DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

Implementation in Bayesian Equilibrium: The Multiple Equilibrium Problem in Mechanism Design

Thomas R. Palfrey

SOCIAL SCIENCE WORKING PAPER 760

December 1990

IMPLEMENTATION IN BAYESIAN EQUILIBRIUM:

THE MULTIPLE EQUILIBRIUM PROBLEM IN MECHANISM DESIGN

Thomas R. Palfrey
California Institute of Technology

ABSTRACT

This paper surveys the literature on implementation in Bayesian Equilibrium.

1. Introduction

Implementation theory links together social choice theory and game theory. At a less abstract level, its application provides an approach to welfare economics based on individual incentives. The underlying motivation for implementation theory is most easily seen from the point of view of a relatively uninformed planner who wishes to optimize a social welfare function which depends on environmental parameters about which relevent information is scattered around in the economy. Thus, the planner wishes to both collect as much of this relevent information as possible, and, with this information, make a social decision (e.g., an allocation of resources). This is the classic problem identified by Hurwicz [1972]. In the twenty years since, we find numerous research agendas falling into the general category of implementation problems: the study of planning procedures, contracts, optimal regulation and taxation, agency relationships, agendas and committee decision making processes, comparative electoral systems, noncooperative foundations of general equilibrium theory, and even much of the recent theoretical work in accounting and the economics of law.

The dilemma such a planner faces is that the individuals from which the information must be collected will not necessarily want to share their information, or worse, they may wish to misrepresent their information. Moreover, exactly how they choose to conceal and misrepresent their information (what we will call their *deception decision*) depends upon three things. First and foremost, it depends on expectations of how the planner intends to put to use the information that is being collected. Second, it depends upon their expectations about the deception decisions of the other agents. Third, it depends on their information, so it is convenient to think of a deception decision as a *plan* of what to reveal as a function of information.

The thoughtful planner, realizing this, takes account of the possibility that there may be deceptions when deciding how to translate the collected information into a social decision. But in order to do this the planner needs to have a basis for predicting how deception decisions vary as a function of the individuals' expectations about how he intends to translate the information into social decisions. Game theory provides a large family of such predictions, in the form of equilibrium concepts, and these each provide a mutually consistent theory of individuals' expectations about the deception decisions of the other individuals. In addition to a predictive theory of behavior, the planner also needs to consider how he can manipulate the individuals' expectations of how he will translate the information into social decisions. The latter problem has, at least until very recently, been finessed in implementation theory by the *commitment assumption*:[1] the planner may commit to any feasible outcome function, which is a rule for translating collected information into social decisions. This assumption implies that the individuals' expectations exactly coincide with whatever outcome function the planner has announced he will use. Finally, the planner must have the ability to control the information collection process. To simplify this problem implementation theory usually imposes *the control assumption*: the planner may choose any message space and the individuals must communicate exactly one message from this message space, may not communicate any additional messages, and may not communicate with each other. The combination of a message space and and outcome function is called a *mechanism.*

Given an equilibrium concept and a class of environments, or domain, one can pose the incentive compatibility question: For what allocation rules does there exist a mechanism for which that allocation rule is an equilibrium outcome? If we can answer that question in the affirmative for some allocation rule, that allocation rule is incentive compatible. For many equilibrium concepts there is a well-known result called *the revelation principle*, which states that if an allocation is incentive compatible, then it can be produced as the equilibrium outcome to a particularly simple kind of mechanism, called a direct mechanism. In a direct mechanism the message spaces are simply the individuals' information sets and the outcome function is the allocation rule itself. If an allocation rule is incentive compatible then it is the "truthful" equilibrium outcome to its associated direct mechansim.

However, incentive compatibility is only half of the implementation problem. The other half is the *multiple equilibrium problem.* While an allocation rule may be incentive compatible, it may have the problem that any mechanism for which that rule is an equilibrium outcome also produces other allocation rules as equilibrium outcomes. An unfortunate implication of the multiple equilibrium problem is that the revelation principle may be of only limited usefulness. In particular, it is concievable (and many plausible examples have been constructed, see Section 2) that under some equilibrium concepts certain incentive compatible allocation rules are uniquely implementable, but the direct mechanism associated with that allocation rule is plagued by multiple equilibria. This suggests (correctly, as it turns out) that the guts of the mechansim design constructions to uniquely implement incentive compatible allocations are typically quite messy – in particular the individuals are asked to report messages that have additional components[2] beyond simply announcements of their private information. Nevertheless, despite this additional complexity, it turns out one can prove a simple characterization theorem establishing that the multiple equilibrium problem cannot be resolved unless a particular set of inequality conditions hold. These inequality conditions are referred to as *monotonicity conditions.* The problem of designing mechanisms to solve both the incentive compatibility problem and the multiple equilibrium problem is sometimes called *full implementation.* For brevity, I drop the word "full" and simply refer to it here as implementation.

To this point, the discussion has focussed on implementing a specific allocation rule. More generally, one can talk about implementing collections of allocation rules, or *social choice sets.* For example, one might wish to implement the set of competitive equilibria rather that one particular selection from the competitive equilibrium correspondence. The theoretical issues of implementing correspondences are in many ways the same as the issues that arise in implementing specific allocation rules (*social choice functions*), so, for simplicity we will mostly concentrate on allocation rules.[3]

The first equilibrium concept to be applied to the implementation problem was *dominant strategy equilibrium.* This equilibrium concept has two important features. First, it circumvents the difficult problem of how to model players' expectations of other players' deceptions. If something is a dominant strategy equilibium, then nobody cares how other individuals are behaving – or at least such expectations will not usually affect behavior. Second, and for more subtle reasons, it trivializes the second half of the implementation problem. Dominant strategy incentive compatibility very nearly implies that any multiple equilibria can be avoided.[4] This in turn implies that little generality is lost in restricting attention to direct mechanisms if dominant strategy equilibrium is the solution concept.

Of course the cost of these simplifications is known to be steep. Specifically, the Gibbard-Satterthwaite theorem tells us that if we consider sufficiently broad domains then practically nothing is implementable in dominant strategies. The reason is the difficulty of satifying incentive compatibility. In their terms, nontrivial *strategy-proof* allocation rules generally do not exist. Domain restrictions are needed before it becomes possible to implement interesting allocation rules.[5]

While it would have been convenient to have obtained more positive results with dominant strategy mechanisms, it is unfortunately not the case. The natural next step was to explore the implications of implementation in Nash equilibrium.[6] Using Nash equilibrium as the solution concept means that individual expectations about other individuals' deceptions are modelled explicitly and are important. In the Nash implementation approach, the assumption that players know nothing about their opponents' private information is replaced by the assumption that everyone's private information is common knowledge among the individuals.[7] For this reason, this line of research is appropriately viewed as "implementation with complete information."

The most promising applications of the complete information approach would probably be to contracting between two agents. In such situations, it may often be the case that the two contracting parties are relatively well informed about each others' preferences and beliefs, but the planner (or the courts that will be enforcing the contract) is poorly informed about their preferences and beliefs. In this case, the mechanism may be thought of as a state-contingent contract that the players agree to *before* the state of the world (including their preferences) is realized. Such an interpretation might also apply to relatively small groups such as committees and legislatures, who interact frequently enough that they become relatively well-informed about each others' preferences. On the other hand, this raises the problem that these static models are *least* applicable where agents interact frequently enough to know each others' preferences, since reputation building and supergame considerations become important.

