.g. Groves (1970), Clarke (1971), Groves and Loeb (1975), Green and Laffont (1979), and Holmstrom (1979). We have not included these two cases in this paper for reasons of space and because our judgement is that they are special cases that do not generalize.

18. Pazner and Wesley (1977, 1978) analyzed the properties of voting procedures for the large number of agents case. Roberts and Postlewaite (1976) investigated the incentive to become a price taker in an exchange economy as the number of agents increases.

REFERENCES

Arrow, K. J. 1963. Social choice and individual values. 2nd ed. New Haven: Yale University Press.

Barbera, S. 1977. The manipulability of social choice mechanisms that do not leave too much up to chance. Econometrica 45: 1573-89.

Blair, D. H. and E. Muller. 1983. Essential aggregation procedures on restricted domains of preferences. Journal of Economic Theory 30: 34-53.

Blin, J.-M. and M. A. Satterthwaite. 1976. Strategy-proofness and Single-Peakedness. Public Choice. 26: 51-58.

Blin, J.-M. and M. A. Satterthwaite. 1978. Individual decisions and group decisions: The fundamental differences. J. of Public Economics 10: 247-67.

Border, K. C. and J. S. Jordan. 1983. Straightforward elections, unanimity and phantom voters. Rev. of Econ. Studies 50 (January): 153-70.

Clarke, E. 1971. Multipart pricing of public goods. Public Choice 11: 17-33.

Feldman, A. 1979. Manipulating voting procedures. Economic Inquiry 17 (July): 452-74.

Gibbard, A. 1973. Manipulation of voting schemes: A general result. Econometrica 41: 587-602.

Gibbard, A. 1978. Straightforwardness of game forms with lotteries as outcomes. Econometrica 46: 595-614.

Green, J. R. and J.-J. Laffont. 1979. Incentives in public decision-making. Studies in public economics, vol. 1. Amsterdam: North-Holland.

Groves, T. 1970. The allocation of resources under uncertainty. Ph.D. diss., Univ. of Calif., Berkeley.

Groves, T. and M. Loeb. 1975. Incentives and public inputs. J. of Public Economics 4: 211-26.

Holmstrom, B. 1979. Groves' Scheme on Restricted Domains. Econometrica 47 (September): 1137-44.

Kalai, E. and E. Muller. 1977. Characterization of domains admitting nondictatorial social welfare functions and nonmanipulable voting procedures. Journal of Economic Theory 16: 457-69.

Kalai, E. and Z. Ritz. 1980. Characterization of the private domains

admitting Arrow social welfare functions. Journal of Economic Theory 22: 22-36.

Kim, K. and F. Roush. 1981. Effective nondictatorial domains. Journal of Economic Theory 24: 40-47.

Maskin, E. 1976. Social choice on restricted domains. Ph.D. dissertation, Harvard University.

Muller, E. 1982. Graphs and anonymous social welfare functions. International Economic Review 23: 609-22.

Pattanaik, P. 1973. On the stability of sincere voting situations. J. of Economic Theory 6: 558-74.

Pazner, E. A. and E. Wesley. 1977. Stability of social choice in inifinitely large societies. Journal of Economic Theory 14: 252-62.

Pazner, E. A. and E. Wesley. 1978. Cheatproofness properties of the plurality rule in large societies. Review of Economic Studies 45: 85-92.

Postlewaite, A. 1979. Manipulation via endowments. Review of Economic Studies 46: 255-62.

Ritz, Z. 1981. On Arrow social welfare functions and on nonmanipulable and noncorruptible social choice functions. Ph.D. dissertation, Northwestern University.

Ritz, Z. 1983. Restricted domains, Arrow social welfare functions and noncorruptible and nonmanipulable social choice correspondence: The case of private alternatives. Mathematical Social Sciences 2: 155-180.

Roberts, D. J. and A. Postlewaite. 1976. The incentive for price-taking behavior in large exchange economies. Econometrica 44: 115-28.

Satterthwaite, M. A. 1973. The existence of a strategy proof voting procedure: A topic in social choice theory. Ph.D. diss., Univ. of Wisconsin, Madison.

Satterthwaite, M. A. 1975. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. J. of Economic Theory 10 (April): 187-217.

Satterthwaite, M. A. and H. Sonnenschein. 1981. Strategy-proof allocation mechanisms at differentiable points. Review of Economic Studies 48: 587-97.

