working paper
department
of economics

The Theory of Implementation in Nash Equilibrium: A Survey

Eric S. Maskin

Number 333

October 1983

massachusetts
institute of
technology
50 memorial drive
cambridge, mass. 02139

The Theory of Implementation in Nash Equilibrium: A Survey

Eric S. Maskin

Number 333 October 1983

D11

The Theory of Implementation in Nash Equilibrium:

A Survey

Eric S. Maskin

MIT

To appear in Social Goals and Social Organization: Volume in Memory of Elisha Pazner, Cambridge University Press.

June, 1983

Revised October, 1983

Financial support from the NSF and the A.P. Sloan Foundation is gratefully acknowledged. I wish to thank David Schmeidler and Hugo Sonnenschein for helpful comments.

The Theory of Implementation in Nash Equilibrium: A Survey

The theory of implementation concerns the problem of designing game forms (sometimes called "mechanisms" or "outcome functions") the equilibria of which have properties that are desirable according to a specified criterion of social welfare called a _social choice rule_. A game form, in effect, decentralizes decision-making. The social alternative is selected by the joint actions of all individuals in society rather than by a central planner.

Formally, a social choice rule assigns a set of alternatives to each profile of preferences (or other characteristics) that individuals in society might have; the set consists of the "welfare optima" relative to the preference profile. A _game form_ is a rule that specifies an alternative (or _outcome_) for each configuration of actions that individuals take. A game form _implements_ (technically, fully implements) a social choice rule if, for each possible profile of preferences, the equilibrium outcomes of the game form coincide with the welfare optima of the social choice rule. Of course, the equilibrium set depends on the particular solution concept being used. Implementation theory has considered a variety of solution concepts, including equilibrium in dominant strategies, Bayesian equilibrium, and Nash equilibrium. Other chapters of this volume treat the first two equilibrium concepts. In the main, this article is confined to implementation in Nash equilibrium, although it relates this theory to those of other solution concepts, dominant strategies in particular.

Nash equilibrium is the noncooperative solution concept _par excellence_, and so it is not surprising that implementation theory should have employed it extensively. Nonetheless, one reason often advanced for the desirability of decentralization is that information is incomplete, and so it may seem strange to use a solution concept of complete information (I am distinguishing here between Nash equilibrium in its original sense, c.f. Nash (1950), and the incomplete information extension due to Harsanyi (1967), commonly called "Bayesian equilibrium"). There are at least three alternative justifications for so doing.

First, as the work of Hurwicz (1972) and Groves and Ledyard (1977) at least implicitly assumes, a Nash equilibrium can be viewed as a stationary point of an iterative adjustment process. In such a process, players may have incomplete information but continually revise their actions until a point is reached where unilateral deviation no longer pays. Such a point is a Nash equilibrium.

There are several difficulties with this interpretation. If an individual believes that others play "naively" in the sense of always adjusting their actions optimally, assuming that the distribution of current actions will continue to prevail, then it will, in general, pay him to act as a Stackelberg leader and allow others to adapt to an action that he does not adjust. But if one or more players attempt to behave as Stackelberg leaders, there is no longer any reason to suppose that a stationary point of the process is a Nash equilibrium.

There are two cases where we might be able to rule out such Stackelberg behavior. One is where society is sufficiently large so that one individual's effect on others is slight enough as to have no appreciable effect on their actions. In that case, the individual would best play in "Nash-like" fashion (see, for example, Roberts and Postlewaite (1976)). The other is where the individual believes that any given iteration is the last (at least with very high probability), in which event, from his perspective, there is no opportunity for influencing future behavior.

Clearly, though, these cases are highly restrictive. When they do not apply, we cannot expect naive behavior. But if all individuals are "sophisticated" then each must realize that, when adjusting his action, he may affect others' (probabilistic) beliefs about his preferences. Since these beliefa, in turn, may affect <u>their</u> behavior, individuals may, again, be induced to behave in a non-Nash-like way.

The second reason for using Nash equilibrium is more satisfactory game theoretically. There are many circumstances where the planner (game form - designer) can be thought of as having highly incomplete information, whereas individuals themselves are well-informed. For example, the individuals may be firms that are experts in research and development and know a great deal about each other, whereas the planner may be the government, who knows next to nothing about R&D but wants to influence firms' behavior. Alternatively, the planner might be a "constitution-designer," who must devise the procedural rules (the game form) by which committee members make decisions long in advance of any particular application. Indeed, the planner may not literally exist as a physical entity; rather he may simply stand for the committee as a whole. But, by the time, any particular decision has to be made, committee members may be well aware of each other's preferences.

In either of these two examples, Nash equilibrium is the appropriate solution concept. It is important in the examples that individuals have good information about each other; otherwise, Bayesian rather than Nash equilibrium pertains. It is equally necessary that the planner have poor information; otherwise, he could simply impose a welfare optimal social alternative by fiat.

Finally, implementation in Nash equilibrium may be thought of as a <u>positive</u> theory. To the extent that the theory can characterize the set of implementable social choice rules, it can predict the kinds of outcomes that can rise as equilibria of already existing (complete information) games.

This article is divided into nine sections. The first introduces notation and the basic concepts. The second presents the fundamental theorem characterizing the set of implementable social choice rules. This theorem is cast in terms of two properties, monotonicity and weak no veto power. Section 3 discusses the so-called "revelation principle" with respect to implementation in Nash equilibrium and several other equilibrium concepts. We clarify the relevance for Nash implementation of the principle, as usually stated, and propose an alternative formulation. Section 4 discusses the connection between implementability and several common properties of social choice rules, viz., weak no veto power, neutrality, and individual rationality. Section 5 exposits the relationship between Nash and dominant strategy implementation. Section 6 treats implementation in a much-studied special case, where preferences are of a "quasilinear" form.

Through Section 6, all analysis assumes noncooperative behavior on the part of individuals. Section 7, however, allows for collusion and studies implementation in strong equilibrium. Section 8 considers an implementation concept, double implementation, that accommodates both noncooperative and cooperative behavior simultaneously. Finally, Section 9 briefly discusses two concepts related to Nash implementation.

1. Notation and Basic Concepts

Let A be a set of social alternatives (A can be either finite or infinite). A utility function, u, on A is a real-valued function

$u\colon A \to R,$

where R denotes the real numbers. Let U_A be the set of all utility functions. For each $i = 1,\ldots,n$, let U_i be a subset of U_A. Then, an n-person social choice rule (SCR) on $(U_1,\ldots,U_n)$ is a correspondence

$f\colon U_1 \times \ldots \times U_n \to A.$[1]

For any profile $(u_1,\ldots,u_n)$ of utility functions, one interprets $f(u_1,\ldots,u_n)$ (sometimes called the choice set and which we assume to be nonempty) as the set of welfare otpimal alternatives. Common examples of social choice rules include the Pareto correspondence, which selects all Pareto optima corresponding to a given profile, and the Condorcet correspondence, which selects all alternatives for which a majority does not prefer some other alternative. Notice that, in principle, we allow the SCR to select two different choice sets for two utility profiles that correspond to the same preference orderings. That is, the choice set may depend on cardinal properties of utility functions. This flexibility will be eliminated below when we discuss implementation. However, our formulation enables the crdinal nature of an implementable SCR to be proved (albeit trivially) rather than postulated.

Given action spaces $S_1,\ldots,S_n$ for each individual, an n-person game form g is a mapping

$g\colon S_1 \times \ldots \times S_n \to A.$

If individuals 1 through n play the action configuration $(s_1,\ldots,s_n)$, the outcome is alternative $g(s_1,\ldots,s_n)$.

For a game form g, let $NE_g(u_1,\ldots,u_n)$ be the set of Nash equilibrium

[1] In this chapter we shall suppose throughout that preferences alone constitute the relevant data about individuals. See the chapter by Postlewaite in this volume for a treatment that allows for other information (e.g., endowments) as well.

outcomes corresponding to the profile $(u_1,\ldots,u_n)$. Slightly diverging from the terminology of Dasgupta, Hammond, and Maskin (1979), we shall say that the game form g weakly implements the SCR f in Nash equilibrium if, for every $(u_1,\ldots,u_n) \in U_1\times\ldots\times U_n$

(1) $NE_g(u_1,\ldots,u_n)$ is nonempty

(2) $NE_g(u_1,\ldots,u_n) \subseteq f(u_1,\ldots,u_n)$.

Thus, if g weakly implements f, an equilibrium always exists, and all equilibria lie in the social choice set.

Requirements (1) and (2) are, by now, the standard requirements in Nash-implementation theory. We shall see below, however, that the analogue of (2) is not always imposed in the corresponding theories for other solution concepts.

If for all $(u_1,\ldots,u_n) \in U_1\times\ldots\times U_n$ and all $a \in f(u_1,\ldots,u_n)$ there exists a game form g that weakly implements f and for which $a \in NE_g(u_1,\ldots,u_n)$, then we say that f is implementable (in Nash equilibrium). The difference between weak and ordinary implementability is that the latter requires every element of every choice set to arise as a Nash equilibrium of some implementing game form. An ostensibly still stronger requirement is that a single game form yield all these equilibria. We shall say that the game form g fully implements the SCR f if for all $(u_1,\ldots,u_n) \in U_1\times\ldots\times U_n$,

(3) $NE(u_1,\ldots,u_n) = f(u_1,\ldots,u_n)$.