Another important difference with Nash implementation as compared with dominant strategy implementation is that the multiple equilibrium problem rears its ugly head. While the problem with dominant strategy implementation was incentive compatibility (strategy-proofness) the problem with Nash implementation is multiple equilibria. This is not surprising, as it is relatively difficult to construct game forms which have dominant strategies across a range of preference profiles, but it is easy to construct games with Nash equilibria. However, it is more difficult to construct games with unique Nash equilibrium.

The incentive compatibility issue becomes trivial for many domains because the redundancy of the individuals' information enables the planner to use the direct mechanism equivalent of "forcing contracts," where a universally bad outcome is enforced unless all the reports of private information agree.[8] For almost any allocation rule, it is possible to use such constructions to make truthful reporting a Nash equilibrium in a game where everyone reports the economy-wide profile of private information. Unfortunately, in some of these constructions almost any profile of mutually consistent deceptions is also a Nash equilibrium. Thus the turn from Dominant Strategy implementation to Nash implementation turns the problem around completely so the focus is almost entirely on the multiple equilibrium problem rather than being almost entirely on the incentive compatibility problem.

Because of this multiple equilibrium problem, it is no longer sufficient to restrict attention to direct mechansims.[9] The constructions required to "weed out" undesirable or extraneous Nash equilibria involve adding nuisance components to individual messages besides simply requesting individuals to submit an announcement of their information. The nature of the message spaces differ from those used in dominant strategy implementation in two ways: (1) players report *profiles* of private information, (2) nuisance messages are used.

Finally, we reach the topic of this paper: Bayesian Nash implementation, or simply Bayesian implementation. This approach combines features of the dominant strategy approach and the Nash approach. Not only is the planner incompletely informed, but now individuals have truly private information – they know things that the other individuals do not know. If this private information is sufficiently exclusive[10] then there is no role at all for forcing contracts. However, unlike the dominant strategy approach some additional structure is imposed on the priors players have about other players' private information. The incentive compatibility conditions of strategy-proofness are replaced by "Bayesian incentive compatibility" conditions in which individuals condition on their priors and on specific expectations about the deceptions of the other individuals. Thus, like Nash equilibrium, optimal equilibrium behavior depends upon expectations about other individuals' behavior. For this reason, the potential difficulty of multiple equilbrium that arose in Nash equilibrium is still present.[11]

Thus, unlike either dominant strategy implementation or complete information Nash implementation, Bayesian implementation distinctly has two components to it – **incentive compatibility** and **multiple equilibrium**. As a result it should not be surprising that there are really two distinct literatures that have developed. One literature, easily the larger and more applied of the two, explores the implications of incentive compatibility. The other line of research, the more recent of the two, explores the implications of the multiple equilibrium problem, attempts to characterize exactly when it can be overcome, and provides some clues about the nature of the indirect mechanisms that need to be used to eliminate undesirable equilbria. The latter approach is also beginning to investigate implementation using refinements of Bayesian equilibrium, an apporach that has been applied in complete information Nash implementation as well. Refinements help mitigate the multiple equilibrium problem, and for some domains solve the problem entirely.

The remainder of this paper focuses on the less well-known and more recent line of work on the multiple equilibrium problem, and for the most part ignores the now vast literature on Bayesian incentive compatibility.[12] Section 2 offers three simple examples to illustrate the multiple equilibrium problem that can arise, and to illustrate how it can or cannot be avoided. Section 3 presents a simple and useful characterization theorem about unique implementation in Bayesian equilibrium, and gives a detailed proof the the result. Section 4 catalogs a variety of alternative characterizations, explains the role of different assumptions, and draws some comparisons with complete infomation Nash implementation. Section 5 investigates some extensions of the usual Bayesian mechanism design approach, where the assumptions of committment and control are relaxed somewhat, and where refinements of Bayesian equilibrium are used. Section 6 raises and discusses without resolution some issues that need to be confronted more carefully and systematically in the near future.

2. Some Examples

Example A

This is based on early example to illustrate the multiple equilibrium problem in Bayesian implementation that appeared in Postlewaite and Schmeidler [1986]. There are three agents and two equally likely states of the world, s, s'. Agent 1 is uninformed. Agents 2 and 3 are perfectly informed.[13] There are three feasible alternatives, $A = \{a, b, c\}$. State-contingent preferences are strict, with the rankings given by

P_1	P_1'	$P_2 = P_3$	$P_2' = P_3'$
b	a	a	b
c	c	b	a
a	b	c	c

Thus, player 1's preference between a and b is always the opposite of the informed players' preferences between a and b. Everyone has a von Neumann-Morgenstern utility function at each state which assigns a utility index of 1 to the first choice, 0 to the last choice, and $v > ½$ to the middle choice. The planner is player 1's twin brother and he wishes to implement x, given by $x(s) = b$, $x(s') = a$.

If a direct mechanism is used where truth is an equilibrium strategy, then only players 2 and 3 report messages, and the outcome function g must look like

g:

Agent 2's report \ Agent 3's report	s	s'
s	b	c
s'	c	a

Figure 1. Direct game

If the direct game were not as in Figure 1, then either truth would not be an equilibrium strategy or $g(s,s) \neq b$ or $g(s',s') \neq a$.

But this game also has a non-truthful equilibrium in which agents 2 and 3 always lie, as well as two pooling equilibria, where 2 and 3 either always reports s or always report s'. In fact, the lying equilibrium makes both informed agents better off and makes the planner (and agent 1) worse off.[14] Therefore, we see that some indirect reporting is necessary to avoid the bad equilibria.

The way to accomplish this is by giving player 1 some messages which change the outcome function $g(m_2, m_3)$. It has been observed by Mookherjee and Reichelstein [1990] that it may be possible to "selectively eliminate" the bad equilibria by giving player 1 exactly one message for each equilibrium (of course these new messages may add new equilibria, but that will be dealt with later). Matsushima [1990a] has observed that we may then label these new messages by the equilibrium strategies of the players in the direct game. In this example, it turns out that *all* of the non-truthful equilibria can be eliminated in the same way so that we only need to give agent 1 two strategies, which we call "truth" and "not truth." One may think of these strategies of player 1 as being an announcement of whether he knows the other players are using the good equilbrium (truth) or one of the bad equilibria (not truth). The outcome function is:

"truth"

Agent 3's Report \ Agent 2's report	s	s'
s	b	c
s'	c	a

"not truth"

Agent 3's Report \ Agent 2's report	s	s'
s	c	b
s'	a	c

Figure 2. Indirect mechanism to implement
$F(s) = b$, $F(s') = a$ in example.

It is easily verified that this indirect mechanism solves the implementation problem in this example. However, notice that there are still two equilibria. One corresponds to the truthful strategy of the direct game (with agent 1 saying "truth"). The other has agent 1 saying "not truth," agent 2 always lying, and agent 3 always telling the truth. Both equilibria produce x as the equilibrium outcome.

It is critical in the above example that $v > .5$, that is, it is more valuable to player 1 to receive c in both states than it is for player 1 to receive either a in both states or b in both states. If this is not the case, then it may be impossible to implement the desired social choice function because only one of the pooling equilibria can be selectively eliminated. This problem is vividly illustrated in example **C** below. However, as we will see in section 3, the existence of a transfer good (money) with no income effects will usually solve the problem even if $v < .5$.

