

SUBGAME PERFECT IMPLEMENTATION

BY JOHN MOORE AND RAFAEL REPULLO¹

We examine the use of stage mechanisms in implementation problems, and we partially characterize the set of choice rules that are implementable in subgame perfect equilibria. The conditions we derive will usually be satisfied if there is at least one private good. Our conclusion is that in these “economic” environments, almost *any* choice rule can be implemented—even if the rule is nonmonotonic, even if it is single-valued, and, under quite reasonable circumstances, even if there are just two agents. Moreover, the mechanism need have no more than three stages.

To illustrate the power of this approach, we discuss a number of models in which it is possible to implement the first-best (although it would have been impossible to do so without using stage mechanisms). The diversity of these models suggests that subgame perfect implementation may find wide application: public goods, contracts, agency theory, constitution design. In several cases, the mechanisms we present have the added attraction that agents’ moves are sequential, not simultaneous.

Finally, we show that the Walras correspondence is implementable in subgame perfect equilibria.

KEYWORDS: Implementation, subgame perfect equilibrium, choice rules, public goods, contracts, agency theory, voting.

1. INTRODUCTION

THE TITLE OF THIS PAPER is dry, but apposite: we shall partially characterize the set of choice rules that can be implemented in subgame perfect equilibria. But our prime objective is to show, by way of several examples, that what we are doing here may have practical relevance—or, at very least, that the approach we are adopting is potentially applicable in a wide number of areas. To begin, though, it may be helpful to introduce some basic concepts, and to sketch the implementation problem in its conventional, albeit abstract, form.

(This will be a rather long Introduction. For the reader who is familiar with the implementation problem, and who wishes to skip the motivation, please turn directly to the paragraph preceding Figure 1.)

We consider an environment with $i = 1, \dots, I$ agents, and a set of feasible outcomes, or decisions, A . There are a number of possible states of nature, Θ . In a given state, the profile of agents’ preferences over A is indexed by $\theta \in \Theta$. The traditional social choice literature would take θ to index the agents’ utility functions; i.e., states of nature would be distinguished solely by changes in agents’ underlying preferences. We would like to broaden this interpretation. States of nature may also be distinguished by changes in endowments or

¹ We are very grateful to the anonymous referee for his thorough criticisms. In addition to several seminar participants, we would like to thank Oliver Hart, David Kreps, Tom Palfrey, Sanjay Srivastava, and, especially, Andrew Postlewaite for helpful comments and discussions. Andrew Postlewaite was largely responsible for the example in the Introduction. We acknowledge financial support from the British Council and the Suntory Toyota International Centre for Economics and Related Disciplines at the London School of Economics.

technology.² A choice function f associates an outcome $f(\theta)$ to each θ in Θ . (More generally, f may take on many values, but for the moment we ignore choice correspondences.³ We shall come to these in Section 2.) A mechanism, or game form, g endows each agent with a strategy set, and maps each vector of strategies chosen by the agents into an outcome. The implementation problem can be posed in the following terms. Given a choice function f , does there exist a mechanism g such that when the agents with a preference profile indexed by θ play the corresponding game, the unique equilibrium outcome is $f(\theta)$? Notice that what gives the implementation problem its bite is that the mechanism g cannot be tailored to the profile θ ; the same g has to cope with all profiles in Θ .

This framework encompasses an extraordinarily diverse number of situations. The most immediate example is that of a planner deciding on the level of provision of a public good together with who should contribute what. In this case, θ indexes how much each household is willing to pay, and $f(\theta)$ denotes both the scale of provision of the public good and the vector of contributions. Auction models, monopoly pricing, and many optimal taxation problems can all be couched in similar terms. Typically, in these examples the agents' (or households') individual preferences will be privately known, and so it would be appropriate to design the mechanism g so that $f(\theta)$ is a Bayesian equilibrium outcome under profile θ .

There are, however, circumstances in which the agents may know each other's preferences, even though the planner does not. For example, consider a manager devising an incentive scheme for a team of I employees. θ indexes the employees' collective working environment. One might reasonably suppose that all the employees know θ even though the manager does not. In this paper, we shall for the most part focus on such "complete information" environments.

More compelling examples of environments with complete information arise when there is no planner (auctioneer, monopolist, government, or manager).

² For example, suppose A comprised the set of stochastic transfers between agents. An agent's risk preferences over A will be affected by a change in his wealth (except in the case of constant absolute risk aversion)—even though his underlying von Neumann–Morgenstern utility function is fixed. Here, "outcomes" are to be interpreted narrowly as the endogenously-determined stochastic transfers; θ summarizes all other (exogenous) payoff-relevant information, including changes in wealth. Consider a second example. Suppose there are two firms (the agents) and an "outcome" is the quantity they trade together with the price. A firm's preferences over A will vary with its productivity/technology, which are indexed by θ . But again, notice that the firm's underlying preference—for profit maximization—is fixed.

In this connection, it is important to observe that A does not vary with θ : any outcome in A is assumed to be feasible—though it may have a very low payoff. In the first example (risk-sharing), the von Neumann–Morgenstern utilities must be defined over a suitably large domain. In the second example, the two firms must in principle be able to trade any of the quantities (at their respective prices) in A , without violating either a technological or a bankruptcy constraint. Of course, it would be desirable to generalize the analysis to the case where A depends on θ , but the difficult problem is how to deal with infeasible outcomes off the equilibrium path.

³ Actually, it is in general more difficult to implement a choice function (using single-stage mechanisms) than to implement a choice correspondence. See, for example, Dasgupta, Hammond, and Maskin (1979).

Consider a club with I members designing its constitution—a mechanism g for making future decisions. Here $f(\theta)$ denotes the collective decision which they would like to take in the event that the members' future preferences are indexed by θ . Typically g will be a voting procedure. It is designed ex ante by the I members for use ex post—i.e., once θ has been realized. Notice that we do not delve into the question of how the choice function f is chosen ex ante; we might presume that either the entire function f is a social standard on which all the club members agree a priori, or that it is the outcome of ex ante negotiation and compromise. The question arises: why not simply write a contract specifying that $f(\theta)$ should be enforced conditional on the realized θ ? We assume that such a contract is infeasible because θ is never observed by outsiders—in particular, the courts.⁴ Thus the club members have to resort to an indirect way to implement f , by using a mechanism g . The point is that g can be contractually enforced (e.g., by penalizing a member for not voting), whereas f cannot.

Of course, the agents could avoid specifying any mechanism, and instead negotiate an outcome once θ is known. However, although this may be ex post efficient, in general there will be inefficiencies ex ante. Consider two firms that are not vertically integrated. Let θ_1 index the upstream firm's ex post situation (which may include such things as input prices, productivity, and technology), and let θ_2 index the downstream firm's circumstances (e.g., its productivity, technology, and the state of demand for the final product). As in the club example, we suppose that although both firms observe $\theta = (\theta_1, \theta_2)$, this is not observable to outsiders; so a contract cannot be conditioned on θ . $f(\theta)$ denotes the desired price and quantity of the goods traded, together with their quality, time of delivery, and so forth. These will be chosen to ensure not just efficient ex post trading, but also to take into account certain key ex ante efficiency considerations such as risk sharing and/or inducing the firms to make appropriate (noncontractible) specific investments after they have signed the contract. The firms can settle the overall division of surplus by means of a side-payment at the time of contracting.⁵ Once this is settled, there is typically no ambiguity about the optimum f . But the point is that the firms must specify a mechanism g which implements f ; leaving decisions to ex post bargaining would in general not lead to the desired outcomes.

This last example is especially pertinent for us, since the mechanism g defined in it is likely to be played in stages (e.g., the downstream firm sends a letter of requirements, to which the upstream firm reacts, and so on). Presumably, the firms will play this game knowing that only credible threats matter. Thus, they

⁴ To use the jargon, the θ 's are observable (by the members) but not verifiable (by outsiders). There is another reason why a contingent contract may be infeasible, namely that the set Θ of all possible θ 's is too complicated to be written down in advance. But we do not want to rely on this idea, because the mechanism g may require specifying Θ .

⁵ If there are other firms with whom to contract, then this division may be determined in an ex ante market.

will implicitly restrict attention to those equilibria which satisfy a criterion of perfectness.⁶ We will come back to this later on.

Common to all the above examples is the problem of finding the best choice function that can be implemented. To make this task operational, it is necessary to pin down the set over which one is optimizing—that is, to characterize the set of implementable choice functions.

The Revelation Principle is often a useful tool. It states that if a choice function f can be implemented by a mechanism g , then there exists a revelation mechanism h (i.e., one in which the strategy space of each agent i is simply the set of all his possible types⁷) for which truth-telling is an equilibrium, whose outcome is $f(\theta)$. From this it follows that any implementable choice function must satisfy an explicit set of truth-telling, or “incentive compatibility” constraints. Thus we have a necessary condition for implementability. But the Revelation Principle does not provide a sufficient condition, since the revelation mechanism h may have other, unwanted, nontruthful equilibria.⁸

In our present context, where agents have complete information, the Revelation Principle has almost no cutting power. The incentive compatibility constraints are usually trivial to satisfy. Consider a revelation mechanism h with the following payoffs. Everyone announces a profile. If they agree on some profile θ , then $f(\theta)$ is implemented. If they disagree, then a “bad” outcome (from all the agents’ perspectives) is implemented. Clearly, given a true profile θ^* , say, it is a Nash equilibrium for everyone to announce the truth. The snag is that every other $\theta \neq \theta^*$ is also an equilibrium. In sum, the necessary condition provided by the Revelation Principle is almost vacuous (though the h we have just constructed does require a bad outcome).

Thus we need to resort to nonrevelation, or indirect, mechanisms. The seminal paper in this area is by Eric Maskin (1977) (also see his 1985 survey). His approach was to take the revelation mechanism h , and construct an indirect mechanism in which the unwanted equilibria are knocked out by giving agents

⁶ Hart and Moore (1988) analyze how an incomplete contract may be renegotiated. Renegotiation occurs before trade, and any disputes are resolved afterwards through the courts. It is crucial to the analysis that the equilibria of this model are assumed to be (subgame) perfect.

⁷ Note that “type” in a revelation mechanism refers not just to an agent’s personal preferences, but also to all the information he possesses. Here, he is assumed to know θ , which indexes all the other agents’ preferences too.

⁸ See Dasgupta, Hammond, and Maskin (1979) and Repullo (1986). Arguably, researchers who have appealed to the Revelation Principle have not always taken the problem of multiple equilibria adequately into account. One retort might be that in a revelation game, if truth-telling is an equilibrium, then it is a focal point: each agent is likely to assume that other agents are going to report the truth, and therefore he will do so too. However, this argument is suspect for a number of reasons, the most telling of which is that $f(\theta)$ may, from the agents’ perspectives, be Pareto inferior to some other nontruthful equilibrium. This is a distinct possibility when there is a conflict of interest between a planner and the agents. A case in point is the above example of a manager providing incentives for a team of employees. He may want them to work hard, but they prefer to shirk. If there is a (nontruthful) equilibrium of the revelation mechanism h in which they shirk, then it is reasonable to suppose that they will choose it. Demski and Sappington (1984) give an example of this in a Bayesian model where the employees’ characteristics are individually observed but correlated.

additional (“nuisance”) strategies. This approach has been adopted in most of the literature.