We shall see below (Section 4) that, in fact, implementability and full implementability are equivalent.

2. The Fundamental Characterization Theorem

To characterize those SCR's that are implementable, we must first define two properties of SCR's. We shall argue that the first of these is in many circumstances extremely weak.

Weak No Veto Power: An SCR f satisfies weak no veto power if, for all $(u_1,\ldots,u_n) \in U_1\times\ldots\times U_n$ and $a \in A$, $a \in f(u_1,\ldots,u_n)$ whenever there exists i such that for all $j \neq i$ and all $b \in A$ $u_j(a) \geq u_j(b)$.

In words, an SCR satisfies weak no veto power if whenever all individuals except possibly one agree that an alternative is top-ranked - i.e., no other alternative is higher in their preference orderings - then that alternative is in the social choice set; the remaining individual cannot veto it. The hypothesis that the alternative be top-ranked is what distinguishes this property from other no veto conditions and what makes it so weak. Indeed, in many circumstances the hypothesis cannot be satisfied at all. Suppose, for example, that we equate a social alternative with an allocation of goods across consumers. Assume also that at least one of these goods is a divisible private good that all individuals find desirable. Then no two individuals will agree that any given alternative is top-ranked, since each would like all the private good to himself. Thus if there are at least three individuals, our weak no veto power condition is satisfied vacuously.

Our other condition is considerably stronger, although quite standard. It sometimes goes under the name "strong positive association" (see Muller and Satterthwaite (1977) and Moulin and Peleg (1982)).

Monotonicity: An SCR f is monotonic if, for all $(u_1,\ldots,u_n)$, $(\overline{u}_1,\ldots,\overline{u}_n) \in U_1\times\ldots\times U_n$ and $a \in A$, $a \in f(\overline{u}_1,\ldots,\overline{u}_n)$ whenever (i) $a \in f(u_1,\ldots,u_n)$ and, (ii) for all $b \in A$ and i, $u_i(a) \geq u_i(b)$ implies $\overline{u}_i(a) \geq \overline{u}_i(b)$.

In words, an SCR is monotonic if, whenever an alternative a is in the choice set for a profile of preferences, and then those preferences are altered in a way such that a does not fall in anyone's preference ordering relative to any other alternative, it remains in the choice set.

Clearly, monotonicity is a purely ordinal property, and an SCR that satisfies it will reflect only ordinal properties of utility functions. That is, if, for all i, $\bar{u}_i = h_i \circ u_i$, where $h_i : R \to R$ is strictly increasing, then a monotonic f satisfies $f(u_1,\ldots,u_n) = f(\bar{u},\ldots,\bar{u}_n)$. Thus monotonicity rules out the interpersonal comparisons inherent in, say, utilitarianism or the Rawlsian difference principle. Moreover, as we shall see below (see section 5), it amounts to something very close to independence of irrelevant alternatives in the sense of Arrow (1951). Nonetheless it is satisfied by such common SCR's as the Pareto and Condorcet correspondences and, in economic contexts, by the correspondence that selects core allocations.

Monotonicity does not require that all Pareto optimal alternatives be in the choice set (the Condorcet correspondence is a counterexample), but, if f is onto A, it does imply that a <u>subset</u> of Pareto optimal alternatives is in the choice set, namely, those that are top-ranked by all individuals:

<u>Lemma 1</u>: Suppose that f is monotonic and onto A. For any $(\bar{u}_1,\ldots\bar{u}_n) \varepsilon U_1 \times \ldots \times U_n$ and $a \varepsilon A$ if, for all b and i, $\bar{u}_i(a) \geq \bar{u}_i(b)$, then $a \varepsilon f(\bar{u}_1,\ldots,\bar{u}_n)$.

<u>Proof</u>: Because, by assumption, f is onto A, there exists $(u_1,\ldots,u_n) \varepsilon U_1 \times \ldots \times U_n$ such that $a \varepsilon f(u_1,\ldots,u_n)$. If, for all i and b, $\bar{u}_i(a) \geq \bar{u}_i(b)$, then, from monotonicity, $a \varepsilon f(\bar{u}_1,\ldots,\bar{u}_n)$.

Q.E.D.

We can now state the fundamental characterization result.

Theorem 1: (Maskin (1977)): Suppose that f is an n-person SCR. If f is implementable in Nash equilibrium, then it is monotonic. Furthermore, if $n \geq 3$ and f satisfies weak no veto power and monotonicity, then it is fully implementable.

Proof: To see that implementability implies monotonicity, suppose that f is not monotonic. Then there exist $(u_1,\ldots,u_n)$ and $(\bar{u}_1,\ldots,\bar{u}_n) \in U_1\times\ldots\times U_n$ and $a \in A$ such that $a \in f(u_1,\ldots,u_n)$ and, for all $b \in A$ and all i,

(4) $u_i(a) \geq u_i(b)$ implies $\bar{u}_i(a) \geq \bar{u}_i(b)$

but

(5) $a \notin f(\bar{u}_1,\ldots,\bar{u}_n)$.

Now, if f is implementable, there exists a game form $g: S_1\times\ldots\times S_n \to A$ and a configuration of strategies $(s_1^*,\ldots,s_n^*)$ such that $g(s_1^*,\ldots,s_n^*) = a$ and $(s_1^*,\ldots,s_n^*)$ is a Nash equilibrium for profile $(u_1,\ldots,u_n)$. But from (4), $(s_1^*,\ldots,s_n^*)$ is also a Nash equilibrium for $(\bar{u}_1,\ldots,\bar{u}_n)$, which, in view of (5), contradicts (2). Hence, f is not implementable.

We only sketch the proof that weak no veto power and monotonicity imply that f is fully implementable. For the omitted details see Maskin (1977). For any $a \in A$ and $u_i \in U_i$ let

$$L(a,u_i) = \{b \in A \mid u_i(a) \geq u_i(b)\}.$$

$L(a,u_i)$ is the lower contour set of u_i at a, i.e., the set of alternatives that someone with utility function u_i does not prefer to a. For each i, let

(6) $S_i = \{(u_1,\ldots,u_n,a) \mid (u_1,\ldots,u_n) \in U_1\times\ldots\times U_n \text{ and } a \in f(u_1,\ldots,u_n)\}$.

That is, each player's action consists of announcing a profile of utility functions and an alternative that is in the choice set with respect to that profile. Define $g\colon S_1\times\ldots\times S_n \to A$ so that:

(7) if $\bar{s}_1 = \ldots = \bar{s}_n = (u_1,\ldots,u_n,a)$, then $g(\bar{s}_1,\ldots,\bar{s}_n) = a$;

(8) if $\bar{s}_j = (u_1,\ldots,u_n,a)$ for all $j \neq i$, then

$\{b \in A | b = g(s_i,\bar{s}_{-i}),\ s_i \in S_i\} = L(a,u_i)$[2]

and

(9) if, for given i, there exist j and k, with $j \neq i \neq k$, such that $\bar{s}_j \neq \bar{s}_k$, then

$\{b \in A | b = g(s_i,\bar{s}_{-i}),\ s_i \in S_i\} = A.$

That there exist game forms satisfying conditions (6)-(9) is demonstrated in Maskin (1977). We claim that any such game form fully implements f.

To see this, first choose $(u_1,\ldots,u_n) \in U_1\times\ldots\times U_n$ and $a \in f(u_1,\ldots,u_n)$. From (7), if all individuals take the action $(u_1,\ldots,u_n,a)$, the outcome is a. Furthermore if $(u_1,\ldots,u_n)$, in fact, <u>are</u> individuals' utility functions, then, from (8), each individual cannot obtain an alternative he prefers to a by varying his action unilaterally. Hence, all individuals' taking the action $(u_1,\ldots,u_n,a)$ is a Nash equilibirum for the profile $(u_1,\ldots,u_n)$. This establishes that for all $(u_1,\ldots,u_n)$, $f(u_1,\ldots,u_n) \subseteq NE_g(u_1,\ldots,u_n)$.

To establish the opposite inclusion, suppose that $(\bar{s}_1,\ldots,\bar{s}_n)$ is a Nash equilibirum of g for the profile $(\bar{u}_1,\ldots,\bar{u}_n)$ and that $a = g(\bar{s}_1,\ldots,\bar{s}_n)$. We

---

[2] The notation "$g(s_i,\bar{s}_{-i})$" is shorthand for $g(\bar{s}_1,\ldots,\bar{s}_{i-1},s_i,\bar{s}_{i+1},\ldots,\bar{s}_n)$.

must establish that $a \in f(\bar{u}_1,\ldots,\bar{u}_n)$. There are three cases to consider: (α) $\bar{s}_1 = \ldots = \bar{s}_n$; ($\beta$) there exist i and action s such that for all $j \neq i$ $\bar{s}_j = s$ but $\bar{s}_i \neq s$; and (γ) all other configurations.