Example B: Double auctions

In simple two-agent bargaining problems where the use of double auctions has been explored, a single seller owns an indivisible object which he values at v_s. A single buyer values it at v_b. These values are each private information to the respective party. An interesting special case is when it is common knowledge that v_s and v_b are independently drawn from a uniform distribution on $[0, 1]$. An allocation rule is a mapping which assigns a probability of a transfer $p(v_s, v_b)$ and a transfer payment $t(v_s, v_b)$ for any $(v_s, v_b) \in [0, 1]^2$. It is well known (Chatterjee and Samuelson [1983], Myerson and Satterthwaite [1983]) that of all incentive compatible allocation rules, the one that maximizes the expected sum of buyer surplus plus seller surplus is:

$$p(v_s, v_b) = \begin{cases} 1 \text{ if } v_s \le v_b - \frac{1}{4} \\ 0 \quad \text{otherwise} \end{cases}$$

$$y(v_s, v_b) = \begin{cases} \frac{1}{6} + \frac{1}{2}(v_b + v_s) \text{ if } v_s = v_b - \frac{1}{4} \\ 0 \quad \text{otherwise} \end{cases}$$

This is equivalant in (interim) expected utility to the allocation rule resulting from a linear bidding equilibrium of the split-the-difference double auction, where a buyer submits bid b and a seller submits an offer s and the buyer gets the object if and only if $b \ge s$. If trade occurs, a transfer of $\frac{b+s}{2}$ is made from the buyer to the seller. The equilibrium bidding functions are $B(v_b) = \min(v_b, \frac{2}{3}v_b + \frac{1}{12})$ $S(v_s) = \max(v_s, \frac{2}{3}v_s + \frac{1}{4})$.

Unfortunately, the double auction with these rules, or in fact *any* direct mechanism yielding an allocation rule which is interim equilivant to this one is plagued by multiple equilibria. Satterthwaite and Williams [1989], and Leinger, Linhart and Radner [1989] show that there is a continuum of multiple equilibria.

In turns out that the efficiency properties of this allocation rule allow us to implement it, essentially uniquely. All of the extraneous, inefficient equilbria can be selectively eliminated. Details will be given in section 5.

Example C: Public goods

Consider the following simple public goods problem without side payments. There are two feasible public good decisions, $A = \{a, b\}$. There are 3 agents, each of whom either strictly prefers a to b (type t_a) or strictly prefers b to a (type t_b). Each knows his own type. It is common knowledge that each player's type was an independent draw with prob $(t_a) = q$, prob $(t_b) = 1 - q$. An allocation rule is a mapping from the set of eight possible type profiles to a probability p of deciding "a".

An optimal (by many criteria) allocation rule x, selects a with probability 1 if at least two players are type t_a and selects b with probability 1 if at least two players are type t_b. This is the majority rule solution, it maximizes ex ante welfare, it is anonymous, incentive compatible, and all sorts of other nice things. *It is even fully implementable in dominant strategies!!* However, for some values of q, x is not Bayesian implementable. The reason is that if q is large, then the allocation rule "always a" is a Bayesian equilbrium outcome in any mechanism for which x is a Bayesian equilibrium outcome. If q is small, then "always b" is an equilbrium outcome in any mechanism where x is an equilbrium outcome. Therefore, x is implementable only if q is sufficiently close to $\frac{1}{2}$ – i.e. only if players have poor information about each other.

It is instructive to see exactly why this is true. Suppose we have a mechanism for which x is an equilibrium outcome. Then it must be that equilibrium strategies exist which produce outcomes as in Figure 3. (Player 3 is the row player. Player 2 is the column player, and Player 1 is the matrix player.)

$\sigma_1(t_a)$:

Player 3 \ Player 2	$\sigma_2(t_a)$	$\sigma_2(t_b)$
$\sigma_3(t_a)$	a	a
$\sigma_3(t_b)$	a	b

$\sigma_1(t_b)$:

Player 3 \ Player 2	$\sigma_2(t_a)$	$\sigma_2(t_b)$
$\sigma_3(t_a)$	a	b
$\sigma_3(t_b)$	b	b

Figure 3. Equilibrum resulting in x,
with players using σ.

Consider the strategy profile $\hat{\sigma}$, where $\hat{\sigma}_i(\cdot) = \sigma_i(t_a)$ is a constant strategy for all i. That is everyone uses σ, but acts as if their types were always t_a. This produces the outcomes given in Figure 4.

$\sigma_1(t_a)$:

Player 3 \ Player 2	$\sigma_2(t_a)$	$\sigma_2(t_b)$
$\sigma_3(t_a)$	a	a
$\sigma_3(t_b)$	a	a

$\sigma_1(t_b)$:

Player 3 \ Player 2	$\sigma_2(t_a)$	$\sigma_2(t_b)$
$\sigma_3(t_a)$	a	a
$\sigma_3(t_b)$	a	a

Figure 4. Equilibrum with players using
$\hat{\sigma}$, yielding "always a"

We now can show that if $q > \sqrt{.5}$ then $\hat{\sigma}$ will be an equilibrium. Suppose not. Then some player, say player 1, must have a message available to him, m, so that $g(\hat{\sigma}_{-1}, m) = b$. But this means that $g(\sigma_{-1}, m)$ produces the allocation rule in figure 5.

		Player 2	
		$\sigma_2(t_a)$	$\sigma_2(t_b)$
Player 3	$\sigma_3(t_a)$	b	?
	$\sigma_3(t_b)$	?	?

Figure 5. Outcome if players 2 and 3 use
σ and player 1 uses m.

Now observe that no matter what the "*?*" entries are in Figure 5, player 1 is better of using m instead of $\sigma_1(t_b)$ when he is type t_b, if everyone else uses σ and $q > \sqrt{.5}$. Therefore σ is not an equilibrium, a contradition. If $1 - q > \sqrt{.5}$, a similar argument shows that the constant strategy $\hat{\sigma}(\cdot) = \sigma(t_b)$ is a Bayesian equilibrium.

3. A Simple and Useful Characterization Theorem

As of the writing of this survey, there are quite a few different characterizations of social choice correspondences that are implementable by Bayesian equilibrium (Postlewaite and Schmeidler [1986], Palfrey and Srivastava [1989a, 1989b], Jackson [1991], Mookherjee and Reichelstein [1990], Matsushima [1990a, 1990b], and others). These characterizations differ from each other in their assumptions about the number of individuals, the existence of money transfers, the finiteness of the space of individual private information, the amount of conflict between individual preferences, the amount of redundancy in the individuals' private information, and a number of other specifics. The next section will discuss differences between the various characterizations in more detail. In this section, a very simple one is presented in detail. This characterization is new. It is suggested by example A of the previous section in which the presence of an uninformed agent is exploited to "break" the mutual deceptions of the informed agents.