Another candidate for knocking out equilibria is to add stages to the mechanism, and impose the requirement that off-the-equilibrium-path strategies (threats) must be credible. In the present paper, where agents have complete information, this additional requirement will amount to subgame perfection. A closely related concept is that of dominance solvability, which has long been used in voting theory, starting with the pioneering work of Farquharson (1969) who termed it “sophisticated voting”, and extensively developed and placed in the modern mechanism design framework by Moulin (1979). This work examines successive voting by binary choices. A parallel line of enquiry concerns successive voting by veto methods; see, e.g., Mueller (1978), Moulin (1980), and Dutta (1980). Crawford (1977) independently analyzed the classical divide-and-choose procedure—which is perhaps the most familiar of stage mechanisms—and in Crawford (1979) he extended this idea by, among other things, introducing the possibility of auctioning off the role of the divider. Crawford’s mechanism was further developed by Demange (1984). Moulin has contributed a great deal in all these areas: his 1983 book provides an excellent synthesis. Other references are to be found there, although we should mention a later paper, Moulin (1984). So our approach is not new. However, the focus of the work just cited has been on specific environments, specific choice functions, and/or special kinds of mechanisms. Our aim in the present paper is to extend these authors’ insights to abstract environments and choice functions, and to work with more general stage mechanisms.

It turns out that the approach of adding stages is extremely effective, especially if simultaneous moves are allowed at each stage. Our conclusions are particularly clearcut for “economic environments”, by which we mean environments in which there is at least one divisible private good (e.g., money) from which agents derive positive benefit.⁹ A proper definition will be given in Section 2, but the essential characteristic of an economic environment, for our purposes, is that there is a specific conflict of interest among the agents. We shall demonstrate that if there are three or more agents *any* choice function can be implemented in economic environments. Moreover, we shall show that if there is the possibility of sufficient free disposal, or wastage, of the private good(s), then this conclusion can hold even when there are just two agents.

We now present a simple example. Although artificial, it neatly illustrates the power of stage mechanisms for implementation. Consider the two-person exchange economy depicted in Figure 1. There are two possible states θ , which are privately observed by the agents: either $\theta = C$ (both agents have Cobb–Douglas preferences; the solid indifference curves) or $\theta = L$ (both agents have Leontief preferences; the dashed indifference curves). $f(C)$ and $f(L)$ are the respective

⁹ Thus a voting procedure designed to select from a finite set of candidates is an example of a “noneconomic environment”—which, it must be quickly admitted, is an unfortunate description!

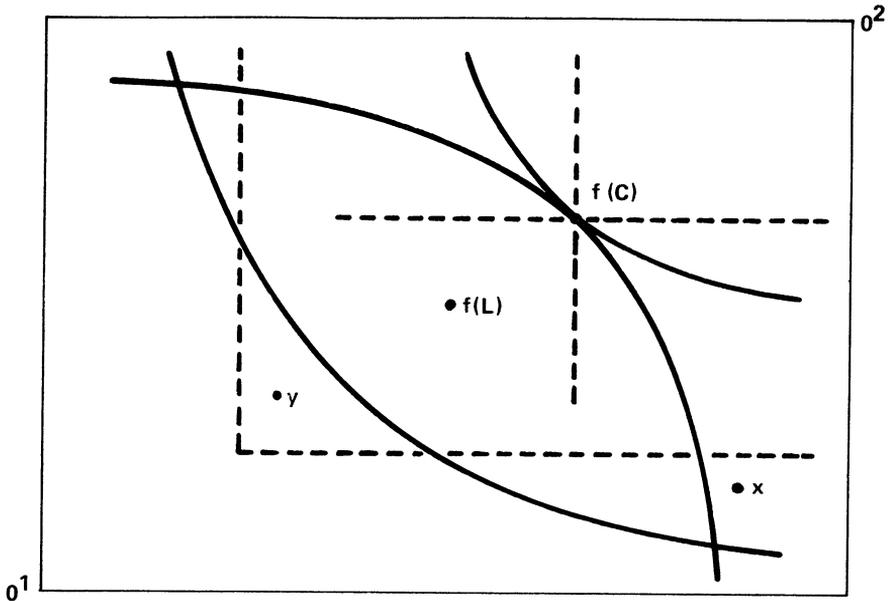


FIGURE 1

allocations of a choice function f . Note that at $f(C)$, the C -lower contour sets are contained in the L -lower contour sets.

We shall show below that f cannot be implemented as the unique Nash equilibrium outcome of a single-stage mechanism. However, f can be implemented as the unique subgame perfect equilibrium outcome of the following three-stage mechanism:

- STAGE 1: Agent 1 announces:
 either $L \Rightarrow$ implement $f(L)$; STOP
 or $C \Rightarrow$ Stage 2.
- STAGE 2: Agent 2 can:
 either agree \Rightarrow implement $f(C)$; STOP
 or challenge \Rightarrow Stage 3.
- STAGE 3: Agent 1 has to choose between x and y ; STOP.

Since agent 1 always prefers $f(C)$ to $f(L)$, he will always try to get away with announcing C in Stage 1 if he can. The question is: will agent 2 challenge at Stage 2? If the true state is L , then on being challenged agent 1 would choose y in preference to x at Stage 3; agent 2 prefers y to $f(C)$ and so in this state he will challenge. Knowing this, agent 1 will not try to dissemble at Stage 1, since he prefers $f(L)$ to y . On the other hand, if the true state is C , then on being challenged agent 1 would choose x ; agent 2 prefers $f(C)$ to x and so in this state he will not challenge. Agent 1 can therefore announce the truth with equanimity.

The key point is that if the game were in normal form, there would be a second equilibrium in state L . Namely, agent 1 announces C and threatens to choose x in preference to y ; and agent 2 doesn't challenge. This second equilibrium involves agent 1 making an incredible threat: if preferences were Leontief and Stage 3 were ever reached, he would in fact choose y . The unwanted equilibrium has been knocked out by appealing to the fact that threats have to be credible: the (subgame) perfectness requirement.

This begs the question: does there exist some other single-stage mechanism g whose unique Nash equilibrium outcome in state $\theta = C, L$ is $f(\theta)$? In short, is f Nash implementable?

Suppose that there were such a mechanism g . Then in state C , the agents have equilibrium strategies s^1 and s^2 , say, such that $g(s^1, s^2) = f(C)$. That is, in state C , agent 1 weakly prefers $f(C)$ to $g(\hat{s}^1, s^2)$ for all \hat{s}^1 , and agent 2 weakly prefers $f(C)$ to $g(s^1, \hat{s}^2)$ for all \hat{s}^2 . But from the way the lower contour sets are nested at $f(C)$, these preferences continue to hold in state L , implying that (s^1, s^2) is a Nash equilibrium in that state too—with outcome $f(C)$. This contradicts the assumption that in state L , g has a unique Nash equilibrium outcome, $f(L)$. Hence f is not Nash implementable.

Maskin (1977) provided a partial characterization of the Nash implementable choice functions. He began by showing that any Nash implementable choice function must satisfy a condition he termed *monotonicity*. Consider some profile θ , with desired outcome $f(\theta)$. And take another profile ϕ such that for every agent the outcomes that strictly dominate $f(\theta)$ under ϕ all strictly dominate $f(\theta)$ under θ too. In short, the lower contour sets of $f(\theta)$ under ϕ contain those under θ . (Roughly speaking, $f(\theta)$ moves up in everyone's rankings when switching from θ to ϕ .) If f is monotonic, then $f(\phi)$ must equal $f(\theta)$.¹⁰ The choice function in the above example is nonmonotonic and hence not Nash implementable. Maskin's remarkable contribution was next to show that if there are three or more agents, monotonicity is almost a sufficient condition for a choice function to be Nash implementable. (The "almost" refers to an additional condition termed *no veto power*, which is vacuously satisfied in economic environments if $I \geq 3$.)

How serious a drawback is the requirement of monotonicity?¹¹ At first glance, monotonicity appears to be innocuous. If it is decided to build a bridge (say) given some configuration of households' preferences, then surely the bridge ought still to be built if their preferences change in a way that makes the bridge a yet more attractive proposition to everyone? So, is the above example perverse (in that one would never wish to implement a choice function for which $f(L) \neq f(C)$)? No, not necessarily. The simple public good example may be misleading.

¹⁰ Were f a choice correspondence, rather than a function, this would need to be modified. If outcome a is the element of $f(\theta)$ having the nested lower contour sets property on switching from θ to ϕ , then $f(\phi)$ must contain outcome a .

¹¹ For application to contractual models, one clearly requires a sufficiency theorem dealing with the case $I = 2$. (And anyway, the no veto power condition on f would be hopelessly strong if there were two agents.) We have recently proved a characterization theorem for the set of Nash implementable choice functions in a two-agent model—see Moore and Repullo (1988).

It overlooks richer questions involving distribution as well as efficiency. For example, consider a choice function that specifies not only the decision whether to build the bridge, but also how much each household should contribute. Then if households' payments are positively related to their willingness to pay for the bridge, the choice function will be nonmonotonic. To see why, suppose that in state θ the vector of households' willingness to pay is $(1, 1, \dots, 1)$, whereas in state ϕ the vector is $(100, 1, \dots, 1)$. If the cost of the bridge is less than I , it is efficient to build the bridge in both states. But then monotonicity does not allow that the first household pays more in the second state—since it requires that $f(\phi) = f(\theta)$.¹²

The great attraction of the above stage mechanism is that it is so simple. There are just three stages, and the agents take turns to move. Our goal in this paper is to see how far this idea generalizes; see Section 3 for a necessary condition and Section 4 for sufficiency conditions. But we should emphasize that the mechanisms we construct (in Section 4) to deal with a general environment are far from simple: agents move simultaneously at each stage, and their strategy sets are unconvincingly rich. We present such mechanisms to demonstrate what is possible, not what is realistic.

The question then is: how simple can a mechanism be for specific applications? Section 5 answers this question for a public good model which has been prominent in the literature: namely, where agents have transferable utility. In such a model it turns out that one can implement *any* public choice function, with any vector of contributions, by using a simple, "balanced" mechanism which has sequential, not simultaneous, moves. It is not even necessary that agents have complete information about each other; it is enough that, for each agent i , only one other agent knows his preferences. We would urge a reader who is not interested in the minutiae of our general construction to turn directly to this section.

Section 6 presents three applications of our general characterization theorems (from Section 4). The first deals with a pure risk-sharing contract; the second is a production contract, or Principal-Agent, model. In both of these applications, we show that a first-best contract is nonmonotonic (and hence cannot be Nash implemented), but satisfies the sufficiency condition for subgame perfect implementation. The third application has a rather different slant: implementing the Walras correspondence.

Finally, in Section 7 we briefly remark on some more recent work by Tom Palfrey and Sanjay Srivastava, who have proceeded to investigate the use of undominated Nash and undominated Bayesian equilibria—with striking results.

We begin by introducing the notation and general definitions.

¹² Also, in a more traditional "social choice" setting, monotonicity can imply that the choice function is dictatorial. See Dasgupta, Hammond, and Maskin (1979) or Roberts (1979), where it is assumed that the domain of f contains all strict preferences, and that its range has at least three distinct outcomes.

2. NOTATION AND DEFINITIONS

Let A denote the set of possible outcomes. The profile of agents' preferences over outcomes is indexed by $\theta \in \Theta$: agent $i = 1, \dots, I$ has preference ordering $R^i(\theta)$ on the set A . Let $P^i(\theta)$ be the strict preference ordering corresponding to $R^i(\theta)$, and let $I^i(\theta)$ be the indifference relation. Given two distinct profiles θ and ϕ in Θ , there must be a pair of outcomes x and y , say, which exhibit *preference reversal* (more accurately, preference change): i.e., there is some agent i for whom $xR^i(\theta)y$ and $yR^i(\phi)x$, and at least one of these preferences is strict.