Consider case (α) first. Suppose that $\bar{s}_i = (u_1,\ldots,u_n,a)$ for all i. We have already observed that, from (7) and (8), $g(\bar{s}_1,\ldots,\bar{s}_n) = a$ and that $(\bar{s}_1,\ldots,\bar{s}_n)$ is a Nash equilibrium for the profile $(u_1,\ldots,u_n)$. For any i consider b such that $u_i(a) \geq u_i(b)$, i.e., such that $b \in L(a,u_i)$. From (8) there exists $s_i \in S_i$ such that $g(s_i,\bar{s}_{-i}) = b$. Hence $\bar{u}_i(a) \geq \bar{u}_i(b)$; otherwise, $\bar{s}_i$ could not be an equilibrium action for utility function $\bar{u}_i$, contrary to our assumption. Therefore, the hypotheses of the monotonicity condition are satisfied, and we conclude that $a \in f(\bar{u}_1,\ldots,\bar{u}_n)$, as required.

Next, consider case (β). Suppose that, for all $j \neq i$, $\bar{s}_j = (u_1,\ldots,u_n,a)$ and that $\bar{s}_i \neq (u_1,\ldots,u_n,a)$. Since, for each $k \neq i$, $\bar{s}_k \neq \bar{s}_i$ and $n \geq 3$, (9) implies that, for all $j \neq i$ and all $b \in A$, there exists $s_j \in S_j$ such that $g(s_j,\bar{s}_{-j}) = b$. Hence, because $(\bar{s}_1,\ldots,\bar{s}_n)$ was assumed to be a Nash equilibrium for $(\bar{u}_1,\ldots,\bar{u}_n)$, we can conclude that $\bar{u}_j(a) \geq \bar{u}_j(b)$ for all $j \neq i$ and all $b \in A$. Our weak no veto power condition then implies that $a \in f(\bar{u}_1,\ldots,\bar{u}_n)$, as required.

Finally, in case (γ), for all i, there exist j and k, with $j \neq i \neq k$, such that $\bar{s}_j \neq \bar{s}_k$. Hence, as in case (β), weak no veto power implies that $a \in f(\bar{u}_1,\ldots,\bar{u}_n)$, completing the proof.

Q.E.D.

The proof of Theorem 1 is constructive. Given an SCR satisfying weak no veto power and monotonicity, we produce a game form that fully implements it. It may be helpful to summarize the construction in words. An action consists of announcing a profile of utility functions and an alternative that is in the choice set for that profile. Condition (7) says that if all individuals announce the same profile $(u_1,\ldots,u_n)$ and alternative a, then a is the outcomme. Condition (8) says that if all individuals but one play the same action $(u_1,\ldots,u_n,a)$, then, by varying his action, the remaining individual can "trace out" the entire lower contour set corresponding to the utility function the others announce for him and to the alternative that they announce. Condition (9) stipulates that if, in a configuration of actions, two individuals' actions differ, then any third individual can trace out the entire set A by varying his action.

As we have noted, the Pareto correspondence is monotonic. Also, it obviously satisfies weak no veto power. Theorem 1 implies, therefore, that the Pareto correspondence is implementable for $n \geq 3$, even when the U_i's are unrestricted (i.e., equal to U_A). This result, however, does not obtain when n = 2, as Theorem 2 demonstrates.

<u>Pareto Optimality</u>: An SCR $f: U_1 \times\ldots\times U_n \to A$ is Pareto optimal if for all $(u_1,\ldots,u_n) \in U_1\times\ldots\times U_n$ and all $a \in f(u_1,\ldots,u_n)$, a is weakly Pareto optimal with respect to $(u_1,\ldots,u_n)$, i.e., there does not exist $b \in A$ such that, for all i, $u_i(b) > u_i(a)$.

<u>Dictatorship</u>: An SCR $f: U_1\times\ldots\times U_n \to A$ is dictatorial if there exists an

individual i such that, for all $(u_1,\ldots,u_n) \in U_1\times\ldots\times U_n$ and all $a \in A$, $u_i(a) \geq u_i(b)$ for all $b \in A$ if $a \in f(u_1,\ldots,u_n)$. That is, an SCR is dictatorial if there exists an individual (the dictator) who always gets his way.

Theorem 2: Let $f\colon U_A \times U_A \to A$ be a two-person, Pareto optimal SCR. Then f is implementable in Nash equilibrium if and only if f is dictatorial.

Proof: See Maskin (1977) and Hurwicz and Schmeidler (1978).

The hypothesis that the U_i's are equal to U_A is crucial to the validity of Theorem 2. As we shall see in Section 7, many two-person, Pareto optimal, and nondictatorial SCR's on restricted domains are implementable.

Given a set of SCR's satisfying the hypotheses of Theorem 1, we can generate new implementable SCR's:

Corollary to Theorem 1: For $n \geq 3$, suppose that $\{f_1,f_2,\ldots\}$ is a sequence of n-person monotonic SCR's. Then, if one of the f_i's satisfies no veto power $\bigcup_{i=1}^{\infty} f_i$ is fully implementable in Nash equilibrium, and if each of f_i's satisfies weak no veto power $\bigcap_{i=1}^{\infty} f_i$ is fully implementable (assuming $\bigcap_{i=1}^{\infty} f_i(u_1,\ldots,u_n)$ is nonempty for all profiles).

Proof: The proof simply consists of verifying that $\bigcup_{i=1}^{\infty} f_i$ and $\bigcap_{i=1}^{\infty} f_i$ both satisfy monotonicity, that $\bigcup_{i=1}^{\infty} f_i$ satisfies weak no veto power if one of f_i's does, and that $\bigcap_{i=1}^{\infty} f_i$ satisfies weak no veto power if all the f_i's do.

3. The Revelation Principle

Let us temporarily broaden the idea of an SCR. Rather than limiting its domain to sets of utility functions, we shall define it to be a correspondence on $\Theta_1 \times \ldots \times \Theta_n$, where Θ_i is individual i's space of possible "characteristics." A characteristic θ_i not only describes i's preferences, but perhaps also his endowment, information about others, and whatever else might be relevant.

Suppose that the SCR $f: \Theta_1 \times \ldots \times \Theta_n \to A$ is weakly implemented by a game form $g: S_1 \times \ldots \times S_n \to A$ according to some noncooperative solution concept. Thus we require the analogues of (1) and (2) to hold for the solution concept under consideration. Because the solution concept is noncooperative, we can write each individual's equilibrium action as a function $s_i^*(\theta_i)$ of his characteristic. Hence, for all profiles $(\theta_1, \ldots, \theta_n)$, $(s_1^*(\theta_1), \ldots, s_n^*(\theta_n))$ is an equilibrium. Now, define the induced game form

$$g^*: \Theta_1 \times \ldots \times \Theta_n \to A$$

so that, for all $(\theta_1, \ldots, \theta_n)$,

$$g^*(\theta_1, \ldots, \theta_n) = g(s_1^*(\theta_1), \ldots, s_n^*(\theta_n)).$$

Notice that for all $(\theta_1, \ldots, \theta_n)$, the <u>actions</u> $(\theta_1, \ldots, \theta_n)$ constitute an equilibrium[3] for the profile $(\theta_1, \ldots, \theta_n)$ and that, furthermore, $g^*(\theta_1, \ldots, \theta_n) \in f(\theta_1, \ldots, \theta_n)$. This is the revelation principle (see Gibbard (1973), Dasgupta, Hammond, and Maskin (1979), Myerson (1979), (1982), and (1983) and the references cited in this last paper): the observation that

[3] Actually this assertion is a bit too strong. It is true only for solution concepts that have the property that an individual's best action does not change when one deletes from the action spaces of other individuals all actions that are never equilibirum actions for any possible characteristic they might have. This property holds for dominant strategy, Bayesian, and Nash equilibrium, but not for, say, maximin equilibrium. However it does hold for a modified version of maximin equilibrium (see Dasgupta, Hammond, and Maskin (1979)).

if a game form implements an SCR, then there exists a "direct revelation" game form whose action spaces coincide with the characteristic spaces and which has the properties that (1) playing one's true characteristic is always an equilibrium action and (2) such a "truth-telling" equilibrium is in the choice set.

Although the revelation principle is a useful technical device, we must stress that g^* does <u>not</u> necessarily implement f. That is because, although $g^*(\theta_1,\ldots,\theta_n)$ is in the choice set for $(\theta_1,\ldots,\theta_n)$, there may be other equilibrium outcomes that are not, even if g (the original game form) <u>does</u> implement f.

Thus, we cannot conclude from the revelation principle that all one ever need consider are direct revelation game forms. Unfortunately, one may draw that incorrect conclusion from reading much of the literature on implementation in dominant and Bayesian equilibria. For the most part, this literature has implicitly used an implementation concept different from (the analogue of) (1) and (2), viz., namely "truthful implementation"[4] which requires only that the truthful equilibrium of a direct revelation game form be in the choice set. Although the connection between truthful and ordinary implementation has been (partially) elucidated for the case of dominant strategy equilibrium, almost nothing is known about it for Bayesian equilibrium. In any case, the Nash implementation theory is the sole implementation literature where much attention has been given to the issue of <u>multiple</u> equilibria. Indeed that is the aspect that lends the literature interest, since for any SCR, it is extremely easy to contruct a direct revelation game form for which, for each profile, the truthful equilibrium

[4] See Dasgupta, Hammond, and Maskin (1979), Laffont and Maskin (1982a), and Sections 5 and 6 below.

is in the choice set. All we have to do is satisfy (7), which is possible for <u>any</u> SCR.