Schmeidler, D. and H. Sonnenschein. 1978. Two proofs of the Gibbard-Satterthwaite theorem on the possibility of a strategy-proof social choice function. In Proceedings of a conference on decision theory and social ethics at Schloss Reisenberg, ed. H. Gottinger and W. Ensler. Reidel Publishing Co.

Sen, A. and P. Pattanaik. 1969. Necessary and sufficient conditions for rational choice under majority decision. J. of Economic Theory 1: 178-202.

TABLE 2.1
Restrictions on F(•, A) Imposed by the Pareto Criterion

			AGENT 2					
			1 (xyz)	2 (xzy)	3 (yxz)	4 (yzx)	5 (zxy)	6 (zyx)
	1	(xyz)	x	x	$\neq z^{6}$	$\neq z^{5}$	$\neq y^{2}$	$?^{4}$
A	2	(xzy)	x	x	$\neq z^{7}$	$?^{8}$	$\neq y^{1}$	$\neq y^{3}$
G								
E	3	(yxz)	$\neq z$	$\neq z$	y	y	$?^{12}$	$\neq x^{9}$
N								
T	4	(yzx)	$\neq z$	?	y	y	$\neq x^{11}$	$\neq x^{10}$
1	5	(zxy)	$\neq y$	$\neq y$	?	$\neq x$	z	z
	6	(zyx)	?	$\neq y$	$\neq x$	$\neq x$	z	z

TABLE 2.2
Details of Feldman's Proof

Cell	Assigned Outcome	Alternative Outcome	Manip. Situation	Manip. Agent	Manip. Strategy	Manip. Outcome
2	x	z	F(1,5)=z	one	F(2,5) =	x
3	x	z	F(2,5)=x	two	F(2,6) =	z
4	x	y or z	F(1,6)=y or z	one	F(2,6) =	x
5	x	y	F(1,6)=x	two	F(1,4) =	y
6	x	y	F(1,6)=x	two	F(1,3) =	y
7	x	y	F(2,3)=y	one	F(1,3) =	x
8	x	y or z	F(2,4)=y or z	one	F(1,4) =	x
9	y	z	F(3,6)=z	one	F(2,6) =	x
10	y	z	F(4,6)=z	one	F(3,6) =	y
11	y	z	F(4,6)=y	two	F(4,5) =	z
12	y	x or z	F(3,5)=x or z	one	F(4.5) =	y

TABLE 3.1
Decisiveness Implications

	Domain Conditions		Implication		Name
i.	(abc) ε t(Ω)	(bca) ε t(Ω)	(ca)	(ba)	Direct Branch
ii.	(abc) ε t(Ω)	(bca) ε t(Ω)	(ab)	(ac)	Direct Branch
iii.	(abc) ε t(Ω)	(bca) $\notin$ t(Ω)	(ab) (bc)	(ac)	Joining Branch
iv.	(abc) ε t(Ω)	(bca) $\notin$ t(Ω)	(ca)	(ba) (cb)	Splitting Branch

TABLE 3.2

Nonrational, Strategy-proof Mechanism

AGENT 1

			1	2	3	4	5	6
			(xzyw)	(yzwx)	(yxwz)	(wxzy)	(zwxy)	(wzyx)
	1	(xzyw)	x	y	y	x	z	z
A	2	(yzwx)	y	y	y	y	y	y
G								
E	3	(yxwz)	y	y	y	y	y	y
N								
T	4	(wxzy)	x	y	y	w	z	w
2	5	(zwxy)	z	y	y	z	z	z
	6	(wzyx)	z	y	y	w	z	w

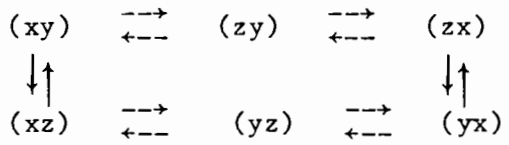

Figure 3.1. $\Omega_1 = \Sigma = \{xyz, yzx, zxy, zyx, yxz, xzy\}$.

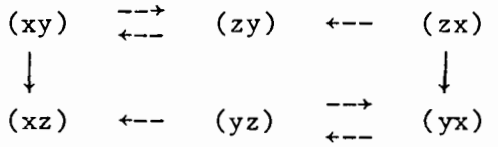

Figure 3.2. $\Omega_2 = \Sigma - \{zyx\} = \{xyz, yzx, zxy, yxz, xzy\}$.