An outcome $a \in A$ is said to be θ -efficient if, for any $d \in A$, $[dR^i(\theta)a$ for all $i]$ implies $[dI^i(\theta)a$ for all $i]$.

A *choice correspondence* f selects a nonempty subset $f(\theta)$ of the set A for each profile $\theta \in \Theta$.

An *extensive game form*—or stage mechanism—with *simultaneous and/or alternating moves* is an array $g = (T, >, D, \delta, w)$, where T is a set of nodes, $>$ is a partial ordering on T that represents precedence, D is a set of possible decisions, δ is a one-to-one function from $T - \{t \in T | t \text{ has no predecessor}\}$ into D that labels each noninitial node with the last decision taken to reach it, and w is a function from $Z = \{t \in T | t \text{ has no successors}\}$ into A that associates to each terminal node the outcome that obtains at this node.¹³

In what follows we assume that there is only one initial node, t_0 , and that the maximum length of the game tree is always finite (i.e., $\sup_{t \in T} |\{t' \in T | t' < t\}| < \infty$). By a slight abuse of notation we write $\delta(t^+) = S^1(t) \times \dots \times S^I(t)$, where t^+ denotes the set of immediate successors of each nonterminal node $t \in T - Z$, and where the set $S^i(t)$ is the set of strategies for i at the node t . If $|S^i(t)| > 1$ and $|S^j(t)| > 1$ then agents i and j move simultaneously at this node, whereas if $|S^i(t)| > 1$ and $|S^j(t)| = 1$ for all $j \neq i$ then agent i is the only one to move.

It should be noted that we are making the following informational assumptions. First, at each node $t > t_0$ all agents know the entire history of the play. Second, if more than one agent moves at t , they do so simultaneously. So the mechanism g may not be one with perfect information (because this would rule out simultaneous moves). But nonetheless, for a given θ the subtree starting at t is a proper subgame. Subgame perfect equilibrium is therefore a natural solution concept.

Now define $S^i = \times_{t \in T - Z} S^i(t)$ to be the strategy set for agent i , and let $S = S^1 \times \dots \times S^I$. For each $s \in S$ and $t \in T$, let $w(s; t)$ be the outcome that obtains when agents use strategy s starting from node t (clearly $w(s; t) \equiv w(t)$ whenever $t \in Z$).

An extensive game form g together with a profile $\theta \in \Theta$ defines an extensive game (g, θ) . A (pure strategy)¹⁴ *Nash equilibrium* for this game is an element

¹³ This notation closely follows that of Kreps and Wilson (1982).

¹⁴ Although in our definition we focus on pure strategy equilibria (notice that preferences are being defined ordinally), it will become clear that our characterization theorems apply equally well to mixed strategy equilibria—with suitably defined von Neumann–Morgenstern utilities. (There is one exception: see footnote 29.) We have chosen not to state our results in this generality so as to keep notation and presentation as simple as possible.

$s = (s^i, s^{-i}) \in S$ such that, for each agent $i = 1, \dots, I$,

$$w((s^i, s^{-i}); t_0) R^i(\theta) w((\hat{s}^i, s^{-i}); t_0)$$

for all $\hat{s}^i \in S^i$. Let $NE(g, \theta)$ denote the set of Nash equilibria of the game (g, θ) and, again by a slight abuse of notation, let $w(NE(g, \theta); t_0)$ be the set of Nash equilibrium outcomes.

A Nash equilibrium $s \in NE(g, \theta)$ is said to be *subgame perfect* if, for each agent $i = 1, \dots, I$,

$$w((s^i, s^{-i}); t) R^i(\theta) w((\hat{s}^i, s^{-i}); t)$$

for all $\hat{s}^i \in S^i$ and $t \in T - Z$ (see Selten (1975)). Let $SPE(g, \theta)$ denote the set of subgame perfect equilibria of the game (g, θ) , with outcomes $w(SPE(g, \theta); t_0)$.

An extensive game form g is said to *implement* the choice correspondence f in *Nash equilibria* if for all $\theta \in \Theta$ we have $w(NE(g, \theta); t_0) = f(\theta)$.

An extensive game form g is said to *implement* the choice correspondence f in *subgame perfect equilibria* if for all $\theta \in \Theta$ we have $w(SPE(g, \theta); t_0) = f(\theta)$.

Finally in this section, we define a restricted class of environments which we call “economic”.¹⁵ In these environments, the two key conditions introduced in the following sections—viz., the necessary Condition C in Section 3, and the sufficient Condition C⁺ in Section 4—will be automatically satisfied.

An Economic Environment

There is at least one divisible private good m (say)—e.g., money—which is in fixed supply but which can be freely redistributed among the I agents provided that each agent receives a nonnegative amount.¹⁶ *Ceteris paribus*, each agent strictly prefers to receive more of good m than less. Let $A^0 \subset A$ be the set of outcomes in which every agent receives a positive amount of good m .¹⁷ It is assumed that for all $\theta \in \Theta$:

- EE(1): for each $a \in f(\theta)$, a is in A^0 ;
- EE(2): for each $a \in f(\theta)$, there is a θ -efficient outcome b in A^0 such that $a R^j(\theta) b$ for some agent j ;
- EE(3): for any $\phi \neq \theta$, there is an agent i and a pair of outcomes x and y in A^0 , such that $x P^i(\theta) y$ and $y P^i(\phi) x$;
- EE(4): for each agent i , A contains a maximal outcome under θ .

¹⁵ The referee has criticized our use of the term “economic environment”, correctly arguing that environments are customarily the domains over which choice functions are defined—without reference to the functions themselves. (This is certainly the sense in which the traditional notion of a “domain restriction” is to be understood.) We agree, and apologize for not sticking to standard terminology. Our reason is that we want to refer to “economic environments” as a catch-all for a number of (fairly weak) assumptions on both the domain *and* the choice function.

¹⁶ This is without loss of generality—since if agent $i = 1, \dots, I$ had to receive at least $m^i \neq 0$ units of the good, it would be tantamount to reducing the fixed total supply by $\sum_{i=1}^I m^i$. In short, we are normalizing A so that $m^i = 0$.

¹⁷ This could be relaxed to “in which at least two of the agents receive a positive amount of the good m .” But see footnote 19.

None of these assumptions is unreasonable. The “interiority” of outcomes a , b , x , and y in the first three assumptions should be understood in the context of the last assumption, EE(4): if an agent is allowed to choose freely from A , then we want him to avoid these particular outcomes.¹⁸ Continuity of preferences over good m will ensure assumption EE(3).¹⁹ EE(2) is perhaps the strongest assumption. The idea is that, because of conflict of interest over the private good, the “contract curve” extends as far as at least one agent’s indifference curve through outcome a .²⁰

3. NECESSARY CONDITION FOR SUBGAME PERFECT IMPLEMENTATION

In this section we prove that if a choice function f can be implemented in subgame perfect equilibria, then it must satisfy Condition C below. Although the condition is rather long, it is in fact very weak. In particular, it will be satisfied in all economic environments (as defined above); we shall justify this assertion after we have proved the theorem.

THEOREM 1: *A choice correspondence f can be implemented in subgame perfect equilibria only if it satisfies:*

CONDITION C: *For each pair of profiles θ and ϕ in Θ , and for each $a \in f(\theta) - f(\phi)$, there exists a finite sequence*

$$\sigma(\theta, \phi; a) \equiv \{ a_0 = a, a_1, \dots, a_k, \dots, a_l = x, a_{l+1} = y \} \subset A,$$

with $l = l(\theta, \phi; a) \geq 1$, such that:

- (1) *for each $k = 0, \dots, l - 1$, there is some particular agent $j(k) = j(k|\theta, \phi; a)$, say, for whom*

$$a_k R^{j(k)}(\theta) a_{k+1}; \text{ and}$$

- (2) *there is some particular agent $j(l) = j(l|\theta, \phi; a)$, say, for whom*

$$[x =] a_l R^{j(l)}(\theta) a_{l+1} [= y] \text{ and } [y =] a_{l+1} P^{j(l)}(\phi) a_l [= x].$$

Further, $l(\theta, \phi; a)$ is uniformly bounded by some $\bar{l} < \infty$.

Before presenting the formal proof, it may be helpful to give the intuition. For simplicity, suppose that there are just two agents. (The argument generalizes.) Consider two profiles θ and ϕ , and an outcome a which is contained in $f(\theta)$ but

¹⁸ If an agent has no best outcome in A —e.g., because A is not compact—then we could instead operate within a compact box B contained in A . See Condition C⁺ in Section 4.

¹⁹ Actually, in Sections 3 and 4, we shall just need $xR^i(\theta)y$ and $yP^i(\phi)x$ for some i . There is another possibility—namely, that the *only* outcomes x and y in A^0 exhibiting preference reversal are those for which $xP^i(\theta)y$ and $yI^i(\phi)x$. But this will not be the case if agent i ’s preferences are continuous in his (interior) allocation of good m at x or y . (If footnote 17 applied, the argument would need to be modified a little, because this agent i ’s allocation of good m at x and y might not be interior.)

²⁰ There is no reason why outcome a should itself be θ -efficient. But if it were, then EE(2) would of course be automatically satisfied.

not in $f(\phi)$. If some mechanism g implements f in subgame perfect equilibria, then it must be the case that under θ there is an equilibrium²¹— s_a , say—of g whose outcome is a . Look at the first stage of g , and construct a “reduced-form” outcome matrix, where each entry is the equilibrium (arising from s_a) of the subgame starting at the corresponding second-stage node. Now examine the row and column containing outcome a , and ask the question: do *all* the outcomes in this row and this column continue to be equilibria of their respective subgames as we switch from θ to ϕ ? If the answer is yes, we are done. Outcome a cannot continue to be an equilibrium of g under ϕ , since a is not contained in $f(\phi)$. So there must be some outcome y , say, contained in this row or this column such that there is preference reversal, for the column or row player respectively, over the pair $(a = x, y)$. That is, letting this player be agent $j(0)$, part (2) of Condition C holds. If the answer is no, then there is some outcome a_1 , say, in either this row or this column which was an equilibrium of the corresponding subgame under θ , but not under ϕ . Under θ , the column or row player—agent $j(\theta)$, say—must weakly prefer outcome a to a_1 , since s_a is an equilibrium. That is, part (1) of Condition C holds for $k = 0$. But now we can repeat the entire argument, turning the spotlight away from outcome a and onto outcome a_1 : why does a_1 cease to be an equilibrium outcome (of its subgame) when switching from θ to ϕ ? The iteration has to stop somewhere because, by assumption, g has a finite number of stages.

PROOF: Take an arbitrary pair of profiles θ and ϕ in Θ , and suppose that there is some outcome $a \in f(\theta) - f(\phi)$. Since g implements f in subgame perfect equilibria, there must be some s_a , say, contained in $SPE(g, \theta)$, such that

$$(\dagger.0) \quad a = w(s_a; t_0) \neq w(s; t_0) \quad \text{for all } s \in SPE(g, \phi).$$

If g has only one stage (i.e., if $t_0^+ - Z = \emptyset$), Condition C is trivially satisfied. To see why, write

$$s_a = s_a(t_0) = (s_a^1(t_0), \dots, s_a^J(t_0)).$$

It must be the case that for some agent i and some $\hat{s}^i(t_0) \neq s_a^i(t_0)$ in $S^i(t_0)$,

$$w((\hat{s}^i(t_0), s_a^{-i}(t_0)); t_0) P^i(\phi) w(s_a(t_0); t_0)$$

—otherwise s_a would belong to $SPE(g, \phi)$, which contradicts $(\dagger.0)$. Let

$$y = w((\hat{s}^i(t_0), s_a^{-i}(t_0)); t_0).$$

Then the sequence $\{a_0 = a, a_1 = x = a, a_2 = y\}$ satisfies (1) and (2) in Condition C. Specifically, $y P^i(\phi) x$; and $x R^i(\theta) y$ since $s_a \in SPE(g, \theta)$.