There <u>is</u>, nonetheless, a version of the revelation principle that is consistent with our definition of Nash implementation. When Nash equilibrium is the solution concept, an individual needs to know not just his <u>own</u> preferences but the preferences of everyone else in order to determine his equilibrium action. Therefore, in the framework of Sections 1 and 2, a characteristic of an individual is an entire profile of utility functions. Indeed, if instead we interpreted individual i's characteristic to be u_i alone, we would, in effect, be requiring dominant strategies (see Theorem 7.1.1 of Dasgupta, Hammond, and Maskin (1979)).

Notice that having individuals announce utility profiles is, essentially, what the game forms in the proof of Theorem 1 do (individuals also announce alternatives, but that is only because f may be multivalued; if f were single-valued, the strategy spaces could be taken to be $U_1 \times \ldots \times U_n$). Thus these game forms may be thought of as ones of direct revelation. Now, as we shall see in Section 4, not all implementable SCR's satisfy weak no veto power. Therefore, Theorem 1 does not quite completely characterize the set of implementable SCR's. Nevertheless, the kind of game form constructed in the proof, only slightly modified, is capable of fully implementing any SCR that can be implemented at all. Thus, in this sense, we need consider only a "canonical" class of SCR's.

Suppose that f is an implementable SCR. For each i and $\tilde{u}_i \in U_i$ let $N_i(\tilde{u}_i) = \{a \in A |$ there exists $\tilde{u}_{-i}$ such that, for all $j \neq i$ and all $b \in A$, $\tilde{u}_j(a) \geq \tilde{u}_j(b)$ but $a \notin f(\tilde{u}_1, \ldots, \tilde{u}_n)\}$. That is, the set $N_i(\tilde{u}_i)$ consists of all the alternatives a that individual i can veto if he has utility function $\tilde{u}_i$ even if a is a top-ranked alternative for everyone else. Clearly, $N_i(\tilde{u}_i)$ is empty if f satisfies weak no veto power. As in the proof of Theorem 1, let

(6) $S_i = \{(u_1,\ldots,u_n,a) | (u_1,\ldots,u_n) \in U_1\times\ldots\times U_n \text{ and } a \in f(u_1,\ldots,u_n)\}$.

Define $g\colon S_1\times\ldots\times S_n \to A$ to satisfy (7),

(8*) if $\bar{s}_j = (u_1,\ldots,u_n,a)$ for all $j \neq i$, then

$\{b \in A | b = g(s_i,\bar{s}_{-i}),\ s_i \in S_i\} = L(a,u_i) - M_i(a,u_i)$,

where $M_i(a,u_i) = \{b \in A|$ there exists $\tilde{u}_i \in U_i$ such that $b \in N_i(\tilde{u}_i)$ and $\tilde{u}_i(b) \geq \tilde{u}_i(c)$ for all $c \in L(a,u_i)\}$, and

(9*) if, for given i, there exist j and k, with $j \neq i \neq k$, such that $S_j \neq S_k$, then

$\{b \in A|\ b = g(s_i,\bar{s}_{-i}),\ s_i \in S_i\} = A - P$,

where $P = \{a \in A|$ there exists $(u_1,\ldots,u_n)$ such that $u_i(a) \geq u_i(b)$ for all i and b but $a \notin f(u_1,\ldots,u_n)\}$. From Lemma 1, if $a \in P$, then a is not in the range of f. Therefore P is empty if f is onto A. To see that such a construction is possible, see Maskin (1977).

Condition (8*) says that if all individuals but i take the same action $(u_1,\ldots,u_n,a)$, then, by varying his action, i can trace out the lower contour set corresponding to u_i and a except for those alternatives b for which there exists a profile $(\bar{u}_1,\ldots,\bar{u}_n)$ such that (α) b is top-ranked by all individuals other than i, (β) individual i (with utility funciton $\bar{u}_i$)

prefers b to all alternatives in the lower contour set corresponding to $L(a,u_i)$, and (γ) b is not in the choice set corresponding to $(\bar{u}_1,..,\bar{u}_n)$. Condition (9*) requires that if, in a configuration of actions, two individuals' actions differ, then any third individual, by varying <u>his</u> action, can trace out the entire set A except for those alternatives a for which there exists a profile in which a is top-ranked by everyone but not in the choice set.

<u>Theorem 3</u>: The Revelation Principle: Suppose that, for $n \geq 3$, f is an n-person SCR that is implementable in Nash equilibrium. Then a game form satisfying (6), (7), (8*), and (9*) exists. Furthermore, f is fully implementable by any such game form.

For the details of the proof, see Maskin (1977). Here we give only an indication of the idea behind the proof by way of an example.

The construction in Theorem 1 will not serve to implement all implementable SCR's. This is because an implementable SCR may fail to satisfy weak no veto power (however some implementable SCR's that violate weak no veto power <u>can</u> be implemented by the Theorem 1 construction, e.g., the individual rationality correspondence of Section 4 below). For example, consider the SCR f that chooses alternative c as optimal unless c is Pareto dominated. If b Pareto dominates c, b is chosen, unless a, in turn, Pareto dominates b, in which case a is chosen. This SCR is clearly monotonic, but it does not satisfy weak no veto power because if individuals 2 and 3 (in a three-person society) both prefer a to b and b to c, and individual 1 prefers b to a and a to c, then b is chosen, even though two out of three individuals top-rank a. Moreover, the construction of Theorem 1 does not implement the SCR.

To see this, suppose, for instance, that individuals' preferences are as just described. However, suppose, in the Theorem 1 construction, that individuals 2 and 3 both play the strategy consisting of announcing the profile

	1	2	3
	b	c	c
(*)	c	a	a
	a	b	b

and the alternative c. If individual 1 does the same, then the outcome is c, since this is the f-optimal alternative. By playing some alternative strategy s', furthermore, individual 1 can obtain alternative a, since a lies in the lower contour set of 1's preference ordering as specified by (*). Individual 1 cannot, however, obtain alternative b. Therefore, a strategy triple where individual 1 plays s' and individuals 2 and 3 each play (*) is a Nash equilibrium with respect to individuals' (true) preferences. Because the corresponding outcome, a, is not optimal for those preferences, we conclude that the game form does not implement f.

However, f is implementable by a game form satisfying (6), (7), (8*), and (9*). Specifically, (8*) guarantees that a non-optimal equilibrium as above cannot arise because, starting from a configuration where all individuals play the same strategy, an individual cannot trace out the whole lower contour set and, in particular, cannot obtain, for any profile of preference, any alternative that is top-ranked by all others and, within his lower contous set, top-ranked for him. Thus, in the example, if individuals 2 and 3 play (*), individual 1 cannot obtain a (in this example, we did not have to invoke (9*), which applies only to SCR's that permit non-Pareto optimal outcomes).

Notice that Theorem 3 establishes that implementability implies full implementability, as we claimed earlier. The theorem can be used to extend the corollary to Theorem 1 to the case of SCR's that do not necessarily satisfy weak no veto power.

Corollary 1: Suppose that, for $n \geq 3$, $f_1, f_2, \ldots$ is a sequence of monotonic SCR's. Suppose one of the f_i's is implementable in Nash equilibrium. Then $\bigcup_{i=1}^{\infty} f_i$ is implementable also.

It remains an open question whether $\bigcap_{i=1}^{\infty} f_i$ is necessarily implementable. However, a case in which the intersection of two implementable SCR's is implementable is where one of the f_i's is the Pareto correspondence.

Corollary 2: For $n \geq 3$, if f_1 is an implementable SCR and f_2 is the Pareto correspondence, then $f_1 \cap f_2$ is implementable if it is nonempty for all profiles.

Closely related to Corollary 2 is the observation that the "Pareto frontier" of an implementable SCR is implementable.

Pareto Frontier of an SCR: The Pareto frontier of an SCR f is the SCR $PF(f)(u_1, \ldots, u_n) = \{a \in f(u_1, \ldots, u_n) | \text{for all } b \in f(u_1, \ldots, u_n), u_i(a) \geq u_i(b)$ for some $i\}$.

Corollary 3: For $n \geq 3$, if f is an implementable SCR, then the Pareto frontier PF(f) is also implementable.

The proofs of Corollaries 1-3 are straightforward applications of Theorem 3 (see Maskin (1977)).

4. No Veto Power, Individual Rationality, and Neutrality

We have already mentioned that weak no veto power is not necessary for implementability. One prominent example of an implementable SCR that violates this property is the individual rationality correspondence. Let a_0 be an element of A. We interpret a_0 to be the "status quo." The individual rationality correspondence, f_{IR}, selects all alternatives that weakly Pareto dominate a_0, i.e.

$$f_{IR}(u_1,\ldots,u_n) = \{a \in A | u_i(a) \geq u_i(a_0) \text{ for all } i\}.$$

Clearly, f_{IR} does not satisfy weak no veto power on all domains of utility functions, because every individual must be guaranteed at least the utility he derives from a_0. Nonetheless it is a simple matter to fully implement f_{IR}. For instance, the construction of Theorem 1 will do the trick. For a simpler example, let $S_i = A$ for all i. Define the game form $g: S_1\times\ldots\times S_n \to A$ so that

$$(10)\quad g(s_1,\ldots,s_n) = \begin{cases} s, \text{ if } s_1 = \ldots = s_n \\ \\ a_0, \text{ otherwise} \end{cases}$$

That is, each individual chooses an alternative as an action. If the alternatives agree, the common alternative is the outcome; otherwise, a_0 is the outcome. It is immediate that g fully implements f_{IR}. Notice that this is true even for n = 2.