Consider an environment with a structure of private information of the following sort. There is a *set of feasible alternatives*, A, and a continuously transferable good that everyone values (call it money), that is in fixed supply, W. Let

$D = \{(w_1, \ldots, w_N) \in I\!R^n \mid w_1 + \cdots + w_N = W\}$, be the set of feasible transfers, so arbitrarily negative transfer payments are feasible. There are $N \geq 2$ individuals, and each individual i may be one of T_i possible types, where T_i is finite. An allocation rule, $x(t) = (a(t), w(t))$, is a mapping from $T = T_1 \times \cdots \times T_N$ into $A \times D$. The set of feasible allocation rules is X. Each individual i has preferences over A that depend on T, and these preferences are represented by a Von Neumann-Morgenstern utility index $V_i(a, t)$ for $a \in A$ and $t \in T$. Given an *allocation* (a, w, t), individual i's utility is $U_i(a, w, t) = V_i(a, t) + w_i$.

The structure of private information is that there is a common prior over profiles of types, denoted $q(t)$ and we assume $q(t) > 0$ for all $t \in T$. This latter assumption is called *diffuse information* (Palfrey and Srivastava [1989a, 1989b]). Individuals follow Bayes rule, so that individual i, conditional on having observed t_i, updates his prior over T_{-i} by $q(t_{-i} \mid t_i) = q(t) / \sum_{\tau_{-i} \in T_{-i}} q(\tau_{-i}, t_i)$.

A *mechanism* consists of a Cartesian product of individual *message spaces*, $M = M_1 \times \cdots \times M_N$ and an *outcome function* $g : M \rightarrow A \times D$. A *strategy*[15] for player i, σ_i, is a function from T_i to M_i. Given a strategy profile $\sigma = (\sigma_1, \ldots, \sigma_N)$, denote $\sigma_{-i} = (\sigma_1, \ldots, \sigma_{i-1}, \sigma_{i+1}, \ldots, \sigma_N)$. The strategy σ is a *Bayesian equilibrium* to the mechanism (M, g) if, for all i, for all $t_i \in T_i$, and for all $m_i \in M_i$

$$\sum_{t_{-i} \in T_{-i}} q(t_{-i} \mid t_i) U_i(g(\sigma(t)), t) \geq \sum_{t_{-i} \in T_{-i}} q(t_{-i} \mid t_i) N_i(g(\sigma_{-i}(t_{-i}), m_i), t).$$

Given an environment, (A, q, T), an allocation rule x is *implementable* in Baysian equilibrum if there exists a mechnism (M, g) with the following properties:

(1) There exists a Bayesian equilibrium to (M, g).

(2) If σ is an equilbrium to (M, g) then $g(\sigma(t)) = x(t)$ for all t.

Summarizing, we are maintaining the following assumptions throughout this section.

Assumption 1: $N \geq 2$

Assumption 2: $|T| < \infty$

Assumption 3: Unlimited balanced transfers without income effects

Assumption 4: Common prior on T

Assumption 5: Diffuse information

Finally, we make the following assumption.

Assumption 6 (An Uninformed Player): $|T_1| = 1$

This assumption, or anything resembling it, cannot be found in any general implementation theorem of which I am aware. Nonetheless, it underlies a number of examples in the literature. The reason that uninformed individuals make the task of eliminating unwanted equilibria so much easier is quite subtle, but very powerful. First, note that the presence of this individual has no effect on the set of incentive compatible allocations. Second, the way a planner eliminates unwanted equilibria is by offering rewards to individuals for acting as stool pigeons and reporting undesirable strategizing by themselves and others to the planner. Of course the problem faced by the planner is how to design incentives for such behavior without encouraging stool pigeons to falsely report that deceptions are being employed. Thus it is quite difficult to design such selective incentives if the planner does not know an individual's preferences. However, if an individual does not have any private information, the planner knows exactly how to design the stool pigeon rewards for that individual.

In order to understand the statement of the theorem, we need to introduce a bit more notation. Suppose we have a direct mechanism, so that $M_i = T_i$ for all i and the outcome function $g: M \rightarrow A \times D$ is therefore simply an allocation rule x. (Henceforth, we will write a direct mechanism (T, x).) Then a strategy for player i is simply any mapping from T_i into itself. We call the identity mapping from T_i to T_i the *truthful strategy* for i, denoted I_i and generally call any strategy for i in the direct mechanism a *deception by i*, denoted α_i. A profile of deceptions by the individuals other than the uninformed individual is called a *joint deception* and is written $\alpha=(\alpha_1, \ldots, \alpha_N)$, and the profile of deceptions by everyone except for i is written α_{-i}. Any joint deception induces a probability distribution over the reported types of the informed individuals, denoted q_α, and also induces a new allocation rule, x_α defined by $x_\alpha(t) = x(\alpha(t))$ for all $t \in T$. Formally q_α is defined by

$$q_\alpha(t) = \sum_{\tau \in \alpha^{-1}(t)} q(t)$$

$$\text{where } \alpha^{-1}(t) = \{\tau \in T \mid \alpha(\tau) = t\}.$$

Following the terminology of Matsushima [1990a], α is called a *consistent deception* if $q_\alpha = q$. That is, a joint deception is consistent if it generates the same distribution of reports by the informed individuals as if they had all used their truthful strategies.

Assumption 7 (No Consistent Deceptions): $q_\alpha \neq q \;\forall\; \alpha \neq I$

It is well known that a necessary condition for Bayesian implementation is that x is implementable only if it is *incentive compatible*. An allocation rule x is incentive compatible if the truthful strategy profile is a Bayesian equilibrium of the direct mechanism (T, x). That is:

Definition 1: An allocation rule x is *incentive compatible* if for all i and for all $t_i, \tau_i \in T_i$

$$\sum_{t_{-i} \in T_{-i}} q(t_{-i} \mid t_i) U_i(x(t), t) \geq \sum_{t_{-i} \in T_{-i}} q(t_{-i} \mid t_i) U_i(x(t_{-i}, \tau_i), t)$$

Theorem 1. Under assumptions 1-7, x is incentive compatible if and only if x is implementable.

Proof: "If" follows from standard results by Harris and Townsend [1981], Myerson [1979] and others. "Only if" is more complicated. As in all characterization proofs in implementation theory, a general mechanism is constructed. In this case the message spaces can be simply given by:

$$M_i = T_i \qquad i = 2, \ldots N$$

$$M_1 = \{\hat{w} : T \rightarrow I\!R^N : \textstyle\sum_{t\in T}\hat{w}_1(t)q(t) < 0 \text{ and } \sum_{i=1}^N \hat{w}_i(t) = 0 \;\forall\, t\} \cup \mathbf{0}$$

That is, all individuals with private information simply send reports of that private information. The uninformed individual requests a supplemental balanced transfer as a *function* of the reported types of the other individuals. This supplemental transfer rule must be either the **0** function or must have the property that the uninformed individual's expected transfer under the *true* type distribution is negative. The outcome function if agents $2, \ldots, N$ report t' and agent reports $\hat{w}(\cdot)$ is:

$$g(m) = (a(t'), w(t') + \hat{w}(t'))$$

We claim that this mechanism uniquely implements x. That is, there is a unique Bayesian Nash equilibrium which produces x as the equilibrium allocation rule. Why does this mechanism work? First observe that since x is incentive compatible, it is a mutual best response for all the privately informed individuals to report truthfully if the uninformed agent reports **0**. It is strictly optimal for the uninformed individual to request **0**, given everyone else tells the truth. Hence x is an equilibrium outcome, and is the *only* equilibrium with truthful reporting by the informed individuals.