The important case is where g has more than one stage (i.e., where $t_0^+ - Z \neq \emptyset$). Write

$$s_a = (s_a(t_0), s_a^+),$$

²¹ To make this paragraph easier to read, we shall simply refer to “equilibria”, rather than subgame perfect equilibrium strategies or outcomes.

where s_a^+ is the set of vectors of strategies specified in s_a for each subgame starting at nodes $t \in t_0^+ - Z$.

Define a subset t_0^{++} of t_0^+ as follows: $t \in t_0^+$ is contained in t_0^{++} if and only if there is some agent i and some $\hat{s}^i(t_0)$ in $S^i(t_0)$ such that

$$\delta(t) = (\hat{s}^i(t_0), s_a^{-i}(t_0)).$$

For each $t \in t_0^{++} - Z$, define $SPE(g|t, \phi)$ to be the set of subgame perfect equilibria under ϕ of the subgame starting at node t .

There are two possibilities to consider.

First, suppose that for each $t \in t_0^{++} - Z$,

$$w((\delta(t), s_a^+); t_0) \in w(SPE(g|t, \phi); t).$$

Then it must be the case that for some agent i and some $\hat{s}^i(t_0) \neq s_a^i(t_0)$ in $S^i(t_0)$,

$$w(((\hat{s}^i(t_0), s_a^{-i}(t_0)), s_a^+); t_0) P^i(\phi) w(s_a; t_0)$$

—otherwise s_a would belong to $SPE(g, \phi)$, which contradicts (†.0). Let

$$y = w(((\hat{s}^i(t_0), s_a^{-i}(t_0)), s_a^+); t_0).$$

Then, as in the case where g has only one stage, the sequence $\{a_0 = a, a_1 = x = a, a_2 = y\}$ satisfies (1) and (2) in Condition C.

Second, suppose that for some $t_1 \in t_0^{++} - Z$,

$$w((\delta(t_1), s_a^+); t_0) \notin w(SPE(g|t_1, \phi); t_1).$$

Let $i = j(0)$ be the agent satisfying $\delta(t_1) = (\hat{s}^i(t_0), s_a^{-i}(t_0))$ for some $\hat{s}^i(t_0) \in S^i(t_0)$. And let

$$a_1 = w((\delta(t_1), s_a^+); t_0) = w(s_a; t_1).$$

Then $aR^{j(0)}(\theta)a_1$, since $s_a \in SPE(g, \theta)$. Moreover, by the choice of t_1 ,

$$(\dagger.1) \quad a_1 = w(s_a; t_1) \neq w(s; t_1) \quad \text{for all } s \in SPE(g|t_1, \phi).$$

Notice that (†.1) is formally identical to (†.0). We can therefore continue iteratively. The iteration must stop since, by assumption, g has a finite number of stages. Eventually, then, we will generate a sequence $\{a_0 = a, a_1, \dots, a_l = x, a_{l+1} = y\}$ satisfying (1) and (2) in Condition C. Moreover, there is an upper bound \bar{l} on l , which is uniform across θ, ϕ , and a . Q.E.D.

We now justify the assertion made at the beginning of this section, that Condition C will be satisfied in all economic environments. Actually, we shall show more: that the upper bound \bar{l} in Condition C is (at most) 2. For future reference, this means that the mechanisms we construct to prove the sufficiency theorems (in the next section) will not have more than 3 (i.e., $\bar{l} + 1$) stages in economic environments.

Take a pair of profiles $\theta, \phi \in \Theta$, and an outcome $a \in f(\theta) - f(\phi)$. By EE(3), $\theta \neq \phi$ implies that there is some “test agent”—say agent i —whose preferences

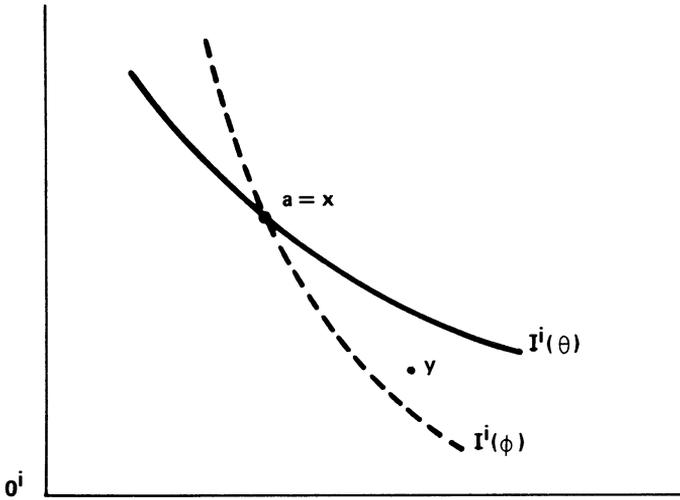


FIGURE 2.1(a)

are reversed somewhere in A^0 : i.e.,

$$xP^i(\theta)y \text{ and } yP^i(\phi)x$$

for some pair of outcomes (x, y) . We shall refer to any such x and y as a “test pair.” (Of course, different test pairs may be associated with different test agents—it may not be agent i .)

There are three possible cases to consider. It is helpful to represent these cases in standard two-commodity diagrams; see Figures 2. (A word of caution, though: apart from the good m , there is no presupposition that the other dimensions of A are necessarily private and/or divisible goods.) In these figures, $I^i(\theta)$ [resp. $I^i(\phi)$] refers to agent i 's indifference curve under profile θ [resp. profile ϕ]. Utility for agent i is increasing away from his “origin” O^i .

CASE 1: *There exists a test pair (a, y) .*

This case arises whenever there is some agent, say agent i , whose ϕ -lower contour set at outcome a does not contain his θ -lower contour set; i.e., if Figure 2.1(a) or 2.1(b) pertains.

If a test pair with $x = a$ can be found, then we define the sequence in Condition C to be simply $\{a_0 = a, a_1 = a, a_2 = y\}$, and $l = 1$.²² Note that $j(0) = j(1) = i$.

If a test pair with $x = a$ cannot be found, then it means that all the agents' ϕ -lower contour sets at outcome a contain their respective θ -lower contour sets.

²² It is convenient, for the next section, to require that the sequence has at least three elements. That is why we put $a_0 = a_1$ in Case 1.

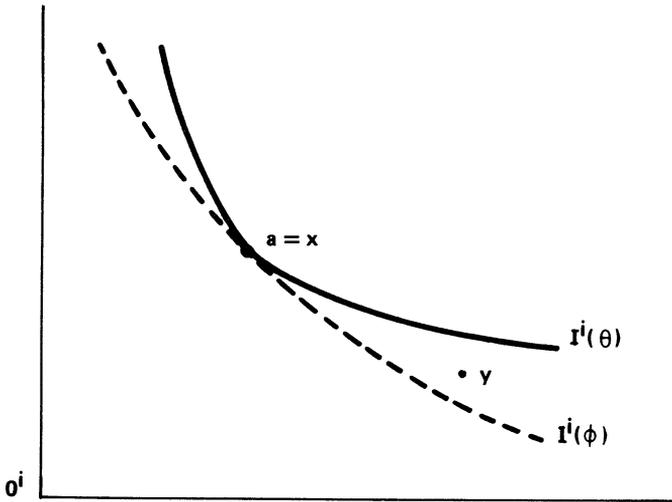


FIGURE 2.1(b)

The sequence in Condition C is then a little more awkward to define:

CASE 2: *There does not exist a test pair (a, y). However, a test pair (x, y) exists for which the outcome x does not strictly Pareto dominate outcome a under θ.*

There are two subcases. If, for some agent i , his θ -lower contour set at outcome a is locally strictly nested inside his ϕ -lower contour set, then the test pair (x, y) can be found in the neighborhood of outcome a —with $aR^i(\theta)x$. See Figure 2.2(a). On the other hand, if no agent's preferences change in the neighborhood of outcome a as we switch from θ to ϕ , then Figure 2.2(b) applies. In this “Edgeworth box” diagram, we are supposing that agent i is the test agent for the pair (x, y) , and $aR^j(\theta)x$ for some agent j . (Agent i could in principle be agent j , though; i.e., it may be that $aR^i(\theta)x$.) We can define the sequence in Condition C to be $\{a_0 = a, a_1 = x, a_2 = y\}$, and as in Case 1, $l = 1$. Note that $j(0) = j$ and $j(1) = i$.

CASE 3: *There does not exist a test pair (a, y). Moreover, the only test pairs (x, y) are those for which outcome x strictly Pareto dominates outcome a under θ.*

This is a fairly unlikely possibility: no agent's preferences change anywhere in his respective lower contour set of outcome a as we switch from θ to ϕ . Nevertheless, the sequence in Condition C can be defined with $l = 2$, thanks to EE(2). This assumption ensures that there exists a θ -efficient outcome b and agent j , say, who weakly prefers outcome a to b under θ . That is, $xP^j(\theta)aR^j(\theta)b$.

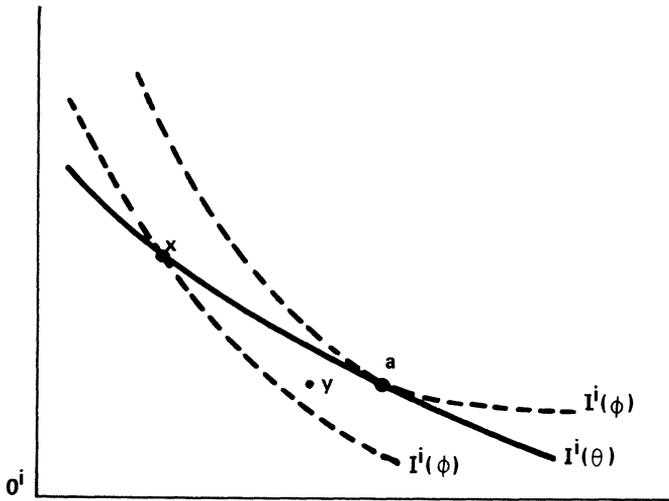


FIGURE 2.2(a)

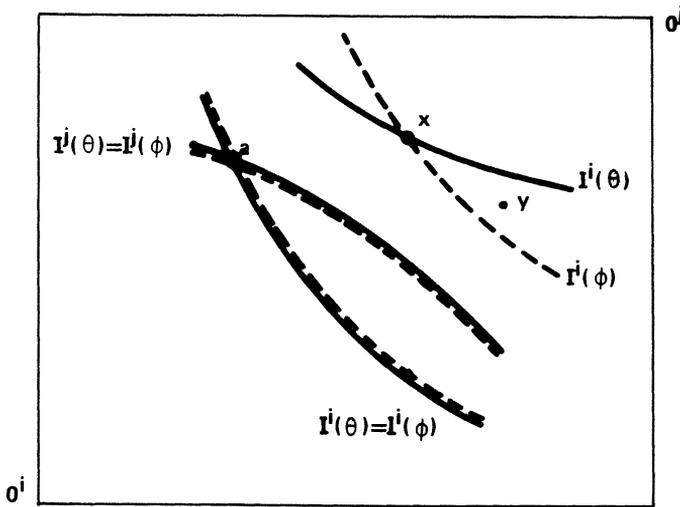


FIGURE 2.2(b)

But since b is θ -efficient, there must be a second agent who prefers outcome b to x under θ . Let us take this second agent to be the test agent, i . (We make this assumption in order to be able to represent the situation in an “Edgeworth box”; see Figure 2.3. In general, the test agent could be different from either of the first two agents.) We can define the sequence in Condition C to be $\{a_0 = a, a_1 = b, a_2 = x, a_3 = y\}$, and $l = 2$. Note that $j(0) = j$, and $j(1) = j(2) = i$.