The SCR f_{IR} is implementable not only by itself but in conjunction with other implementable SCR's.

<u>Corollary 4 to Theorem 3</u>: Suppose that, for $n \geq 3$, f is an n-person SCR that is implementable. Then $f \cap f_{IR}$ is implementable too.

The individual rationality correspondence is highly "non-neutral"; it treats the alternative a_0 very differently from all others. But, just as it is implementable, so is any neutral and monotonic SCR.

<u>Neutrality</u>: An SCR $f\colon U_A \times \ldots \times U_A \to A$ is neutral if for any permutation π: $A \to A$ and any profile $(u_1,\ldots,u_n)$

$$f(u_1 \circ \pi,\ldots,u_n \circ \pi) = \pi \circ f(u_1,\ldots,u_n).$$

Neutrality simply says that an alternative's labelling is irrelevant. Notice that in the formal statement, we have defined f on the unrestricted domain. This is to ensure that f is defined on the permutation profile $(u_1 \circ \pi,\ldots,u_n \circ \pi)$. The following result is another simple application of Theorem 3.

<u>Theorem 4</u> (Maskin (1977)): For $n \geq 3$, an n-person SCR that is monotonic and neutral is implementable in Nash equilibrium.

Theorem 4 and Corollary 4 raise the question of whether weak no veto power is a redundant condition for implementability when $n \geq 3$. In fact, the following example demonstrates that it is not, by exhibiting a three-person monotonic SCR that is not implementable.

<u>Example 1</u> (Maskin (1977)) A nonimplementable, monotonic SCR: Let $n = 3$ and $A = \{a,b,c\}$. For each i, let U_i consist of all utility functions corresponding to strict preference orderings (i.e., $u_i(a) = u_i(b)$ implies $a = b$). Define the SCR $f\colon U_1 \times U_2 \times U_3 \to A$ so that for all $(u_1,u_2,u_3) \in U_1 \times U_2 \times U_3$ and all $x, y \in A$, $x \in f(u_1,u_2,u_3)$ if and only if

(11) x is Pareto optimal

(12) if $x \in \{a,b\}$ $u_1(x) > u_1(y)$ for all $y \neq x$

(13) if $x = c$, there exists $y \in A$ such that $u_1(x) > u_1(y)$.

It is easy to see that f is monotonic. Choose (u_1^*,u_2^*,u_3^*), $(u_1^{**},u_2^{**},u_3^{**})$, and $(u_1^{***},u_2^{***},u_3^{***}) \in U_1 \times U_2 \times U_3$ so that

$$u_1^*(b) > u_1^*(c) > u_1^*(a)$$

$$u_2^*(c) > u_2^*(a) > u_2^*(b)$$

$$u_3^*(c) > u_3^*(a) > u_3^*(b)$$

$u_1^{**}(a) > u_1^{**}(b) > u_1^{**}(c)$

$u_2^{**}(c) > u_2^{**}(b) > u_2^{**}(a)$

$u_3^{**}(c) > u_3^{**}(a) > u_3^{**}(b)$

$u_1^{***}(b) > u_1^{***}(a) > u_1^{***}(c)$

$u_2^{***}(a) > u_2^{***}(b) > u_2^{***}(c)$

$u_3^{***}(a) > u_3^{***}(b) > u_3^{***}(c)$.

Then

(14) $f(u_1^*, u_2^*, u_3^*) = \{b,c\}$

(15) $f(u_1^{**}, u_2^{**}, u_3^{**}) = \{a\}$

(16) $f(u_1^{***}, u_2^{***}, u_3^{***}) = \{b\}$.

If f is implementable, there exists a weakly implementing game form $g\colon S_1 \times S_2 \times S_3 \to A$ and a vector of actions (s_1, s_2, s_3) such that $g(s_1, s_2, s_3) = c$ and (s_1, s_2, s_3) is a Nash equilibrium for the profile (u_1^*, u_2^*, u_3^*). Because $u_1^*(b) > u_1^*(c)$, there does not exist $s_1'' \in S_1$ such that $g(s_1'', s_2, s_3) = b$. If there exists $s_1' \in S_1$ such that $g(s_1', s_2, s_3) = a$, then (s_1', s_2, s_3) is a Nash equilibrium for $(u_1^{***}, u_2^{***}, u_3^{***})$, contradicting (16). If there does not exist $s_1' \in S_1$ with $g(s_1', s_2, s_3) = a$, then (s_1, s_2, s_3) is an equilibirum for $(u_1^{**}, u_2^{**}, u_3^{**})$, contradicting (15). Hence f is not implementable.

5. Nash versus Dominant Strategy Implementation

A dominant strategy is an action that an individual is willing to take regardless of the actions of others. Formally, we have

Dominant Strategy: In a game form $g\colon S_1 \times \ldots \times S_n \to A$, an action $\bar{s}_i$ is a dominant strategy for individual i with utility function u_i if for all $s_i \in S_i$ and $s_{-i} \in \prod_{j \neq i} S_j$

$u_i(g(\bar{s}_i, s_{-i})) \geq u_i(g(s_i, s_{-i}))$.

The definition of implementability in dominant strategies is analogous to that for Nash equilibrium. The game form g: $S_1 \times \ldots \times S_n \rightarrow A$ weakly implements the SCR f if for all profiles $(u_1, \ldots, u_n)$

(17) $DSE_g(u_1, \ldots, u_n)$ is nonempty.

and

(18) $DSE_g(u_1, \ldots, u_n) \subseteq f(u_1, \ldots, u_n)$,

where $DSE_g(u_1, \ldots, u_n)$ consists of all dominant strategy equilibium outcomes corresponding to $(u_1, \ldots, u_n)$. If (18) is an equality, g fully implements f.

As we suggested in Section 3, however, the literature on dominant strategies has emphasized not <u>this</u> definition but rather the concept of <u>truthful</u> implementation. For dominant strategies, a direct revelation game form is a mapping

$$g: \; U_1 \times \ldots \times U_n \rightarrow A.$$

The game form g truthfully implements f in dominant strategies if, for all $(u_1, \ldots, u_n)$, the actions $(u_1, \ldots, u_n)$ constitute a dominant strategy equilibrium with respect to the utility functions $(u_1, \ldots, u_n)$ and

$$g(u_1, \ldots, u_n) \in f(u_1, \ldots, u_n).$$

Clearly, if f is weakly implementable in dominant strategies, it is truthfully implementable. However, it is easy to give examples where the converse does not hold (e.g., Example 4.1.2 in Dasgupta, Hammond, and Maskin (1979)). Nonetheless, there is an important case in which we <u>can</u> deduce the converse; viz., where the U_i's contain only strict preferences.

<u>Lemma 2</u>: Suppose the U_i's contain only strict preferences. If the SCR f: $U_1 \times \ldots \times U_n \rightarrow A$ is truthfully implementable, then it is weakly implementable in dominant strategies.

Proof: See Dasgupta, Hammond and Maskin (1979).

For much of the rest of this section, we will concentrate on SCR's that are single-valued, i.e., whose choice sets contain only a single element. For such SCR's (denoted "SSCR's" for single-valued social choice rules) we can characterize truthful implementability in terms independent person-by-person monotonicity.

Independent Person-by-Person Monotonicity (IPM): An SSCR f satisfies IPM if for all $(u_1,\ldots,u_n) \in U_1 \times \ldots \times U_n$, all i, all $\bar{u}_i \in U_i$ and all a, b $\in$ A such that $a \in f(u_1,\ldots,u_n)$ and $\bar{u}_i(b) > \bar{u}_i(a)$, it must be the case that $b \notin f(\bar{u}_i, u_{-i})$.

Lemma 3: An SSCR $f: U_1 \times \ldots U_n \to A$ is truthfully implementable if and only if it satisfies IPM.

Proof: See Dasgupta, Hammond, and Maskin (1979).

We should point out that IPM does not, in general, imply monotonicity. That is, truthful implementability (even full implementability) in dominant strategies does not imply Nash implementability.