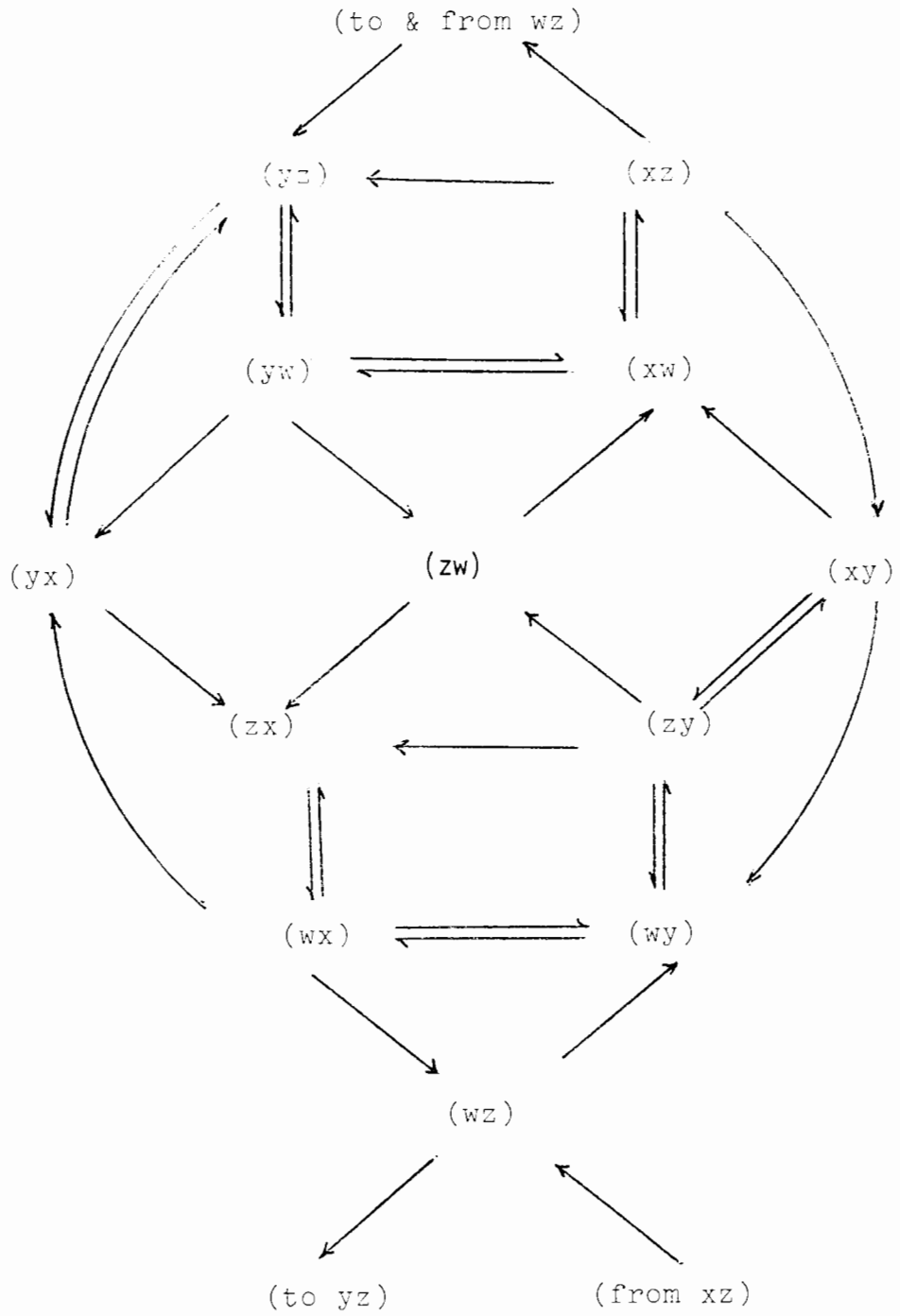

Figure 3.3

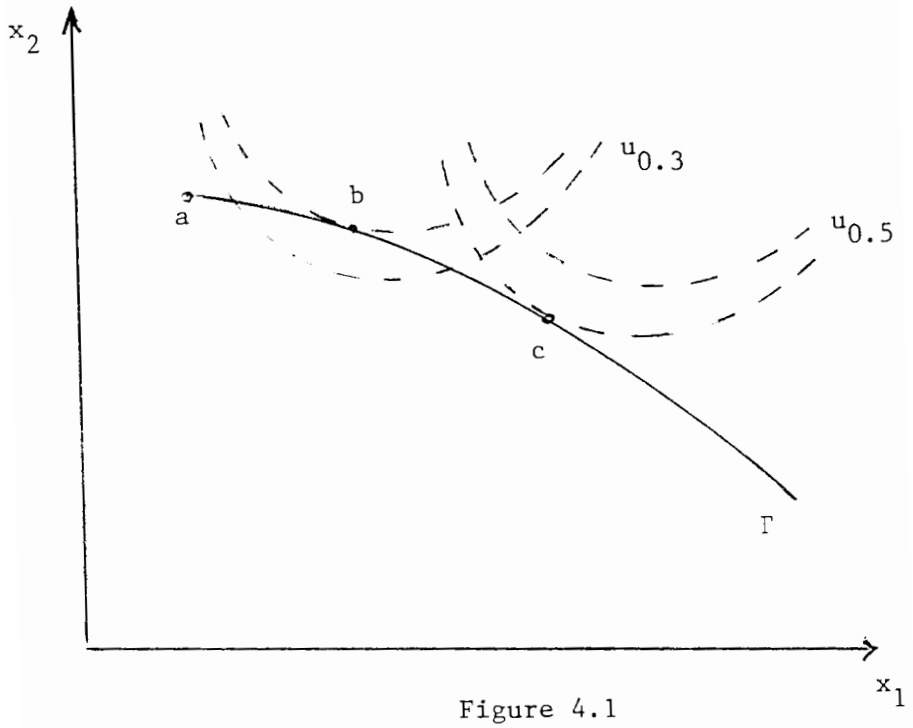