Only in Case 3, then, does l exceed 1, and in no case does l exceed 2. In sum, Condition C is always satisfied, and $\bar{l} = 2$, in economic environments.

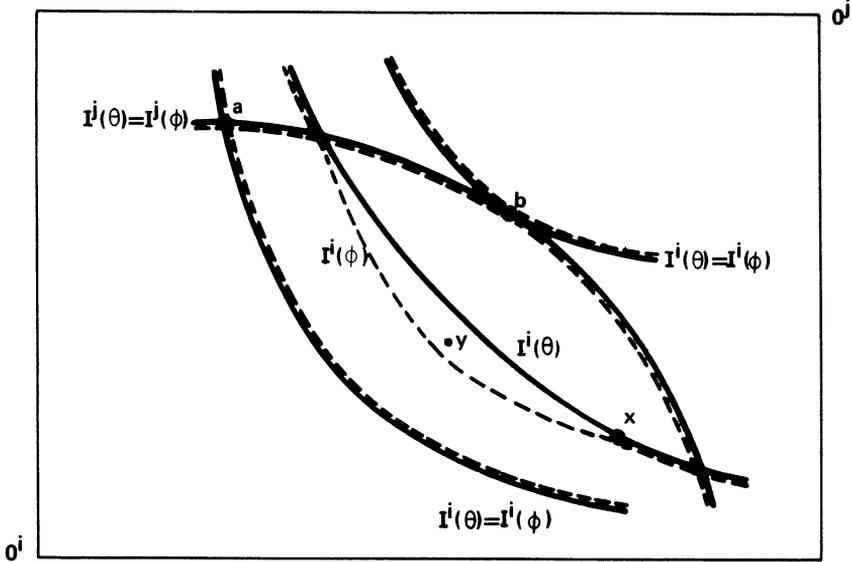


FIGURE 2.3

4. SUFFICIENT CONDITIONS FOR SUBGAME PERFECT IMPLEMENTATION

In this section we find sufficient conditions under which a choice correspondence f can be implemented in subgame perfect equilibria. The discussion is divided into two parts: first, for the case where there are three or more agents ($I \geq 3$); second, for the case where there are just two agents ($I = 2$).

4.1. Three or More Agents

Before stating Condition C^+ below we need to introduce one further piece of notation. Assuming Condition C holds, let $\mathcal{Q}(f)$ be a class of subsets Q of A . A typical Q is defined as follows:

For each pair of profiles θ and ϕ in Θ , and for each $a \in f(\theta) - f(\phi)$, select *one* sequence $\sigma(\theta, \phi; a)$ satisfying (1) and (2) in Condition C. Then let Q be the union of the elements in these sequences.²³

$\mathcal{Q}(f)$ comprises the Q 's constructed from *all* possible selections.

CONDITION C^+ : *Condition C holds. Further, there exists a particular $Q^+ \in \mathcal{Q}(f)$, and a particular set B (in A) containing Q^+ , such that the following is true for each θ in Θ :*

Each agent $i = 1, \dots, I$ has a nonempty maximal set $M^i(\theta)$ in B under θ . These I maximal sets are pairwise disjoint, and none of them intersects with Q^+ .

²³ In the unlikely circumstance that some outcome is contained in $f(\theta)$ for *all* $\theta \in \Theta$, then this outcome should be added to each of the sets Q .

Let the selected sequences $\sigma(\theta, \phi; a)$ in Q^+ be labelled $\sigma^+(\theta, \phi; a)$.

Again, although the statement of Condition C^+ is lengthy, it is in fact weak. Like Condition C , it will be satisfied in all economic environments, as defined in Section 2. The point is quite simple. We can set $B = A$ so that, by EE(4), under any given profile θ , each agent i has a nonempty maximal set $M^i(\theta)$ in B . Moreover, any outcome in this maximal set is such that agent i gets all the private good m (implying that the outcome cannot lie in A^0). So the I maximal sets $\{M^1(\theta), \dots, M^I(\theta)\}$ are pairwise disjoint. Finally, let Q^+ comprise the outcomes contained in the sequences $\{a, a, y\}$, $\{a, x, y\}$, $\{a, b, x, y\}$ which we constructed at the end of the previous section. From EE(1)–(3), we know that all these outcomes are “interior” (i.e., lie in A^0), and therefore Q^+ will not intersect with $M^i(\theta)$ for any i or θ .

THEOREM 2: *If a choice correspondence f satisfies Condition C^+ , and $I \geq 3$, then f can be implemented in subgame perfect equilibria.²⁴*

PROOF: Define a mechanism g as follows.

First Stage (Stage 0): Each agent $i = 1, \dots, I$ announces some triplet (θ^i, a^i, n_0^i) , say, where $\theta^i \in \Theta$, $a^i \in f(\theta^i)$, and n_0^i is a nonnegative integer. There are three possibilities to consider:

- (0.1) If all I agents agree on a common profile θ and outcome $a \in f(\theta)$, i.e., if $\theta^i \equiv \theta$ and $a^i \equiv a$, then outcome a is implemented. STOP
- (0.2) If only $I - 1$ agents agree on a common profile θ and outcome $a \in f(\theta)$, and if the remaining agent i announces a profile ϕ , and
 - (0.2.1) if $a \in f(\phi)$, then outcome a is implemented; STOP
 - (0.2.2) if $a \notin f(\phi)$ but i is not the agent $j(0)$ prescribed in $\sigma^+(\theta, \phi; a)$, then outcome a is implemented; STOP
 - (0.2.3) if $a \notin f(\phi)$ and $i = j(0)$, then go to Stage 1.

²⁴ Since the previous version of this paper (Moore and Repullo (1986)) was circulated, Dilip Abreu and Arunava Sen (1987) have essentially closed the gap between our necessary Condition C and our sufficient Condition C^+ . To understand their new condition, “Condition α ”, refer back to the argument in the paragraph following the statement of Theorem 1, and let B denote the entire set of terminal outcomes of the mechanism g . The argument can be taken further. Specifically, one must be able to find sequences $\sigma(\theta, \phi; a)$ which have the additional property that for $k = 0, \dots, l - 1$,

(\ddagger) a_k is not maximal in B under ϕ for agent $j(k)$

—where $j(k) = j(k|\theta, \phi; a)$ is the agent for whom $a_k R^{j(k)} a_{k+1}$. To see why, consider the case $k = 0$. Recall that $a_0 = a$, which is an outcome contained in $f(\theta)$ but not in $f(\phi)$. By construction, a_0 cannot be an equilibrium of g under ϕ . Therefore in searching for the appropriate agent $j(0)$ —together with outcome a_1 —we can limit ourselves to those agents for whom a_0 is not maximal in B under ϕ . This argument applies at each of the subsequent iterations $k = 1, \dots, l - 1$ of the proof. Abreu and Sen added (\ddagger) to our necessary Condition C . (In fact, their Condition α is a little stronger than this. And their argument is more delicate because they consider arbitrary extensive form games.) They have shown that Condition α , together with no veto power and $I \geq 3$, is sufficient for a choice correspondence to be implementable in subgame perfect equilibria (using the same mechanism as we construct below to prove Theorem 2). This is nicely consonant with Maskin’s sufficiency theorem for Nash implementation (see the Introduction). Moreover, Abreu and Sen’s condition not only avoids the additional complexity of our Condition C^+ , but also enables them to determine whether or not certain choice correspondences from the voting literature are subgame perfect implementable, even though they lie in the “grey area” between our Conditions C and C^+ .

- (0.3) If neither (0.1) nor (0.2) apply, then the agent with the highest integer n_0^i is allowed to choose²⁵ an outcome from B . Ties are broken by, say, selecting from the agents who announced the highest number according to who has the smallest i . STOP

Subsequent Stages (Stages $k = 1, \dots, l$): Each agent $i = 1, \dots, I$ can either raise a “flag”,²⁶ or announce a nonnegative integer n_k^i , say. Again there are three possibilities to consider:

- (k.1) If $I - 1$ or more flags are raised, then the agent $j(k - 1)$ prescribed in $\sigma^+(\theta, \phi; a)$ is allowed to choose an outcome from B . STOP
- (k.2) If $I - 1$ or more agents announce zero, and
- (k.2.1) if the agent $j(k)$ prescribed in $\sigma^+(\theta, \phi; a)$ is one of those who announce zero, then implement outcome a_k from sequence $\sigma^+(\theta, \phi; a)$ [recall that $a_k = x$ if $k = l$]; STOP
- (k.2.2) if $j(k)$ does not announce zero, then
 if $k < l$, go to Stage $k + 1$;
 if $k = l$, implement outcome $a_{i+1}[=y]$ from sequence $\sigma^+(\theta, \phi; a)$. STOP
- (k.3) If neither (k.1) nor (k.2) apply, then the agent who announces the highest integer n_k^i is allowed to choose an outcome from B . (In this comparison, raising a flag counts as -1 , say.) Ties are broken as in (0.3). STOP

Suppose the true profile is some $\theta^* \in \Theta$. Then for each $a \in f(\theta^*)$ there is an equilibrium of g in which all agents announce a common profile at the first stage (Stage 0) together with outcome a . Also, their strategies are to announce zero at any subsequent stage (Stages $k = 1, \dots, l$)—if it were ever reached. In equilibrium, the game stops at Stage 0, and outcome a is implemented.

Moreover, these are the only equilibria of g .

To understand why $f(\theta^*)$ equals the set of subgame perfect equilibrium outcomes of g , we first need to observe:

CLAIM: At any Stage $k = 1, \dots, l$, there are only two possible equilibria: either all the agents raise their flags (denote this event by F_k), or they all announce zero (denote this event by Z_k).²⁷

PROOF OF CLAIM: The “integer game” (k.3) clearly has no equilibria. There are therefore only two other candidates, apart from F_k and Z_k , for equilibrium at Stage k :

- (i) Exactly $I - 1$ agents raise their flags. This cannot be an equilibrium because there is at least one agent $i \neq j(k - 1)$ who must be raising his flag whereas he

²⁵ In principle, this choice could be made part of this agent’s strategy at the outset (of this Stage). We have opted to separate out this choice so as to simplify the presentation of the mechanism. The same applies to the choices made at substages (k.1), (k.3).

²⁶ If one preferred, “flag” could be defined to be the integer -1 .

²⁷ Actually, this is not strictly true when there are exactly three agents. If Z_k —as well as F_k —is an equilibrium at Stage k (see below for circumstances in which this is the case), and if $I = 3$, then there is a mixed-strategy equilibrium too. However, it will be seen that this makes no difference to our final conclusion.

could do strictly better by announcing a sufficiently high integer n_k^i in order to be able to choose from B .

(ii) Exactly $I - 1$ agents announce zero. This cannot be an equilibrium either. There are two cases to consider. If agent $j(k)$ is among those who are announcing zero, the outcome is a_k . But then $j(k)$ can do strictly better by announcing a sufficiently high integer in order to be able to choose from B . On the other hand, if agent $j(k)$ is not announcing zero, then (by assumption) all the other agents are doing so. The game proceeds to the subgame starting at Stage $k + 1$, for which there is some (subgame perfect) equilibrium outcome. It cannot of course be the case that any more than one of the $I - 1$ (≥ 2) agents $i \neq j(k)$ will end up choosing from B in this equilibrium. So there must be at least one agent i who could do strictly better by announcing a sufficiently high integer n_k^i in order to be able to choose from B now (i.e., at Stage k).