Example 2: Let $A = \{a,b,c,d\}$ and $n = 3$. Suppose that each U_i consists of 4 utility functions: u^a, u^{ab}, u^b, u^c, where

$$u^a(a) > u^a(b) > u^a(d) > u^a(c)$$

$$u^{ab}(a) = u^{ab}(b) > u^{ab}(c) > u^{ab}(d)$$

$$u^b(b) > u^b(a) > u^b(d) > u^b(c)$$

$$u^c(c) > u^c(d) > u^c(a) > u^c(b).$$

Define the SSCR $f: U_1 \times U_2 \times U_3 \to A$ so that

$$f(u_1,u_2,u_3) = \begin{cases} \{c\}, \text{ if } u_i = u^c \text{ for some } i \text{ and} \\ \quad \text{a majority prefers } c \text{ to } d. \\ \{d\}, \text{ if } u_i = u^c \text{ for some } i \text{ and} \\ \quad \text{a majority prefers } d \text{ to } c. \\ \{b\}, \text{ if at least two individuals have} \\ \quad \text{utility function } u^b \text{ and no one} \\ \quad \text{has } u^c. \\ \{a\}, \text{ otherwise} \end{cases}$$

One can verify straightforwardly that f is truthfully implementable in dominant strategies. In fact, the direct revelation game form corresponding to f <u>fully</u> implements f (and has the strong property that, truth-telling is dominant even for coalitions). However, f is not monotonic because, for example, $f(u^b,u^b,u^a) = \{b\}$ but $f(u^{ab},u^b,u^a) = \{a\}$ even though, for all $x \in A$, $u^b(b) \geq u^b(x)$ implies $u^{ab}(b) \geq u^{ab}(x)$. Thus f is not implementable in Nash equilibrium. This may seem odd, because the concept of dominant strategy equilibrium is much more demanding than that of Nash equilibrium. The apparent paradox is resolved by remembering that, to implement an SCR, one not only has to ensure that the elements of the choice set can arise as equilibrium outcomes, one has to <u>prevent</u> the existence of equilibrium outcomes <u>outside</u> the choice set. It is easier to meet this second requirement when dominant strategies are the solution concept, since by the very stringency of a dominant strategy equilibrium, a nonoptimal equilibrium is less like to arise.

Nonetheless, when preferences are strict, dominant strategy implementability does imply Nash implementability:

Theorem 5 (Dasgupta, Hammond, and Maskin (1979)): If the U_i's contain only strict preferences, then an SSCR f that is truthfully implementable in dominant strategies is also monotonic.

Proof: From Lemma 3, an SSCR that is truthfully implementable satisfies IPM. Consider $(u_1,\ldots,u_n)$, $(\bar{u}_1,\ldots,\bar{u}_n)$, and $a \in A$ such that $a \in f(u_1,\ldots,u_n)$ and, for all $b \in A$ and i, $u_i(a) > u_i(b)$ implies $\bar{u}_i(a) > \bar{u}_i(b)$. Suppose that $c \in f(\bar{u}_1,u_2,\ldots,u_n)$ for some $c \in A$. If $c \neq a$, then IPM implies that $\bar{u}_1(c) > \bar{u}_1(a)$ and $u_1(a) > u_1(c)$. But $u_1(a) > u_1(c)$ implies $\bar{u}_1(a) > \bar{u}_1(c)$, by hypothesis. Therefore a=c, and so $a \in f(\bar{u}_1,u_2,\ldots,u_n)$. Continuing iteratively, $a \in f(\bar{u}_1,\ldots,\bar{u}_n)$.

Q.E.D.

Not surprisingly, monotonicity does not in general imply IPM. Still, there is a large class of cases where the implication holds. To discuss this class, we need the following definition.

Monotonically Closed Domain[5]: A class U of utility functions is a monotonically closed domain if, for all pairs $\{u,u'\} \subseteq U$ and $\{a,b\} \subseteq A$ such that (i) $u(a) \geq u(b)$ implies $u'(a) \geq u'(b)$ and (ii) $u(a) > u(b)$ implies $u'(a) > u'(b)$, there exists $u'' \in U$ such that for all $c \in A$ (iii) $u(a) \geq u(c)$ implies $u''(a) \geq u''(c)$, and (iv) $u'(b) \geq u'(c)$ implies $u''(b) \geq u''(c)$.

One way of generating a u" satisfying the requirements of the definition is by taking minimums: if $u(a) = u'(a)$ and $u(b) = u'(b)$, then $u'' = \min(u,u')$ will suffice.

[5]A monotonically closed domain is called a "rich domain" in Dasgupta, Hammond, and Maskin (1979).

Clearly, the unrestricted domain U_A is monotonically closed. Trivially, any domain consisting of a single utility function is also monotonically closed. Suppose that A is the set of allocations across individuals of fixed stocks of m divisible commodities. If U consists of all utility functions corresponding to continuous, strictly monotone, strictly convex, selfish (i.e., no externalities) preferences over A, then, as shown by Dasgupta, Hammond, and Maskin (1979), U is monotonically closed as well.

<u>Theorem 6</u> (Dasgupta, Hammond, and Maskin (1979)): If U_i is monotonically closed for all i, then if the SSCR f is implementable in Nash equilibrium, it is truthfully implementable in dominant strategies.

<u>Proof</u>: If f is implementable in Nash equilibrium, then it is monotonic. If f violated IPM, there would exist $(u_1,\ldots,u_n)$, $\bar{u}_i$, a, and, b such that $a \in f(u_1,\ldots,u_n)$ and $\bar{u}_i(a) > \bar{u}_i(b)$ but $b \in f(\bar{u}_i,u_{-i})$. From the monotonic closure of U_i, however, there exists $\bar{\bar{u}}_i \in U_i$ such that for all c

$$u_i(a) \geq u_i(c) \text{ implies } \bar{\bar{u}}_i(a) \geq \bar{\bar{u}}_i(c)$$

and

$$\bar{u}_i(b) \geq \bar{u}_i(c) \text{ implies } \bar{\bar{u}}_i(b) \geq \bar{\bar{u}}_i(c).$$

From monotonicity applied to (u_i,u_{-i}) and $(\bar{\bar{u}}_i,u_{-i})$, we have $a \in f(\bar{\bar{u}}_i,u_{-i})$. But from monotonicity applied to $(\bar{u}_i,u_{-i})$ and $(\bar{\bar{u}}_i,u_{-i})$, $b \in f(\bar{\bar{u}}_i,u_{-i})$, a contradiction of f's single-valuedness. Therefore, f satisfies IPM and so is truthfully implementable in dominant strategies.

Q.E.D.

Theorem 6 implies that if a planner wishes to implement a single-valued SCR, he will get no extra mileage from using the ostensibly weaker concept of Nash implementation if the domain of utility functions is monotonically closed. In particular, we have the following negative result.

Corollary 1 (Dasgupta, Hammond, and Maskin (1979), Roberts (1979): Suppose that A contains at least three elements and that f: $U_A^n \to A$ is an n-person SSCR that is onto A. If f is implementable in Nash equilibrium, it is dictatorial.

Proof: Because U_A is monotonically closed, Theorem 6 implies that f is truthfully implementable in dominant strategies. But then, from the Gibbard (1973)/Satterthwaite (1975) theorem on dominant strategies, f is dictatorial.

Q.E.D.

Roberts (1979) extends Corollary 1 to the case of "conjectural" equilibria, where, rather than taking other players' strategies as given, an individual conjectures that others will respond to his strategy choice. This result is, in turn, closely related to one of Pattanaik (1976).

Another implications we can draw from Theorem 6 is a set of conditions under which an implementable f can be thought of as maximizing a social aggregation function.

Social Aggregation Function: Let B_A be the class of all complete, reflexive, binary relations on A. A social aggregation function (SAF) is a mapping

$$F: U_1 \times \ldots \times U_n \to B_A.$$

If the range of F consists of acyclic relations, F is called a social decision function, and if these relations are also transitive, F is a social welfare function. F satisfies the Pareto property if whenever all individuals strictly prefer a to b (i.e., $u_i(a) > u_i(b)$ for all i) then $F(u_1,\ldots,u_n)$ ranks a above b. F satisfies nonnegative response if, for all $\{a,b\}$ and $\{(u_1,\ldots,u_n), (\bar{u}_1,\ldots,\bar{u}_n)\}$, if, for all i, $u_i(a) \geq u_i(b)$ implies $\bar{u}_i(a) \geq \bar{u}_i(b)$ and $u_i(a) > u_i(b)$ implies $\bar{u}_i(a) > \bar{u}_i(b)$, then that a is ranked weakly (strictly) above b by $F(u_1,\ldots,u_n)$ implies that a is ranked weakly (strictly) above b by $F(\bar{u}_1,\ldots,\bar{u}_n)$.

The SSCR f maximizes F if, for all $(u_1,\ldots,u_n)$, $a \in f(u_1,\ldots,u_n)$ implies that, for all $b \neq a$, a is stricity preferred to b by $F(u_1,\ldots,u_n)$.

Corollary 2: Suppose that the U_i's are monotonically closed and consist only of strict preferences, the SSCR f is implementable in Nash equilibrium if and only if there exists an SAF F satisfying nonnegative response such that f maximizes F. Furthermore, if f is onto A, F satisfies the Pareto property.

Proof: See Dasgupta, Hammond, and Maskin (1979)

Nonnegative response implies independence of irrelevant alternatives (IIA) in the sense of Arrow (1951). Corollary 2, therefore, illustrates the close relationship among monotonicity, IPM, and IIA.

6. Quasilinear Preferences

So far, the only particular domain of utility functions that we have discussed in any detail is the unrestricted domain U_A. We next consider an important restricted domain: the class of quasilinear preferences.

Suppose that a social alternative consists of a public decision d (which is an element of some set D) and a vector $(t_1,\ldots,t_n)$ of transfers of some private good (the t_i's are real numbers). Individual i's preferences are quasilinear if his utility function u takes the form

(20) $v(d) + t_i$.