Can there exist an equilibrium with the informed individuals using a joint deception α? No. If they use the deception α, then the probability distribution of reports is q_α. Since $q_\alpha \neq q$, there exists t such that $q_\alpha(t) > q(t)$. Therefore there exists w such that $Kw\cdot q < 0$ and $Kw\cdot q_\alpha > 0$ for all $K > 0$. Therefore, **0** is not a best response. By choosing K large, we can make $Kw\cdot q_\alpha$ arbitrarly large. But since $Kw\cdot q_\alpha$ can be made arbitrarly large he has no best response when the informed individuals use deception α. Therefore there is no equilibrium with the informed individuals using deception α. QED

This construction has several notable features. First, and most important, is the optional use of supplementary side payments by the uninformed agent. The planner (via a message space restriction) specifically disallows any supplemental request by the uninformed agent that would make him better off were the informed individuals adopting truthful reporting strategies. Since x is incentive compatible, this implies that x is still an equilibrium outcome in the expanded indirect mechanism (and the uninformed player sends **0**). On the other hand, suppose α were an equilibrium in the direct mechansm (T, x). Then α is no longer an equilibrium in the expanded mechanism because $q_\alpha \neq q$ implies that there exists a proposal $\hat{w} \in M_1$ where the uninformed individual prefers $(a_\alpha, w_\alpha + \hat{w}_\alpha)$ to $x_\alpha = (a_\alpha, w_\alpha)$. In the parlance of Mookherjee and

Reichelstein [1990], x_α is "selectively eliminated" from being an equilibrium outcome by allowing the uninformed agent to propose $\hat{w}$. Finally, we have to make sure that the mechanism has been carefully constructed to avoid the addition of any new equilibrium which involves $\hat{w} \neq 0$. No such equilibrium can arise because for any proposal $\hat{w} \neq 0$ the uninformed agent might want to make, the proposal $2\hat{w}$ makes him even better off.

Thus we see that there are three goals to accomplish in the construction. One may make a loose, if somewhat morbid, analogy with the medical practice of surgically removing diseased tissue. Goal number one is to remove the unwanted tissue (bad equilibrium). Goal number two is to *not* remove the healthy tissue (good equilibrium). Goal number three is to make this selective removal sanitary, without creating a "secondary infection" (creating *new* unwanted equilibrium). All constructions of which I am aware have this structure, and the proofs consist of establishing that the constructed mechanism accomplishes these three goals.

In general, the solution is not as simple as I have described, either because there usually does not exist a completely uninformed agent, or because there exist consistent deceptions (as for example would be the case with a continuum of types), or because side payments are infeasible or restricted. These lead to complications in several ways: by necessitating different individuals to play the role of breaking different undesirable equilibrium, depending on which equilibrium is to be excised; by needing to allow proposed allocations that differ in the "public decision" besides simply supplementing the money transfers; by restricting proposals to be ones that make *all possible types of the proposer* worse off if everyone is adopting truthful strategies; and by having some components of joint messages "incompatible"[16] with each other. These and other complications are dealt with in Section 4. We next turn from sufficient conditions to necessary conditions.

Necessary Condition: Monotonicity

It is fairly well-known now in the literature on implementation via Nash equilibrium with complete information that a social choice function must satisfy a monotonitity condition in order to be implementable. This condition, the importance of which relative to implementation theory was first recognized by Eric Maskin [1977], roughly states that if some alternative is the social choice for one profile of preferences, R, and we consider another profile in the domain, R', where that alternative does not go down in anyone's preference ranking relative to any other alternative, then it must also be the social choice for the new profile. This property was originally viewed as a desirable normative criterion without consideration of its implications for equilibrium of noncooperative games.[17]

Formally, we have the following definition (for correspondences):[18]

Definition 2: A *social choice correspondence* F is *monotonic* if, for all R, R'

If: (a) $x \in F(R)$

(b) $\forall\ i, y \quad xR_i y \rightarrow xR_i' y$

Then $x \in F(R')$

What is the connection to noncooperative game theory that Maskin noticed? It is most easily seen by considering a slightly rearranged statement of the definition.

Definition 3: A social choice correspondence F is *monotonic* if for all R, R'

If: (a) $x \in F(R)$

(b) $x \notin F(R')$

$\exists\, i, y$ such that $xR_i y$ and $yP_i' x$.

Henceforth we will call i a "test agent" and y a "test allocation," or (x, y) a "test pair."[19] Now it is almost trivial to see why monotonicity is neccessary for Nash implementation. Suppose F is Nash implementable and let $x \in F(R)$ and x nomem $F(R')$. Then there exists a mechanism in which $g(m) = x$ for some $m \in M$ and m is an equilibrium at R. But m is not an equilbrium at R', so there must be some $i, \hat{m}_i, y$ such that $y = g(m_{-i}, \hat{m}_i) P_i' g(m) = x$. But, since m is an equilibrium at R, $x\ R_i y$. Therefore, we see that there exists a test agent i and a test allocation $y = g(m_{-i}, \hat{m}_i)$ which makes x *not* an equilibrium at R'. This gives:

Theorem 2: If F is Nash implementable then F is monotonic.

Turning to incomplete information, intuition suggests that, since Bayesian Nash equilibrium is simply a more general statement of Nash equilibrium there should exist an appropriate restatement of monotonicity that applies to Bayesian implementation. The first such extension was identified by Postlewaite and Schmeidler [1986], and it captures exactly the notion of a test agent and a test allocation. The critical difference is that the relevant notion of an *outcome* of a Bayesian game is now a function mapping type profiles into feasable allocations whereas one traditionally thinks of the equilibrium outcome of a complete information game as a single alternative. While one can of course build up such functions by concatenating all of the complete information outcomes for different profiles, this is not how monotonicity is usually stated with complete infomation.

This reinterpretation of outcomes of games as allocation rules means that "test allocations" are substituted with "test allocation rules." Since we look at allocation *rules*, we can no longer state monotonicity by simply comparing equilibrium outcomes at two specific type profiles (as we compared R and R' in the original definition of monotonity), but must compare two strategy profiles, or alternatively, two deception. Thus, instead of comparing all pairs (R, R') we consider all **functions** $\alpha = (\alpha^1, \ldots, \alpha^N)$ with $\alpha^i : T^i \rightarrow T^i$. For a given α and a given t, the pair $(t, \alpha(t))$ is analogous to the pair (R, R') in the original definition of monotonicity. Similarly, the preferences, R_i, are replaced by interim preferences $\tilde{R}_i(t_i)$, where $x\ \tilde{R}_i(t_i) y$ iff

$\sum_{t_{-i}\in T_{-i}} q(t_{-i} \mid t_i)U_i(x(t),t) \geq \sum_{t_{-i}\in T_i} q(t_{-i} \mid t_i)U_i(y(t),t)$. A social choice correspondence is then represented as a collection of social choice functions, and so is often called a *social choice set.* With this in mind we have (assuming diffuse information):

Definition 4: F is *Bayesian Monotonic* if $\forall\ \alpha = (\alpha^1, \ldots, \alpha^N)$ with $\alpha^i : T^i \to T^i$, $\forall\ i$:

If: (1) $x : T \to A \in F$

(2) $\forall\ i, s_i, y : T^{-i} \to A \quad x\tilde{R}_i(t_i)y \quad \forall\ t_i \in T_i$

$$\Longrightarrow x_\alpha \tilde{R}_i(s_i) y_\alpha$$

Then $x_\alpha \in F$

Alternatively, we may write, instead of (2)

(2′) $\forall\ i, s_i, y : T \to A \quad x\tilde{R}_i(t_i)y(\cdot \mid \alpha(s_i)) \quad \forall\ t_i \in T^i$

$$\Longrightarrow x_\alpha \tilde{R}_i(s_i) y_\alpha$$

where $x\tilde{R}^i(t_i)y(\cdot \mid \alpha(s_i))$ means

$$\sum_{t^{-i}\in T^{-i}} q_i(t_{-i} \mid t_i)\Big[u_i(x(t),t) - u_i(y(t_{-i},\alpha_i(s_i)),t)\Big] \geq 0$$

As was the case with the original monotonicity condition this can be rewritten in terms of the existence of a test agent i and a test allocation rule y.