The claim is proved.

F_k is *always* an equilibrium at Stage $k = 1, \dots, l$. Moreover, if at Stage $k + 1$ (for $1 \leq k < l$) all agents are going to adopt the strategy of raising their flags, then F_k is the *unique* equilibrium at Stage k . (In these circumstances, Z_k —with outcome a_k —would not be an equilibrium since agent $j(k)$ would be strictly better off not announcing zero in order to be able to choose from B at Stage $k + 1$.)

To reach any Stage $k > 0$, it must have been the case that (0.2.3) applied at the first stage (Stage 0). Let θ be the profile that was (commonly) announced by $I - 1$ agents at Stage 0, and let ϕ be the profile that was announced by the remaining agent.

Can Z_k in fact be an equilibrium of Stage k , if $1 \leq k < l$? Yes, if $\theta = \theta^*$ and at Stage $k + 1$ all agents adopt the strategy of announcing zero. (Agent $j(k)$ no longer has an incentive not to announce zero, because the equilibrium outcome at Stage $k + 1$ is a_{k+1} , which is no better for him than a_k .)

Next we ask the crucial question: is Z_l (in addition to F_l) an equilibrium of the final stage, l ? This depends on how the true profile θ^* compares with profiles θ and ϕ announced at the first stage. Z_l leads to outcome x . If $\phi = \theta^* \neq \theta$, then agent $j(l)$ would strictly prefer not to announce zero in order to implement y ; hence Z_l is *not* an equilibrium. However if $\phi \neq \theta^* = \theta$, then agent $j(l)$ would be as well off announcing zero along with the other $I - 1$ agents; here Z_l is an equilibrium.

Let us sum up what we have learned so far. Suppose that following the agents' announcements at Stage 0, it turns out that (0.2.3) applies. Then one equilibrium outcome of the subgame starting at Stage 1 is that agent $j(0)$ chooses from B (equilibrium F_1). This will be the *unique* equilibrium outcome whenever $\phi = \theta^* \neq \theta$. However if $\phi \neq \theta^* = \theta$, then there is a second equilibrium (equilibrium Z_1), with outcome a_1 .

Now we turn to Stage 0. The "integer game" (0.3) clearly has no equilibria. Also, it cannot be the case that in equilibrium (0.2) applies—i.e., that only $I - 1$ agents announce a common profile θ and a common outcome $a \in f(\theta)$. (The

reasoning is the same as that given above to show that at Stage $k > 0$ it is not an equilibrium for exactly $I - 1$ agents to announce zero.) The only vectors of strategies which are candidates for equilibrium at Stage 0 are therefore ones in which all I agents announce a common profile θ and a common outcome $a \in f(\theta)$. From what we know about the possible equilibria at Stage 1, such a vector cannot be an equilibrium if $\theta \neq \theta^*$ and $a \notin f(\theta^*)$. Why not? Because the agent $j(0)$ prescribed in $\sigma^+(\theta, \theta^*; a)$ can do strictly better by announcing $\phi = \theta^*$ instead of θ in order to be able to choose from B at Stage 1 (where F_1 is the unique equilibrium).

There is only one candidate left. Namely, at Stage 0 all the agents announce a common profile together with a common outcome a contained in $f(\theta^*)$. And outcome a is then implemented.

Finally, then, take any a in $f(\theta^*)$. The following vector of strategies is a subgame perfect equilibrium of g whose outcome is a . Each agent $i = 1, \dots, I$ announces the triplet $(\theta^i, a^i, n_0^i) = (\theta^*, a, 0)$ at Stage 0, and $n_k^i = 0$ at all subsequent Stages $k = 1, \dots, l$. Notice that no agent has an incentive to change his Stage 0 announcement to some $\phi \neq \theta^*$ for which $a \in f(\theta^*) - f(\phi)$ and he is the agent $j(0)$ prescribed in $\sigma^+(\theta^*, \phi; a)$ —since the equilibrium outcome at Stage 1 is a_1 , which is no better for him than a .

Theorem 2 is proved.

Q.E.D.

4.2. Two Agents

For two agents, the sufficient Condition C^+ is strengthened to:

CONDITION C^{++} : $I = 2$ and Condition C^+ holds. Further, in that Condition, Q^+ and B can be chosen so that there exists some particular outcome $z \in A$ which is “bad” in the following sense:

Take any $\theta \in \Theta$. Then for both agents z is strictly dominated under θ by any outcome in Q^+ , and is weakly dominated by any outcome in $M^1(\theta) \cup M^2(\theta)$.

Although Condition C^{++} will not be satisfied for all economic environments (as defined in Section 2), nevertheless it is a reasonable condition. For example, consider the following circumstances. (a) Define Q^+ to comprise the outcomes contained in the sequences $\{a, a, y\}$, $\{a, x, y\}$, $\{a, b, x, y\}$ which we constructed at the end of the previous section. All these outcomes lie in a subset A^{00} of A^0 with the property that neither agent $i = 1, 2$ receives strictly less than $\hat{m}^i > 0$ (say) of good m . (b) Each agent i always has a nonempty maximal set in A^{00} . And (c) each agent i always prefers an outcome with $m^i \geq \hat{m}^i$ of good m to any outcome with none of the good. In these circumstances, we could define z to be some outcome with $m^1 = m^2 = 0$ (assuming, of course, that such an outcome z belongs to A , i.e., that free disposal of good m is feasible). By defining B to equal A^{00} , it would follow that for both agents, and under any profile θ in Θ , outcome z would be strictly dominated by any outcome in Q^+ and by any outcome in the two maximal sets in B . That is, Condition C^{++} would be satisfied.

THEOREM 3: *If a choice correspondence f satisfies Condition C^{++} , and $I = 2$, then f can be implemented in subgame perfect equilibria.*

PROOF: Define a mechanism g as follows.

First Stage (Stage 0): Each agent $i = 1, 2$ announces some triplet (θ^i, a^i, n_0^i) , say, where $\theta^i \in \Theta$, $a^i \in f(\theta^i)$, and n_0^i is a nonnegative integer. There are four possibilities to consider:

- (0.1) If the two agents do not agree on a common outcome a which is contained in $f(\theta^1) \cap f(\theta^2)$, then the bad outcome z is implemented. STOP
- (0.2) If the two agents agree on a common outcome a which is contained in $f(\theta^1) \cap f(\theta^2)$, and if both announce zero, then outcome a is implemented. STOP
- (0.3) If the two agents agree on a common outcome a which is contained in $f(\theta^1) \cap f(\theta^2)$, and if exactly one agent, say agent i , announces a positive integer (i.e., if $n^i > 0$ but $n^{-i} = 0$), then he is asked to choose²⁸ another profile ϕ . And
 - (0.3.1) if $a \in f(\phi)$, then outcome a is implemented; STOP
 - (0.3.2) if $a \notin f(\phi)$ but i is not the agent $j(0)$ prescribed in $\sigma^+(\theta, \phi; a)$, then outcome a is implemented; STOP
 - (0.3.3) if $a \notin f(\phi)$ and $i = j(0)$, then go to Stage 1.
- (0.4) If none of (0.1)–(0.3) applies, then the agent with the highest integer n_0^i is allowed to choose an outcome from B . Ties are broken by, say, selecting agent 1. STOP.

Subsequent Stages (Stages $k = 1, \dots, l$): Each agent $i = 1, 2$ can either raise a “flag”, or announce a nonnegative integer n_k^i , say. Again there are four possibilities to consider:

- (k.1) If two flags are raised, then the agent $j(l-1)$ prescribed in $\sigma^+(\theta, \phi; a)$ is allowed to choose an outcome from B . STOP
- (k.2) If one flag is raised, then the bad outcome z is implemented. STOP
- (k.3) If no flags are raised, and if at least one of the agents announces zero, and
 - (k.3.1) if the agent $j(k)$ prescribed in $\sigma^+(\theta, \phi; a)$ announces zero, then implement outcome a_k from sequence $\sigma^+(\theta, \phi; a)$ [recall that $a_k = x$ if $k = l$]; STOP
 - (k.3.2) if $j(k)$ announces a positive integer, then
 - if $k < l$, go to Stage $k + 1$;
 - if $k = l$, implement outcome $a_{l+1}[= y]$ from sequence $\sigma^+(\theta, \phi; a)$. STOP
- (k.4) If none of (k.1)–(k.3) applies (i.e., if both agents announce positive integers $n_k^i > 0$), then the agent who announces the highest integer is allowed to choose an outcome from B . Ties are broken as in (0.4). STOP

²⁸ Footnote 25 also applies here, and a⁺ substages (0.4), (k.1), (k.4).

The detailed demonstration of why, for each $\theta^* \in \Theta$, $f(\theta^*)$ equals the set of subgame perfect equilibrium outcomes of g , is similar to that given for Theorem 2.²⁹ Rather than repeat it here, we briefly comment on the need for the bad outcome z . As can be seen from the proof of Theorem 2, when $I \geq 3$, any one agent can claim that the other agents are cheating. Specifically, if we have $I - 1$ (≥ 2) agents announcing a common outcome a contained in some $f(\theta)$, then the remaining agent can accuse them of lying by announcing a different profile ϕ for which $f(\phi)$ does not contain outcome a . When there are only two agents this approach breaks down: how could one tell which agent was accusing the other of lying? We therefore have to construct a more delicate argument, which exploits the bad outcome z . Basically, if agents disagree on their initial announcements, then outcome z is implemented. Their (common) veracity is ensured by using the numbers n_0^1 and n_0^2 in conjunction with what would happen at subsequent stages.

5. A SIMPLE MECHANISM: THE CASE OF TRANSFERABLE UTILITY

The mechanisms we constructed for our general sufficiency Theorems 2 and 3 were complicated. In this section we give an example, concerned with the provision of a public good, which illustrates that for specific applications the mechanism can be refreshingly simple and yet still highly effective. Our example takes the case where agents, or households, have transferable utility. (We shall think of it as money.) This case has been prominent in the literature.

Suppose that there are two agents. (A priori, one might expect that the two-agent case is the most difficult. It will be seen that the mechanism we design is easily adapted to more than two agents.) Their respective utility functions are:

$$\text{Agent 1: } u_1(d, \theta_1) - t_1,$$

$$\text{Agent 2: } u_2(d, \theta_2) + t_2$$

—where d is the public good decision, t_1 and t_2 are monetary transfers (from agent 1 and to agent 2), and parameters (θ_1, θ_2) are contained in $\Theta_1 \times \Theta_2$. Of course, t_i may be negative ($i = 1, 2$). The agents observe each others' θ_i 's, but these are not publicly known.

Consider how one might implement some general choice function $f = (D, T_1, T_2)$ —where for each profile (θ_1, θ_2) in $\Theta_1 \times \Theta_2$, the public good decision is given by $d = D(\theta_1, \theta_2)$, and the transfers are given by $(t_1, t_2) =$

²⁹ There is one difficulty. Suppose that there are two distinct outcomes, a and a' , which are both contained in $f(\theta)$ and $f(\theta')$ for some pair of profiles θ and θ' . (θ and θ' need not be distinct.) Then, under θ or θ' there is an unwanted mixed strategy equilibrium in which the both agents announce zero, but randomize between announcing outcomes a and a' at Stage 0. (Of course, as Theorem 3 states, under either θ or θ' , there are also two pure strategy equilibria with outcomes a and a' .) This is a problem which will not arise if f is a choice function.