Let U_i^{QL} be the class of all preferences of form (20). This class has been the object of much study in the dominant strategy implementation literature (see, for example, Clarke (1971), Groves (1973), Green and Laffont (1979)). Rather less has been done with it in the Nash implementation literature (see, however, Laffont and Maskin (1982a) and (1982b) and Roberts (1979)).

It is readily verified that the domain U_i^{QL} is not monotonically closed. Therefore, Theorem 10 does not apply, and we cannot conclude that the sets of Nash- and dominant strategy-implementable SSCR's are the same. Nevertheless, as Roberts (1979) has shown, the public decision parts of the SSCR's are identical.

In view of (20) we can express an SSCR as a function of the public parts of individuals' utility functions. Write

$$f(v_1,\ldots,v_n) = (d(v_1,\ldots,v_n),\ t(v_1,\ldots,v_n),\ldots,t_n(v_1,\ldots,v_n)),$$

where $u_i = v_i + t_i$.

Theorem 7: Suppose that D, the public decision space is finite. Let f: $U_1^{QL}\times\ldots\times U_n^{QL} \to A$ be an SSCR such that d( ) is onto D. Then if f is either Nash-implementable or truthfully implementable in dominant strategies, there exist $v_0 : D \to R$ and numbers $\alpha_1,\ldots,\alpha_n$ such that $\alpha_i \geq 0$ for all i and $\sum_{i=1}^{n} \alpha_i = 1$ such that $d(v_1,\ldots,v_n) = \arg\max_d (v_0(d) + \sum_{i=1}^{n} \alpha_i v_i(d))$.

Proof: See Roberts (1979).

Laffont and Maskin (1982a) place more structure on the problem by assuming that

$$D = [0,1]$$

and that the individuals' v_i functions are concave and differentiable and take their maxima in the interior of D. Let V be the class of such functions. They also assume that the public decision function d( ) is weakly efficient (if $v_1 = \ldots = v_n$, then $d(v_1,\ldots,v_n) = \arg\max v_i$), and neutral ($d(\hat{v}_1,\ldots,\hat{v}_n) = d(v_1,\ldots,v_n) + c$, where for all i, $\hat{v}_i(d) = v_i(d-c)$).

Theorem 8: Let f be an SSCR on $V\times\ldots\times V$ that is either Nash-implementable or truthfully implementable in dominant strategies. If d is weakly efficient and neutral then

(i) there exists a continuous and semi-strictly increasing[5] function h: $R^n \to R$ such that $h(0,\ldots,0) = 0$ and $d(v_1,\ldots,v_n)$ solves $h(v_1'(d),\ldots,v_n'(d)) = 0$, where primes denote derivatives;

(ii) if f is Nash implementable, t_i is a function of the numbers $d(v_1,\ldots,v_n)$ and $v_1'(d(v_1,\ldots,v_n)),\ldots,v_n'(d(v_1,\ldots,v_n))$;

(iii) if f is truthfully implementable in dominant strategies, then

$$t_i = -\int_0^{d(v_1,\ldots,v_n)} h_i(v_i'(t))dt + H_i(v_{-i}),$$

where h_i: $\mathbb{R}^{n-1} \to \mathbb{R}$ satisfies

[5] By "semi-strictly increasing" we mean that if x is bigger in evry component than y, then $h(x) > h(y)$.

$h(h_i(a_i), a_{-i}) = 0$, if there exists a_i with $h(a_i, a_{-i}) = 0$

$h_i(a_{-i}) = 0$, otherwise.

Proof: See Laffont and Maskin (1982a).

Notice that the set of implementable public decisions is defined by varying h, whether it be Nash or dominant strategy implementation. When, for example, $h = \sum_{i=1}^{n} \lambda_i v_i'$.

the public decision becomes

$$d(v_1, \ldots, v_n) = \arg\max_d \sum_{i=1}^{n} \lambda_i v_i(d).$$

The form of the transfers, however, depends on the type of implementation. Nash implementation demands that the transfers be a function of the optimal public decision and the derivatives of individuals' utility functions evaluated at the optimum. Dominant strategy implementation requires that an individual's transfer be the sum of two terms: a term depending on the derivitives of the utility functions and the public decision, and a term depending only on the utility functions of the other individuals.

7. Strong Equilibrium

Nash equilibrium is a noncooperative concept; it implicitly assumes that individuals do not act in concert. When individuals can collude, strong equilibrium may be a more appropriate solution concept.

Strong Equilibium: A strong equilibrium for the game form $g: S_1 \times \ldots \times S_n \to A$ with respect to the profile $(u_1, \ldots, u_n)$ is a configuration $(\overline{s}_1, \ldots, \overline{s}_n)$ such that for all coalitions $C \subseteq \{1, \ldots, n\}$ and all $s_C \in \prod_{j \in C} S_j$ there exists $i \in C$ such that $u_i(g(\overline{s})) \geq u_i(g(s_C, \overline{s}_{-C}))$.

By analogy with Nash equilibrium, a game form g fully implements the f in strong equilibrium if for all profiles $(u_1,\ldots,u_n)$

$$SE_g(u_1,\ldots,u_n) = f(u_1,\ldots,u_n),$$

where $SE_g(u_1,\ldots,u_n)$ consists of the strong equilibrium outcomes of g for the profile $(u_1,\ldots,u_n)$.

We should note that if g fully implements f in strong equilibrium, it does <u>not</u> necessarily implement f in Nash equilibrium. The reason for this apparent anomaly is that g may possess Nash equilibria that are not strong and which, futhermore, do not lead to outcomes in the choice set. For example, consider the following two-person game form, where individual 1 chooses rows as actions, and individual 2, columns:

	s_2	s_2'
s_1	a	a
s_1'	a	b

This game form fully implements the SSCR $f^*\colon\ U_1 \times U_2 \to \{a,b\}$ in strong equilibria, where the U_i's contain the strict preferences on $\{a,b\}$ and

$$f^*(u_1,u_2) = \begin{cases} b, & \text{if both individuals prefer b to a} \\ a, & \text{otherwise.} \end{cases}$$

However, the game form does <u>not</u> implement f^* in Nash equilibrium, because (s_1,s_2) is a non-f^*-optimal Nash equilibrium when both individuals prefer b to a.

We have seen that monotonicity is a necessary condition for implementability in Nash equilibirum. The same is true for strong

Theorem 9: (Maskin (1979b): If an SCR f is implementable in strong equilibrium, it is monotonic.

On the other hand, weak no veto power, which played an important role in establishing positive results for Nash implementation, prevents implementation in strong equilibrium when the number of individuals does not exceed the number of alternatives and the domain is unrestricted.

Theorem 10: If the n-person SCR $f\colon U_A \times \ldots \times U_A \to A$ is onto A, n is less than or equal the cardinality of A but greater than or equal to three, and f is implementable in strong equilibrium, f does not satisfy weak no veto power.

Proof: The proof consists of considering a "cyclic" profile of preferences $(u_1, \ldots, u_n)$, where

$$u_1(a_1) > u_1(a_2) > \ldots > u_1(a_n)$$
$$u_2(a_2) > u_2(a_3) > \ldots > u_2(a_1)$$
$$\vdots$$
$$u_n(a_n) > u_n(a_1) > \ldots > u_n(a_{n-1}).$$

Such a profile exists because there are at least as many alternatives as individuals. But then it is a straightforward to show that no alternative can be a strong equilibrium, since no single individual has veto power. For the details, see Maskin (1979b).

Q.E.D.

Theorem 10 is false if the number of individuals exceeds the number of alternatives, as the following example shows.

Example 3: Let n = 3, A = {a,b}, and U_i consist of the strict preferences on A. Let f be majority rule, i.e., an alternative is in the choice set if and only if it is top-ranked by two or more individuals. The following game form implements f:

a	a
a	b

a	b
b	b

where individual 1 chooses rows, 2 columns, and 3 matrices. A large class of other examples has been constructed by Moulin and Peleg (1982).

Clearly if an SCR f is onto its range and fully implementable, it must be Pareto optimal. In Section 4 we demonstrated that the SCR that selects all Pareto optimal and individually rational alternatives is implementable in Nash equilibirum. In fact, this is the <u>only</u> individually rational SCR on the unrestricted domain that is fully implementable in strong equilibrium.

<u>Individually Rational SCR</u>: If $a_0 \varepsilon A$ is the status quo, an SCR $f: U_1 \times \ldots \times U_n \rightarrow A$ is individually rational if for all $(u_1, \ldots, u_n)$ and all $a \varepsilon f(u_1, \ldots, u_n)$ $u_i(a) \geq u_i(a_0)$ for all i.

<u>Theorem 11</u> (Maskin (1979b)): Let f_Q: $U_A \times \ldots \times U_A \rightarrow A$ be the SCC such that for all $(u_1, \ldots, u_n)$

$$f_Q(u_1, \ldots, u_n) = \{a \ \varepsilon \ A \mid \text{for all } j \ u_j(a) \geq u_j(a_0) \text{ and, for all i, and for all } b \ \varepsilon \ A, \text{ there exists i such that } u_i(a) \geq u_i(b)\}.$$

Then f_Q is the unique individually rational SCC on $U_A \times \ldots \times U_A$ that is implementable in strong equilibrium.