Definition 5: F is *Bayesian monotonic* if for all α and for all $x \in F$, if $x_\alpha \notin F$ then $\exists$ $i, s_i \in T^i, y : T \to A$ such that

$x\tilde{R}^i(t_i)y(\cdot \mid \alpha(s_i)) \quad \forall\ t_i \in T^i$

but $y_\alpha \tilde{P}^i(s_i)x_\alpha$.

To prove that Bayesian monotonicity is necessary for implementation in Bayesian Nash equilibrium involves essentially the same argument used in Theorem 2. Suppose $x \in F$, $x_\alpha \notin F$, and F is Bayesian implementable. Then since $x \in F$, there exists a set of equilibrium strategies which produce x, call it σ. The joint strategy σ_α defined by $\sigma^i_{\alpha_i}(t_i) = \sigma^i(\alpha_i(t_i))$ for all i and t_i yields outcome x_α. Since this is not an equilibrium, some player i has some type s_i and some alternative message m_i such that $g(\sigma^{-i}_{\alpha_{-i}}(t_{-i}), m_i)\tilde{P}_i(s_i)x_\alpha$. Let y be the allocation defined by g when everyone else uses σ^{-i} and i always sends the message m_i, so $y = g(\sigma^{-i}(t_{-i}), m_i)$. Then $x\tilde{R}^i(t_i)y\ \forall\ t_i$ since σ is an equilbrium, so y is a test allocation rule that satifies the required inequality conditions. Observe that by the way y was constructed (with i using a "constant" strategy m^i, this means that we may write $y_{-i}: T_{-i} \rightarrow A$ instead of $y : T \rightarrow A$ and this generates the definition of Bayesian monotoniciy given in (2). Summarizing:

Theorem 3: If F is Bayesian Nash implementable then F is Bayesian monotonic.

Returning to Theorem 1 briefly, we can see exactly why Bayesian monotonicity is satisfied under Assumptions 1-7. Since we assumed that every deception α generates a probability of reported types different from q (the "no consistent deceptions" assumption), this guarantees that for any α, the uninformed agent (agent 1) is one such test agent, and there will always exist a test allocation rule $y = \hat{w}$ simply involving a type contingent transfer scheme ($y : T^{-1} \rightarrow A$) that staisfies $x\tilde{R}^1 y$ and $y_\alpha \tilde{P}^1 x_\alpha$. Notice that since 1 is uninformed – he has only one type – the dependence of R^1 on t_1 is suppressed.

Also observe that all players utility functions are such that their preferences for lotteries over transfers do not depend on their type. Thus, if types are *independently* distributed, so we may write $q(t) = q_1(t_1) \cdots q_N(t_N)$ for any $t \in T$, a similar result holds, using a nearly identical construction.[20]

4. Other Characterizations

The characterization in section 2 is special in a number of ways. The characterizations that have appeared in the literature are usually stated so that they apply to a much broader class of environments than ones that satisfy Assumptions 1-7.

This section explores a few of these alternative characterizations. First, we consider pure exchange environments where income effects and limited transfers do not permit the same kind of construction as the one used in the last section. Second we consider generalizations of pure exchange environments called "economic" environments. The basic ingredient in such environments is that there are always differences in preference between some individuals in the population. In particular, too much agreement on a most-preferred alternative is ruled out. Third, we explore the implications of relaxing the assumption of diffuse information. This provides results for the "grey area" between complete and incomplete information.

A. Pure exchange environments

A feasible allocation of resources in a private goods economy with L commodities and N agents is a collection of individual allocations $a = (a_1, \ldots, a_N) \in I\!R_+^{NL}$ such that $\sum_{i=1}^{N} a_{il} \leq \omega_l$ for all l, where ω_l is the aggregate endowment of commodity l. Observe that exact balancing of resources is not imposed, although many of the results below can also be proved when exact balancing is required.[21] We denote by A the set of all such feasible allocations. There are a set of states of the economy, S. Given a state $s \in S$, we denote individual i's utility of a_i is state s by $U^i(a_i, s)$ and assume it is strictly increasing in a_{il} for all l, and $s \in S$. The state also captures information players may have about each other. To maintain consistency with the notation introduced in section 3, there are a finite set of types for each i, denoted T_i, so the set of states of the economy are the set of all vectors of types.[22] Let $X = \{x : T \rightarrow A\}$.

The definitions of incentive compatibility and Bayesian Monotonicity are the same for pure exchange economies as for the environments discussed in the previous section. However, because (a) we are not guaranteed the existence of an uninformed agent, (b) there are limitations on transfers and (c) there are income effects, it becomes very helpful to require at least three agents. While this is not a necessary condition for implementation, it simplifies the problem considerably, as will be apparent below.

For the case where information is diffuse, the set of Bayesian implementable allocations is characterized as a special case of a result in Palfrey and Srivastava [1989b].[23] One version of that result, adapted to our current notation and assumptions is:

Theorem 4: If T is finite, information is diffuse and A consists of pure exchange allocations then x is Bayesian implementable if and only if it is incentive compatible and Bayesian monotonic.

Proof: "Only if" follows from Theorem 2. "If" requires construction of a general mechanism. The following one works. For each i, let $M_i = X \times T^i \times \{0, 1, \ldots\}$, so each individual announces a social choice function (allocation rule), a type, and a non-negative integer, n_i. Let t' be the profile of reported types. If everyone reports x, and at least $n-1$ agents report 0, then the outcome is $x(t')$. If everyone except i reports x and 0 but agent i reports $y \neq x$, and any $n_i \in \{0, 1, \ldots\}$, then the outcome is:

$$x(t') \quad \text{if} \quad y\,\tilde{P}^i(t_i)x \quad \text{some } t_i \in T^i$$
$$y(t') \quad \text{if} \quad x\,\tilde{R}^i(t_i)y \quad \forall\, t_i \in T^i$$

For any other profile of reports the outcome is determined by an "integer game." The individual who announces the highest[24] $n_i \in \{0, 1, \ldots\}$ receives ω and everyone else receives nothing.

This mechanism produces x as the unique equilbrium outcome. Clearly it is an equilibrium for everyone to send the message $(x, t_i, 0)$ when type $t_i \in T$, since x is incentive compatible. There can be no equilibrium with some i announcing $y \neq x$, or $n_i \neq 0$ at some $t_i \in T_i$. First, note that because of the integer game no such equilibrium could arise with i announcing a y which gives *all* resources to some $j \neq i$. Therefore any $j \neq i$ is better off announcing an integer higher than any integers announced by anyone at any type (recall T is finite!).