Figure 4.1

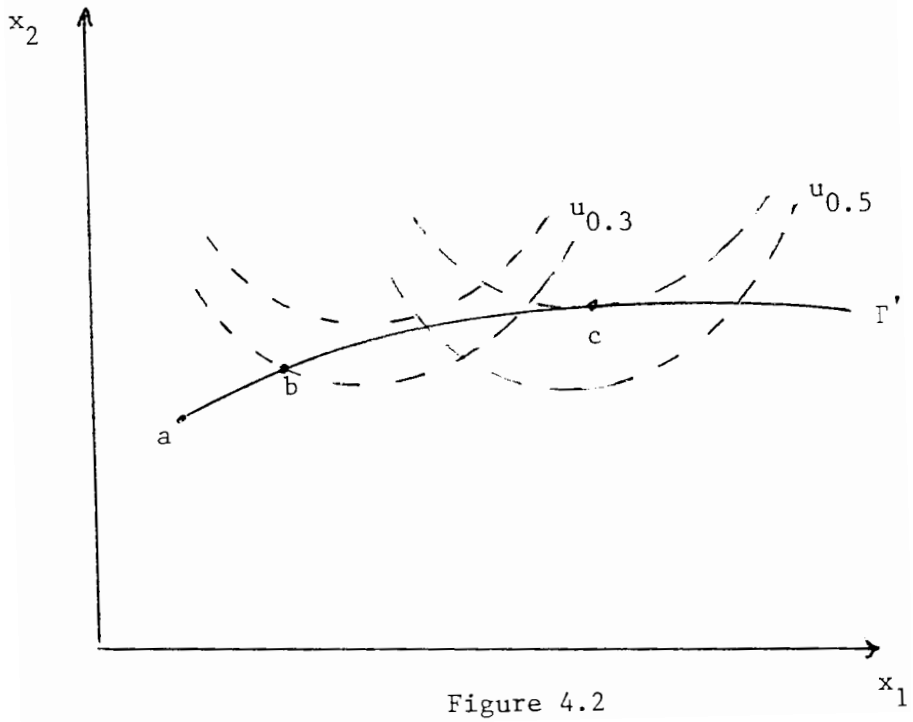

Figure 4.2