<u>Proof</u>: It is immediate to verify that f_Q is fully implemented by the game form (10) (which, interestingly, also implements the individual rationality correspondence in Nash equilibrium). That f_Q is the <u>only</u> implementable individually rational SCR on the unrestricted domain follows from an argument in Maskin (1979b).

8. Double Implementation

Whereas implementation in Nash equilibrium ignores the possibility of collusion, implementation in strong equilibrium may, in effect, require coalitions to form. To see this, consider the game form (10). In order to obtain any alternative other than a_0, all individuals have to take the same action. Clearly, there are many (non-strong) Nash equilibria in which different individuals take different actions, and to avoid ending up in one of these presumably involves some coordination. That is, collusion is necessary.

Because the game form designer may not know the extent to which collusion can or will take place, it is desirable to have an implementation concept that does not posit any particular degree of collusion. One possibility is to require a game form to fully implement simultaneously in both strong and Nash equilibrium. This game form would yield optimal outcomes regardless of collusion. We shall say that such a game form (fully) doubly implements the SCR.

Of course, double implementation is a very demanding requirement. Not very surprisingly, when the number of alternatives is at least three and the domain of utility functions is unrestricted, the only SCR's that are onto A and doubly implementable are dictatorial.

Theorem 12 (Maskin 1979a): Suppose A contains at least three elements and $f\colon U_A \times \ldots \times U_A \to A$ is an n-person SCR that is onto A. If f is doubly implementable, then it is dictatorial.

The results are more encouraging, however, when preferences are restricted. Suppose, in particular, that there exists (at least) one divisible and transferable private good that all individuals find desirable and that does not create externalities (i.e., one individual's allotment of this good does not affect any other's utility). Let us express a social alternative a as $(b,t_1,\ldots,t_n)$, where t_i is the transfer of this private good to individual i, and b represents all other social decisions inherent in a. We shall call b the "public decision," although it may itself entail the allocation of private goods. Denote the status quo, a_0, by $(b_0,0,\ldots,0)$. Suppose that the private good is sufficiently desirable (and that consumers have enough of it in the status quo so that, for all i and all public decisions b, there exist $(\bar{t}_1,\ldots,\bar{t}_n)$ such that

(21) $(b,\bar{t}_1,\ldots,\bar{t}_n) \in A$ and, for all $t_i < \bar{t}_i$ and all u_i, $u_i(a_0) >$
$$u_i(b,t_i,\bar{t}_{-i}).$$

Condition (21) provides for the existence of "punishments." It says that regardless of the public decision, it is always possible to take away enough of the private good from individual i to make him worse off than under the status quo. We have the following result.

Theorem 13: Assume the existence of a desirable and divisible private good. If (21) is satisfied, then any individually rational and Pareto optimal SCR is fully doubly implementable.

Proof: See Maskin (1979a).

9. Related Concepts

This paper has discussed Nash, strong Nash, and "double" implementation. We should, however, mention two related lines of work.

Farquharson (1969) proposed the concept of a "sophisticated" equilibrium. This is a refinement of Nash equilibrium in which weakly dominated strategies are successively eliminated. For example, consider the following two player game:

	d	e	f
a	2,2	1,1	1,1
b	0,0	1,1	2,1
c	0,0	1,2	0,0

The strategy configurations (a,d), (b,e), (c,e), and (b,f) are all Nash equilibria. However, strategies c and f are weakly dominated for players I and II. If we delete them, the game becomes

	d	e
a	2,2	1,1
b	0,0	1,1

Notice that here strategies b and e are weakly dominated. Once these are deleted, the players have one strategy each. Hence (a,d) forms a <u>sophisticated</u> or <u>dominance solvable</u> equilibrium.

The theory of implementation in dominance solvable equilibrium has been developed largely by Moulin (see Moulin (1979a), (1979b), (1979c), (1980), (1981)). Although a full characterization of the implementable SCR's is not available, there are by now many examples of Pareto optimal, neutral, and anonymous SCR's that can be implemented, including some that are not Nash implementable.

An SSCR can itself be thought of a a game form; a player's strategy is the announcement of a utility function (not necessarily his true one) and the outcome is the alternative optimal with respect to the announced preferences. An SSCR is said to be consistent if for any profile of (true) preferences there exists a strong equilibrium of the SSCR (when viewed as a game form) whose outcome is optimal with respect to those (true) preferences. Notice that the qualification about optimality is not superfluous since the strategies played in equilibrium may themselves be untruthful. The concept of consistency is due to Peleg (1977). Besides Peleg, contributors to the subject include Dutta and Pattanaik (1978).

REFERENCES

Arrow, K. (1951), Social Choice and Individual Values, Cowles Foundation Monograph 12.

Clarke, E. (1971), "Multipart Pricing of Public Goods," Public Choice, 8:19-33.

Dasgupta, P., P. Hammond, and E. Maskin (1979), "The Implementation of Social Choice Rules," Review of Economic Studies, 46:185-216.

Dutta, B. and P. Pattanaik (1978), "On Nicely Consistent Voting Systems," Econometrica, 46:163-170.

Farquharson, R. (1969), The Theory of Voting, New Haven: Yale University Press.

Gibbard, A (1973), "Manipulation of Voting Schemes: A General Result," Econometrica, 41:587-602.

Green, J. and J.J. Laffont (1979), Incentives in Public Decision-Making, Amsterdam:North-Holland.

Groves, T. (1973), "Incentives in Teams," Econometrica, 41:617-31.

Groves, T. and J. Ledyard (1977), "Optimal Allocation of Public Goods: A Solution to the Free Rider Problem," Econometrica, 45:783-810.

Harsanyi, J. (1967), "Games with Incomplete Information Played by 'Bayesian' Players," Management Science, 14:159-82, 320-34, 486-502.

Hurwicz, L. (1972), "On Informationally Decentralized Systems," in Decision and Organization, edited by C.B. McGuire and R. Radner, Amsterdam:North-Holland.

Hurwicz, L. and D. Schmeidler (1978), "Outcome Functins which Guarantee the Existence and Pareto Optimality of Nash Equilibria," Econometrica, 46:144-74.

Laffont, J.J. and E. Maskin (1982a), "Nash and Dominant Strategy Implementation in Economic Environments," Journal of Mathematical Economics, 10:17-47.

Laffont, J.J. and E. Maskin (1982b), "The Theory of Incentives: An Overview," in Advances in Economic Theory, edited by W. Hildenbrand, Cambridge:Cambridge University Press.

Maskin, E. (1977), "Nash Equilibrium and Welfare Optimality," to appear in Masthematics of Operations Research.

Maskin, E. (1979a), "Incentive Schemes Immune to Group Manipulation," mimeo.

Maskin, E. (1979b), "Implementation and Strong Nash Equilibrium," in Aggregate and Revelation of Preferences, edited by J.J. Laffont, Amsterdam:North-Holland.

Moulin, H. (1979a), "Dominance-Solvable Voting Schemes," Econometrica, 47:1337-51.

Moulin, H. (1979b), "Implementing Just and Efficient Decision Making," forthcoming in the Journal of Public Economics.

Moulin, H. (1979c), "Prudence versus Sophistication in Voting Strategy," forthcoming in the Journal of Economic Theory.

Moulin, H. (1980), "Implementing Efficient, Anonymous, and Neutral Social Choice Functions," Journal of Mathematical Economics, 7:249-269.

Moulin, H. (1981), The Strategy of Social Choice, Paris: Laboratoire d'Econometrie de l'Ecole Polytechnique.

Moulin, H. and B. Peleg (1982), "Cores of Effectivity Functions and Implementation Theory," Journal of Mathematical Economics, 10:115-45.

Muller, E. and M. Satterthwaite (1977), "The Equivalence of Strong Positive Association and Strategy-Proofness," Journal of Economic Theory, 14:412-18.

Myerson, R. (1979), "Incentive Compatibility and the Bargaining Problem," Econometrica, 47:61-73.

Myerson, R. (1982), "Optimal Coordination Mechanisms in Principal-Agent Problems," Journal of Mathematical Economics, 10.

Myerson, R. (1983), "Implementation via Bayesian Equilibria," in Social Goals and Social Organization: Volume in Memory of Elisha Pazner, Cambridge University Press, forthcoming.

Nash, J. (1950), "Equilibrium Points in n-person Games," Proceedings of the National Academy of Sciences, 36:48-50.

Pattanaik, P. (1976), "Counter-threats and Strategic Manipulation under Voting Schemes," Review of Economic Studies, 43:11-18.

Peleg, B. (1977), "Consistent Voting Systems," Econometrica, 46:153-162.

Roberta, J. and A. Postlewaite (1976), "The Incentives for Price-taking Behavior in Large Exchagne Economies," Econometrica, 44:115-28.

Roberts, K. (1979), "The Characterization of Implementable Choice Rules," in Aggregation and Revelation of Preferences, edited by J.J. Laffont, Amsterdam:North-Holland.

Satterthwaite, M. (1975), "Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions," Journal of Economic Theory, 10:187-217.

Date Due

MAR ~ 1991		
DEC 2 1991		
NOV. 20 1992		
		Lib-26-67