Finally, if there were an equilibrium with individuals deceving via α, so that each j reported $(x, \alpha_j(t_j), 0)$ at t_j, then it must be that $x_\alpha = x$. Otherwise, since x is Bayesian monotonic there would be an agent i and a type $s_i \in T^i$ who could report y such that $x\tilde{R}^i(t_i)\,y(\cdot \mid \alpha(s_i)) \;\forall\, t_i \in T$ but $y_\alpha \tilde{P}^i(s_i)\,x_\alpha$, so that individual would be better off reporting y than reporting x when he is type $s^i \in T^i$. ■

Matsushima [1990a] has produced a proof that uses an alternative construction where undesirable (deceptive) equilibria are broken by a designated test agent $i(\hat{\alpha})$ who may report the joint deception $\hat{\alpha}$ being used, and which results in a pre-defined test allocation $y(\hat{\alpha})$. A related idea underlies the construction by Mookherjee and Reichelstein [1990]. The point is that one may either allow agents to announce alternative allocations y and allow y if it passes the test in the Bayesian monotonicity definition, or the planner can simply define a function that maps reported joint deceptions into the appropriate test allocations.

B. "Economic" environments

Many economies do not fit exactly into the pure exchange setting described above, such as environments with public goods, or production and consumption externalities. Nevertheless, in many of these problems of interest, there is always some minimal degree of conflicting preferences between the individuals in the environment.

For a broad class of these environments, virtually identical results to Theorem 3 can be proved. The basic insight, originally due to Maskin's [1977] complete information results, is that implementation problems may arise if a social choice correspondence fails to select nearly unanimous "best alternatives." For complete information environments, we say an alternative is *nearly unanimous at state* s if at least $N - 1$ agents rank it at the top of their preference ordering at s. A social choice correspondence satisfies *No Veto Power* (NVP) if $x \in F(s)$ whenever x is nearly unanimous at s. NVP is a very convenient sufficient condition for proving implementability of a monotonic SCC when $N \geq 3$.

In pure exchange environments with $N \geq 3$, there is nothing even close to a nearly unanimous outcome, and consequently no auxilliary assumption is required. To generalize this, Jackson [1991] identified what he calls *economic environments*. This requires that ". . . for any given social choice function and state, there are at least two agents who prefer to alter the social choice function at that state." (p. 9) In other words, no outcome is nearly unanimous at any state.

For the case of diffuse information, we have, formally:

Definition 6: The environment is *economic* (E) if, for any $a \in A$ and $t \in T$, there exist $i \neq j$, and $b_1, b_2 \in A$ such that $b_1 P_i(t_i) a$ and $b_2 P_j(t_j) a$.

Assumption 8: The enviornment is economic.

This assumption is also made implicitly in Matsushima [1990b p. 8] who assumes that every type of every individual has a *most preferred alternative* that does not depend on the other players' types. This assumption is automatically guaranteed in pure exchange enviornments and also plays a role in the earlier work by Palfrey and Srivastava [1989a, 1991a] on implementation using undominated Nash equilibrium. Formally, this assumption is:

Assumption 9 (Matsushima [1990b]): (Existence of known best element). The environment is economic and for every i and every $t_i \in T_i$ there exists $b_i(t_i) \in A$ such that $U_i(b_i(t_i), t) \geq U_i(a, t)$ for all $t_{-i} \in T_{-i}$, and for all $a \in A$.

Notice that Assumption 9 combines two features. First, it requires that everyone have a most preferred alternative that doesn't depend on the other players' private information. Second, the implementation problem must be "economic." These two components jointly require that $N \geq 3$.

Theorem 5: If the implementation problem is economic, $N \geq 3$, $T < \infty$, and information is diffuse, then x is Bayesian implementable if and only if x is incentive compatible and Bayesian monotonic.

Remarks about Theorem 5: Essentially the same proof as the one used to prove Theorem 4 works here. Jackson [1991] proves a more general version of Theorem 5 for correspondences in which he allows agents to have non-exclusive information. His proof also uses a different integer game, which he calls a matching game. It has the attractive feature that if A is finite then the message space is finite. Matsushima [1990b] proves a version of Theorem 5 using Assumption 9.

C. Non-Exclusive Information

The initial work on Bayesian implementation by Postlewaite and Schmeidler [1986, 1987] and Palfrey and Srivastava [1987] assumed non-exclusive information (NEI).[25] If states are represented by profiles of types, and players share a common prior q, then:

Definition 7: There is *non-exclusive information* (NEI), if, for all i and for all t_{-i} there exists $\hat{t}_i(t_{-i})$ such that, for all $t_i \neq \hat{t}_i(t_{-i})$, $q(t_i, t_{-i}) = 0$.

Thus non-exclusive information changes the assumption that $q(t) > 0$ for all $t \in T$, in a very special way. In particular, it implies that player i's type can be precisely inferred from knowledge of all the other players' types. That is, $q(\hat{t}(t_{-i}) \mid t_{-i}) = 1$ for all i and for all t_{-i}. The reason for the initial focus on these information structures was simple: incentive constraints are never binding in pure exchange economies. That is, for *any* allocation rule, x, one can always[26] design a forcing contract so that x is an equilibrium outcome. Therefore, essentially the only hurdle for implementation is the multiple equilibrium problem. In this way, there is a close resemblance to Nash implementation with complete information.

A useful insight gained from the case of non-exclusive information is that it encompases complete information as a special case.[27] This means that Bayesian implementation is exactly the natural generalization of Nash implementation that one hoped it would be. What is perhaps a little surprising and unfortunate about NEI is that the simplicity gained by circumventing incentive compatibility is lost in the complexity of stating a more complicated Bayesian monotonicity condition, needing a more complicated constructive proof, and requiring an additional condition, called *closure.*

Closure is a property of a social choice correspondence. Any new allocation rule that is an appropriate mixture of some collection of allocation rules that are in the social choice set must also lie in the social choice set. These mixtures involve splicing[28] together a portion of one allocation rule for all type profiles in one common knowledge event of T with a portion of another allocation rule for all remaining type profiles in T. Therefore, closure only has bite if (a) the common knowledge partition of T consists of at least two nonempty events and (b) the social choice set F contains at least two allocation rules.[29]

Closure is most easily illustrated by an example with complete information and two type profiles, so $T = \{t, t'\}$. Suppose that $x, y \in F$ and $x(t) = a$, $x(t') = a'$, $y(t) = b$, $y(t') = b'$ and a, a', b, b' are all distinct elements of A. Then closure says $\hat{x}$ and $\hat{y}$ must also be in F, where $\hat{x}(t) = a$, $\hat{x}(t') = b'$, $\hat{y}(t) = b$, $\hat{y}(t') = a'$. Notice that closure implies, among other things, that with complete information a social choice set may be obtained as the cross product of the images of a social choice correspondence. That is, with complete information, if F is a social choice set satisfying closure, then there exists a social choice correspondence $\hat{F} : T \rightarrow A$ such that $F = \times_{t \in T} \hat{F}(t)$.