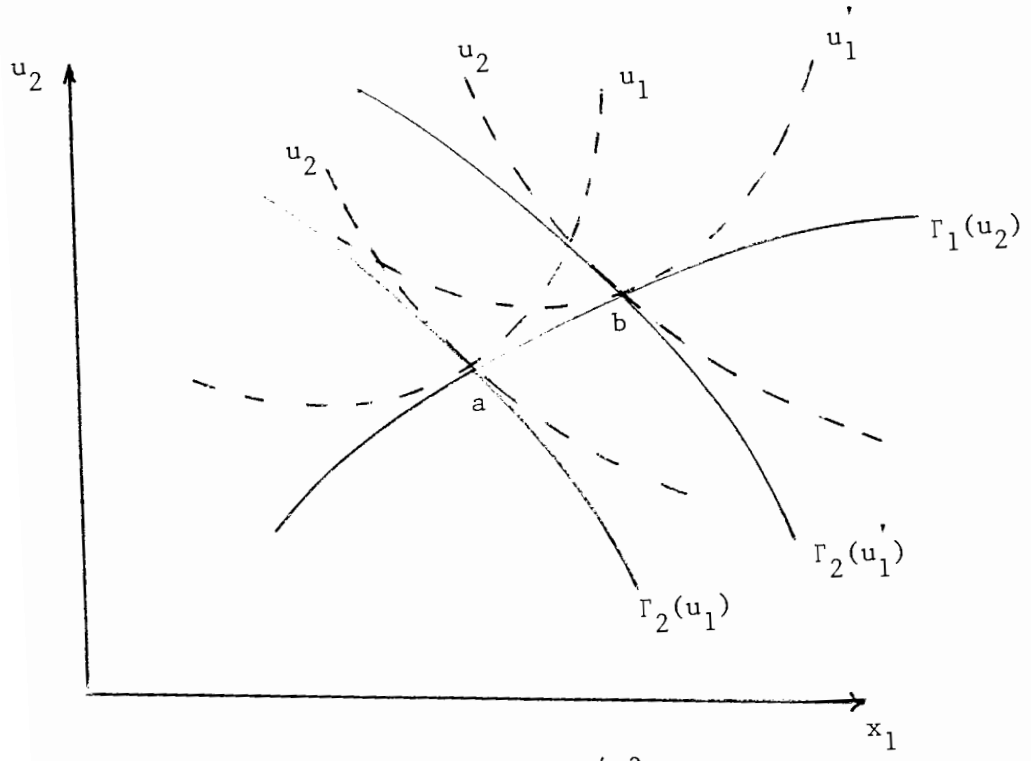

Figure 4.3

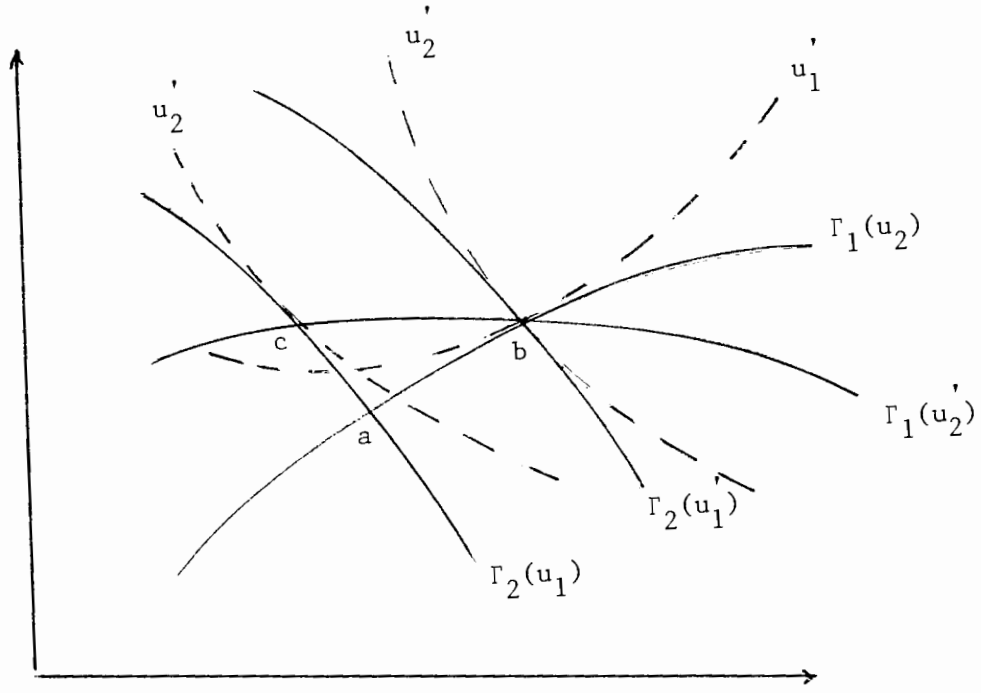

Figure 4.4