If f is a choice correspondence, however, the problem could be partially alleviated by introducing an earlier stage (Stage “-1”) to the game which effectively makes a random selection of a choice function from f . By doing this, we can guarantee that all final equilibrium outcomes in either state θ or θ' will lie in $f(\theta)$ or $f(\theta')$ respectively—where “final” means after the random selection has been made at Stage -1.

$(T_1(\theta_1, \theta_2), T_2(\theta_1, \theta_2))$, say. We place no restrictions on f, u_1 or u_2 , other than to assume that utility is bounded: see (*) below. So f may be nonmonotonic and therefore not Nash implementable. Nevertheless, f can be implemented as the (unique) subgame perfect equilibrium outcome of the following mechanism, provided $\Delta t \gg 0$ is sufficiently large:

STAGE 1 (eliciting agent 1's preferences):

Stage 1.1: Agent 1 announces his preferences $\theta_1 \Rightarrow$ Stage 1.2.

Stage 1.2: Agent 2 can:

either agree \Rightarrow Stage 2.

or challenge, by announcing some $\phi_1 \neq \theta_1 \Rightarrow$ Stage 1.3.

Stage 1.3: Agent 1 has to choose between $\{d; t_1\} = \{x; t_x + \Delta t\}$ and $\{y; t_y + \Delta t\}$ —where $x = D_x(\theta_1, \phi_1)$, $y = D_y(\theta_1, \phi_1)$, $t_x = T_x(\theta_1, \phi_1)$, and $t_y = T_y(\theta_1, \phi_1)$ are such that

$$u_1(x, \theta_1) - t_x > u_1(y, \theta_1) - t_y, \quad \text{and}$$

$$u_1(x, \phi_1) - t_x < u_1(y, \phi_1) - t_y.$$

Further, if agent 1 chooses $\{x; t_x + \Delta t\}$, then agent 2 receives $t_2 = t_x - \Delta t$ (and the planner nets $2\Delta t$)—whereas if agent 1 chooses $\{y; t_y + \Delta t\}$, then agent 2 receives $t_2 = t_y + \Delta t$ (and the planner breaks even). STOP

STAGE 2: Same as Stage 1, except that the roles of agents 1 and 2 are switched, in order to elicit agent 2's preferences.

If at Stage 2.1 agent 2 has announced his preferences as θ_2 , say, and at Stage 2.2 agent 1 has agreed, then implement the public good decision $d = D(\theta_1, \theta_2)$, and the transfers $t_1 = T_1(\theta_1, \theta_2)$ and $t_2 = T_2(\theta_1, \theta_2)$. STOP

Our bounded utility assumption is as follows. For $i = 1, 2$ let \mathcal{R}_i denote the range of (D, T_i) from the domain $\Theta_1 \times \Theta_2$, and for $z = x, y$ let \mathcal{R}_z denote the range of (D_z, T_z) from the domain $(\Theta_1)^2 \cup (\Theta_2)^2$. Then for $i = 1, 2$:

$$(*) \quad u_i(d, \theta_i^*) + (-1)^i t_i \text{ is uniformly bounded for all } \theta_i^* \text{ in } \Theta_i$$

$$\text{and for all } (d, t_i) \text{ in } \mathcal{R}_i \cup \mathcal{R}_x \cup \mathcal{R}_y.$$

This assumption will automatically be satisfied if $\Theta_1 \times \Theta_2$ is finite.

The key point of the mechanism is that at Stage i ($i = 1, 2$), agent i will announce his true preferences θ_i and the other agent will not challenge. Consider for example Stage 1. If agent 1 lies, then agent 2 can challenge him with the truth, and at Stage 1.3 agent 1 will choose $\{y; t_y + \Delta t\}$. For large enough Δt , this must be worse (for agent 1) than whatever the choice function f offers—recall our assumption (*). Equally, for large enough Δt , agent 2 will be satisfied with his reward of Δt (he receives a transfer of $t_y + \Delta t$). On the other hand, if agent 1 tells the truth, then agent 2 will not (falsely) challenge, since agent 1 would now choose $\{x; t_x + \Delta t\}$ at Stage 1.3, which incurs a penalty of Δt for agent 2 (he receives a transfer of $t_x - \Delta t$).

It should be clear that this construct could be used for more than two agents. Indeed, there is an advantage to having more. With just two agents, if one of them ever falsely challenged the other, the planner nets $2\Delta t$. (Of course, this only happens out of equilibrium.) With more agents, the $2\Delta t$ could go to a third agent rather than to the planner. The mechanism is thus balanced in equilibrium for two agents (but notice that out of equilibrium the planner never has to pay out); and it is always balanced for three or more agents. (“Balanced” here means that the net aggregate transfer provides the funding for the public decision. There is no presupposition that the net aggregate transfer is zero.)

Moreover if there are more than two agents it is enough that, for each agent i , only one other agent knows his type θ_i —since this other agent can be used to test him at Stage $i.2$. We consider this to be a strong and surprising result: any choice function with transferable utility can be implemented when each agent’s preferences are known to at least one other person.

We can go further. Unlike in much of the literature which delicately exploits the assumption of transferable utility, we do not actually need it in our stage mechanism. Roughly speaking, all that is required is that one agent can be adequately rewarded by sufficiently penalizing another agent. In effect, if the range of the choice function is contained in a “payoff box,” then more extreme allocations, outside this box, must be feasible. (Cf. Conditions C^+ and C^{++} .) This is a plausible enough requirement.

6. APPLICATIONS

In none of the examples below is the choice function (or correspondence) monotonic. Nash implementation is therefore ruled out. However, it is possible to implement it in subgame perfect equilibria, thanks to Theorems 2 or 3. Conditions C^+ or C^{++} apply in each case.

6.1. Risk Sharing

Two agents participate in a project with uncertain returns:

Agent 1 receives $\tilde{r}_1 = r_1(\tilde{\theta})$, and

Agent 2 receives $\tilde{r}_2 = r_2(\tilde{\theta})$

—where $\tilde{\theta}$ is a random variable with known distribution on a support Θ , and $r = (r_1, r_2)$ is an invertible function with bounded range. The realization θ of $\tilde{\theta}$ is only known to these two agents, so a contract cannot be usefully conditioned on θ directly. We assume that agent $i = 1, 2$ has a von Neumann–Morgenstern utility function u_i , say, defined over the relevant domain of “final wealth” e .³⁰ Their

³⁰ If agent i ’s utility were defined over (e^i, ∞) , where $e^i > -\infty$, then we would have to ensure that all terminal outcomes of g were feasible across states $\theta \in \Theta$. We investigated such a model in Examples 3 and 4 of Moore and Repullo (1986). (The specific utility function we used was $\log(1 + e)$, with $e > -1$.) Example 3 was of independent interest: we were able to implement a first-best risk-sharing contract as the unique subgame perfect equilibrium of a simple three-stage mechanism in which the agents take turns to move.

preferences exhibit decreasing absolute risk aversion (DARA): $-u_i''(e)/u_i'(e)$ is strictly decreasing in e .

A first-best risk-sharing contract will have the agents' marginal utilities in constant proportion across $\tilde{\theta}$. Let us suppose that some particular first-best contract requires that, if $\tilde{\theta} = \theta$, agent 2 pays agent 1 a transfer of $f(\theta)$ (which in general will take positive and negative values). $f(\theta)$ is an "outcome" in A .

$f(\theta)$ is a nonmonotonic choice function. If it were monotonic, it would have to be independent of θ , even if the set of outcomes (A) were broadened to include lotteries over the transfers. To see why, suppose, for example, that the "initial wealth" r_1 (i.e., prior to the transfer) of agent 1 decreases from $r_1(\theta)$ to $r_1(\phi)$. Then his lower contour set at the deterministic transfer $f(\theta) = \bar{p}$ (say) expands because of DARA: given any (stochastic) $\tilde{p} \neq \bar{p}$,

$$\begin{aligned} Eu_1(r_1(\theta) + \tilde{p}) &\leq u_1(r_1(\theta) + \bar{p}) \quad \text{implies} \\ Eu_1(r_1(\phi) + \tilde{p}) &< u_1(r_1(\phi) + \bar{p}). \end{aligned}$$

A similar argument holds for agent 2.

So $f(\theta)$ cannot be implemented in Nash. This is hardly surprising, since there are just two agents, and their preferences are diametrically opposed: it would seem a tall order to agree (uniquely) upon a transfer which depends on θ . Nevertheless, $f(\theta)$ can be implemented in subgame perfect equilibria.

We appeal to Theorem 3. Since r is invertible, for any θ, ϕ for which $f(\theta) \neq f(\phi)$, the initial wealth of at least one of the agents—agent 1, say—changes: $r_1(\theta) \neq r_1(\phi)$. DARA implies a change in his risk preferences over stochastic transfers. That is, there exists a test pair (\tilde{x}, \tilde{y}) exhibiting preference reversal:

$$\begin{aligned} Eu_1(r_1(\theta) + \tilde{x}) &> Eu_1(r_1(\theta) + \tilde{y}) \quad \text{and} \\ Eu_1(r_1(\phi) + \tilde{y}) &> Eu_1(r_1(\phi) + \tilde{x}). \end{aligned}$$

Given that the range of f must be bounded (the range of r is bounded), it is straightforward to confirm that the rest of Condition C^+ is satisfied. In particular, a compact "box" B of stochastic transfers can be found that contains the range of f and all the selected distributions (\tilde{x}, \tilde{y}) . Finally, a bad outcome \tilde{z} can be defined as a stochastic transfer whose support is wide enough to make \tilde{z} less attractive to either agent than any of the transfers in B . Condition C^{++} is therefore also satisfied, and Theorem 3 applies: this first-best risk-sharing contract f can be implemented in subgame perfect equilibria.

6.2. Production Contract (Principal-Agent)

Two firms that are not vertically integrated sign a production contract over the quantity to be traded ($q \geq 0$) and the transfer payment (p). This may be viewed as a Principal-Agent model: the downstream firm is the Principal, and the upstream firm is the Agent. The cost of production for the upstream firm is uncertain ex-ante, and is indexed by $\theta = (\alpha, \beta)$; α has support $[0, \bar{\alpha}]$ where $\bar{\alpha} < 1$, and β has support $[\underline{\beta}, \infty)$ where $\underline{\beta} > 0$. We assume that although both firms learn

θ ex-post, this is not observable to outsiders. The contract therefore cannot be conditioned on θ directly.

The parties' von Neumann–Morgenstern utilities are given by

Downstream firm (Principal): $u(q - p)$, with $u' > 0$ and $u'' \leq 0$;

Upstream firm (Agent): $v(p - \alpha q - \beta q^2/2)$, with $v' > 0$ and $v'' < 0$.

Notice that the Agent is risk averse—and the Principal may be too. For simplicity, let us assume that u and v are defined over the entire range of profit, $(-\infty, \infty)$, so that we avoid any feasibility problems.

A first-best contract (denoted by $*$) will have efficient trading

$$q = q^*(\alpha, \beta) = (1 - \alpha)/\beta,$$

and efficient risk-sharing

$$p = p^*(\alpha, \beta), \text{ say,}$$

where the exact form of p^* depends on u and v .

Now the states $\theta = (\alpha, \beta)$ can be partitioned into equivalence classes as follows. For each $\xi \geq \underline{\beta}$, define the equivalence class Θ_ξ :

$$\Theta_\xi \equiv \{ \theta = (\alpha, \beta) \mid 0 \leq \alpha \leq \bar{\alpha} \text{ and } \beta = \xi(1 - \alpha)/(1 - \bar{\alpha}) \}.$$

Within some class Θ_ξ , the efficient quantity $q^*(\alpha, \beta)$ equals $(1 - \bar{\alpha})/\xi$, which is a constant, q_ξ^* , say; yet it can be easily confirmed that the efficient transfer $p^*(\alpha, \beta)$ varies. Moreover, the Principal's preferences are independent of (α, β) , and the Agent's lower contour sets at (q_ξ^*, p) —for any p —are nested for $\theta = (\alpha, \beta) \in \Theta_\xi$.

To see this, note that for any $(\alpha, \beta) \in \Theta_\xi$, the slope of the Agent's indifference curve at some point (q, p) in the quantity/transfer space is given by

$$\alpha + \beta q = \alpha + \frac{\xi q(1 - \alpha)}{(1 - \bar{\alpha})} = 1 + \frac{\xi(q - q_\xi^*)(1 - \alpha)}{(1 - \bar{\alpha})}.$$

That is, as α increases (and β decreases commensurately to keep (α, β) in Θ_ξ), the slope of the indifference curve at any point to the right (resp. left) of (q_ξ^*, p) becomes shallower (resp. steeper). [Of course at (q_ξ^*, p) , the slope equals unity, as required for efficient production.]

Therefore the optimal contract $f(\theta) = (q^*(\alpha, \beta), p^*(\alpha, \beta))$ is nonmonotonic, and not Nash implementable. But by Theorem 3, f can be implemented in subgame perfect equilibria, since Condition C^{++} holds. In particular, the bad outcome z can be defined as a pair (q_z, p_z) , say, where q_z is large enough (i.e., excessive overproduction) to make both parties worse off at z than anywhere in the compact "box" B .

6.3. Walras Correspondence

Consider an I -person exchange economy in which preferences are continuous, semi-strictly convex and monotonic, and where endowments are strictly positive.

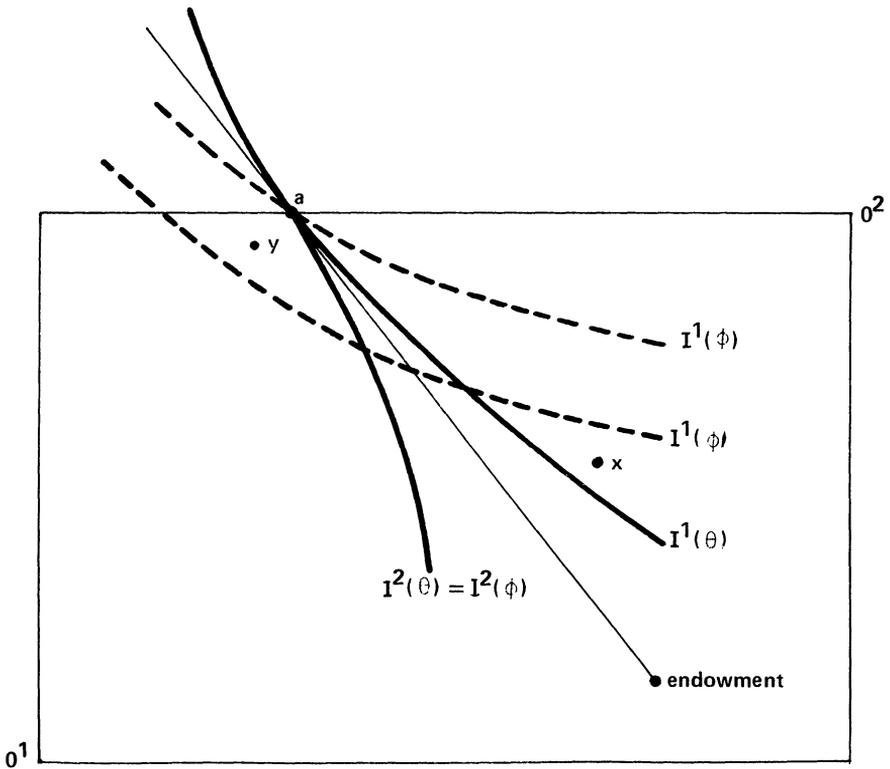


FIGURE 3

Let f_W be the correspondence that associates to each preference profile θ in Θ the set $f_W(\theta)$ of Walrasian equilibrium allocations corresponding to that profile.

It is well-known that the correspondence f_W is nonmonotonic, and so it cannot be implemented in Nash; e.g., Postlewaite (1985). To see this, consider the profiles θ and ϕ depicted in Figure 3. (For the purpose of illustration, we take $I = 2$.) Clearly, allocation a is a Walrasian equilibrium under the profile θ , and the agents' θ -lower contour sets at allocation a are contained in their respective ϕ -lower contour sets. (We are using the fact that no allocation outside the Edgeworth box is feasible.) However, allocation a is not a Walrasian equilibrium under ϕ , which contradicts monotonicity.

Nevertheless, under the above assumptions, an exchange economy is obviously an "economic environment" (as defined in Section 2) par excellence. So Condition C^+ holds. (In particular, the sequence $\sigma(\theta, \phi; a)$ pertaining to Figure 3 would be $\{a, x, y\}$.) From Theorem 2 we conclude that if $I \geq 3$, the Walras correspondence f_W can be implemented in subgame perfect equilibria.³¹ Finally, it is clear that if the set A of feasible allocations also contains the zero allocation (i.e., if free disposal is allowed), then by Theorem 3, f_W could also be implemented when $I = 2$.

³¹ For another stage mechanism that implements the Walras correspondence (though not with our informational assumptions), see the recent paper by Gale (1986).

7. FINAL REMARKS

The power of stage mechanisms (together with the requirement of subgame perfection) immediately suggests that certain other closely-related refinements of Nash equilibrium ought to yield strong implementation results too. In particular, one could stick to single-stage mechanisms and confine attention to those equilibria that are trembling-hand perfect—or even to those equilibria meeting the less stringent requirement that no agent's strategy is weakly dominated. Palfrey and Srivastava (1986) have now pursued this alternative route. Their remarkable finding is that for three or more agents essentially *any* choice correspondence satisfying no veto power is implementable in undominated Nash equilibria. Thus, the strength of their results is that they can implement choice correspondences which do not satisfy our necessary Condition C.

In light of the fact that our sufficient Condition C^+ is satisfied in what we have termed “economic environments”, to us the most persuasive examples of choice correspondences that they are able to implement (and we are not) arise in the context of voting when there is no divisible private good.

Of course, the key question is to what extent one can be confident that agents will *never* play weakly dominated Nash equilibrium strategies—even when, for example, the resultant outcome would be Pareto superior to that of some other (undominated) Nash equilibrium. In contrast, one could argue that the merit of the stage mechanisms introduced in Sections 1 and 5 above is that the way they will be played is more clear-cut: since agents move in turns it is rather compelling to believe that they will in fact play the (unique) subgame perfect equilibrium strategies. We think that this clarity is an inherent potential advantage of stage mechanisms, which gives them independent interest—albeit that we now know from Palfrey and Srivastava's work that undominated Nash implementation covers a wider class of choice correspondences.³²

On the other hand, an advantage to looking at single-stage mechanisms is that in environments with incomplete information, it is much easier to analyze the equilibria. (There is no need to worry about off-the-equilibrium-path beliefs, for example.) In a yet more recent paper, Palfrey and Srivastava (1987) have found that in private values models with three or more agents, it is possible to implement a very wide class of choice functions in undominated Bayesian equilibria.

London School of Economics, London WC2A 2AE, England
and
Bank of Spain, 28014 Madrid, Spain

Manuscript received May, 1986; final revision received October, 1987.

³² But we must again stress that although in many examples it may be possible to avoid simultaneous moving, this is not a feature of the general mechanisms used to prove our Theorems 2 and 3.

REFERENCES

- ABREU, D., AND A. SEN (1987): "Subgame Perfect Implementation: A Necessary and Almost Sufficient Condition", mimeo, Harvard University.
- CRAWFORD, V. P. (1977): "A Game of Fair Division," *Review of Economic Studies*, 44, 235-247.
- (1979): "A Procedure for Generating Pareto-Efficient Egalitarian-Equivalent Allocations," *Econometrica*, 47, 49-60.
- DASGUPTA, P., P. HAMMOND, AND E. MASKIN (1979): "The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility," *Review of Economic Studies*, 46, 185-216.
- DEMANGE, G. (1984): "Implementing Efficient Egalitarian-Equivalent Allocations," *Econometrica*, 52, 1167-1177.
- DEMSKI, J. S., AND D. SAPPINGTON (1984): "Optimal Incentive Contracts with Multiple Agents," *Journal of Economic Theory*, 33, 152-171.
- DUTTA, B. (1980): "Further Results on Voting by Veto," *Proceedings of the Conference on Collective Choice* (Caen, September 1980), ed. by P. Pattanaik and M. Salles. Amsterdam: North Holland.
- FARQUHARSON, R. (1969): *The Theory of Voting*. New Haven, Conn.: Yale University Press.
- GALE, D. (1986): "Bargaining and Competition Part I: Characterization," *Econometrica*, 54, 785-806.
- HART, O., AND J. MOORE (1988): "Incomplete Contracts and Renegotiation," *Econometrica*, 56, 755-785.
- KREPS, D. M., AND R. WILSON (1982): "Sequential Equilibria," *Econometrica*, 50, 863-894.
- MASKIN, E. (1977): "Nash Equilibrium and Welfare Optimality," mimeo, M.I.T.
- (1985): "The Theory of Implementation in Nash Equilibrium: A Survey," in *Social Goals and Social Organization*, ed. by L. Hurwicz, D. Schmeidler, and H. Sonnenschein. Cambridge: C.U.P., pp. 173-204.
- MOORE, J., AND R. REPULLO (1986): "Subgame Perfect Implementation," STICERD Theoretical Economics Discussion Paper 86/134, London School of Economics.
- (1988): "Nash Implementation: A Full Characterization," STICERD Theoretical Economics Discussion Paper 88/175, London School of Economics.
- MOULIN, H. (1979): "Dominance-Solvable Voting Schemes," *Econometrica*, 47, 1337-1351.
- (1980): "Implementing Efficient, Anonymous and Neutral Social Choice Functions," *Journal of Mathematical Economics*, 7, 249-269.
- (1983): *The Strategy of Social Choice*. Amsterdam: North Holland.
- (1984): "Implementing the Kalai-Smorodinsky Bargaining Solution," *Journal of Economic Theory*, 33, 32-45.
- MUELLER, D. (1978): "Voting by Veto," *Journal of Public Economics*, 10, 57-75.
- PALFREY, T., AND S. SRIVASTAVA (1986): "Nash Implementation Using Undominated Strategies," mimeo, Carnegie-Mellon University.
- (1987): "Mechanism Design with Incomplete Information: A Solution to the Implementation Problem," mimeo, Carnegie-Mellon University.
- POSTLEWAITE, A. (1985): "Implementation via Nash Equilibria in Economic Environments," in *Social Goals and Social Organization*, ed. by L. Hurwicz, D. Schmeidler, and H. Sonnenschein. Cambridge: C.U.P., pp. 205-228.
- REPULLO, R. (1986): "The Revelation Principle under Complete and Incomplete Information" in *Economic Organizations as Games*, ed. by K. Binmore and P. Dasgupta. Oxford: Basil Blackwell, pp. 179-195.
- ROBERTS, K. (1979): "The Characterization of Implementable Choice Rules" in *Aggregation and Revelation of Preferences*, ed. by J.-J. Laffont. Amsterdam: North Holland, pp. 321-348.
- SELTEN, R. (1975): "Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 4, 25-55.