In pure exchange environments, several results are known for the case of non-exclusive information. First, Theorem 4 extends in a natural way, with the exception that incentive compatibility is not required and, accordingly, joint deceptions that are incompatible[30] can safely be ignored. Thus monotonicity can be stated by

Definition 8 (NEI monotonicity): x is *Bayesian Monotonic* if, for all compatible α such that $x_\alpha \neq x$, there exists i, t_i such that $x\, R^i(\alpha_i(t_i))\, y$ and $y_\alpha P^i(t_i)\, x_\alpha$.

A second result is that if we restrct attention to pure exchange economies where preferences are strictly concave and satisy a condition guaranteeing interior Walrasian equilibrium, then the Rational Expectations equilibrium correspondence is Bayesian monotonic and therefore implementable when information is non-exclusive.[31] Wettstein [1986] shows that if the interiority condition is dropped, we may still implement the "constrained" Rational Expectations equilibrium, where demands are exogenously bounded by the aggregate endowment. Third, allocations which are interim envy-free are implementable with non-exclusive information. These and other applications of Bayesian monotonicity to pure exchange environments with non-exclusive information are found in Palfrey and Srivastava [1987].

Jackson [1991] extends these results with either exclusive or non-exclusive information in general economic environments (Assumption (8)) and also provides a very general characterization for finite-type environments without the "economic" condition. He identifies a modification to Bayesian monotonicity which incorporates an interim version of the No Veto Power condition used to prove sufficiency in Nash implementation. That combined condition is called *Monotonicity-No-Veto* (MNV). In fact MNV reduces to the separate conditions of monotonicity and NVP when information is complete. It is therefore not a necessary condition for implementation.

5. Extensions

In this section we will discuss extensions in three different directions of the Bayesian implementation results summarized in the earlier section:

(A) Refinements of Bayesian equilibrium

(B) Relaxing the committment assumption (renegotiation and individual rationality)

(C) Relaxing the control assumption (message space constraints and preplay communication between the agents).

A. Refinements of Bayesian equilibrium

The most obvious strengthening of Bayesian equilibrium, and the one that has received by far the greatest attention in mechanism design theory is Dominant Strategy Equilibrium. Unfortunately, as noted above, incentive compatibility in dominant strategies is generally very difficult to achieve, and when it can be achieved the multiple equilibrium problem does not arise.

Palfrey and Srivastava [1989a] investigate the equilibrium refinement of *Undominated Bayesian Equilibrium* (UBE), which is a Bayesian Nash Equilibrium in which no player uses a weakly dominated strategy. Thus it combines the best response property of Bayesian Nash equilibrium with a dominance property. This refinement turns out to have a lot of clout.

The strength of this refinement is best seen by returning to the third example of Section 2. In that example, we were trying to implement majority preferences over two alternatives. However, it turns out that for most priors over types, that social choice function is not Bayesian monotonic. Neverthless, we observed that it was implementable in dominated strategies. How do we resolve this apparent paradox? It would seem that if x is implementable in dominant strategies then x should also be implementable in Bayesian equilibrium, but that is unfortunately not true.

The resolution to this apparent paradox is simple: *all the extra equilibria that undermine the implementation of the majoritarian social choice rule involve the use of weakly dominated strategies.* The question then is: What other social choice functions are not Bayesian monotonic but can be implemented in UBE? The answer is that if individual preferences do not depend on other players' types ("private values") then virtually any incentive compatible social choice function is UBE-implementable.

The intuition behind this result borrows heavily from results about implementation with complete information using the undominated refinement of Nash equilibrium (Palfrey and Srivastava [1991a]). There it was shown that if preferences were *value distingushed*, then with at least three agents any social choice function or correspondence satisfying no veto power is implementable in Undominated Nash Equilibrium.

Preferences are *strictly value distinguished* if, for each i, $t_i \neq t_i'$, there exists a test pair of outcomes, (a, b), such that $aP_i(t_i)b$ and $bP_i(t_i')a$. Weak value distinction allows for some weak preferences in the definition of the test pair. Thus value distriction is a condition that says two types are different if and only if their preferences are different.[32]

With incomplete information, it has been proved that if $N \geq 3$, preferences are value distinguished and private, no one is ever completely indifferent over all outcomes, and best and worst elements exist for every type of every player, then a social choice function is incentive compatible if and only if it is implementable in UBE. If preferences are not value distinguished or depend on the entire profile of types, then a necessary condition must be satisfied that is weaker then Bayesian monotonicity, but is quite strong nonetheless.

The proof for "private values" is long but straightforward, at least for the case of strict preferences. The idea is to have players announce their type twice (plus some nuisance messages). In equilibrium, only the first announcement, t_i, matters. However, the mechanism is constructed so that the second announcement, t_i', may (out of equilibrium) trigger the use of a test pair $\{a(t_i, t_i'), b(t_i, t_i')\}$ with the property that $a(t_i, t_i')\ P^i(t_i)\ b(t_i, t_i')\ P^i(t_i')\ a(t_i, t_i')$. If player i announces the pair (t_i, t_i'), for which his corresponding test pair is (a, b), then when the test pair is triggered, the outcome is b, which is preferred to a in state t_i', but not in state t_i. The rest of the mechanism is set up so that, at state t_i', the report (t_i, t_i) is dominated by the report of (t_i, t_i'), which will effectively prevent any joint deception from being an equilibrium. To guarantee that there are no other equilibria, say with i reporting (t_i, t_i'), we introduce a new kind of integer game called "tailchasing." In addition to announcing (t_i, t_i'), each player also announces an integer. Whenever a player reports two *different* types then reporting the number M is weakly dominated by reporting any number larger than M. Details may be found in Palfrey and Srivastava [1989a].

The restriction to private values and value-distinguished-types is quite important. For one thing it rules out some applications which might be important, including bargaining and auctions with common or affiliated values. The reason these assumptions produce such strong results is that in order for a simple construction with test pairs, the choice of the test pair cannot depend on other players' types. And, if types were not value-distinguished (e.g., two types differ only in information about other players) then test pairs do not exist.

The following example illustrates how common values can create implementation problems even when using the UBE refinement. This example is a "common-value majoritarian" social choice function. There are three players and $A = \{a, b\}$. Everyone has identical strict preferences, but different information. Each player can be one of two types, $T_i = \{t_a, t_b\}$. Types are independently drawn, with $q = \text{prob}\{t_a\} > \sqrt{.5}$. Each individual prefers a to b if and only if there are at least two type t_a players, otherwise everyone prefers b to a. The social choice function, x, we would like to implement is the obvious one: select a if everyone prefers a to b; select b if everyone prefers b to a.

Clearly x is incentive compatible, so that in a direct game everyone is better off honestly reporting his type than lying as long as everyone else is honestly reporting his type. However, one can show that there is also a pooling Bayesian equilibrium where everyone always claims to be type t_a, so a is always selected. This additional equilibrium outcome arises in any game for which x is an equilibrium outcome. Furthermore, one can never "refine" this bad outcome away by eliminating weakly dominated strategies and using tailchasing or other elaborate construction techniques. Therefore, x is not implementable in UBE.

The above example suggests that in some seemingly innocuous cases, implementation may require resorting to stronger refinements such as those found in Kohlberg and Mertens [1986], or those based on sequential rationality such as sequential equilibrium or Perfect Bayesian equilibrium. Little is known about implementability using these refinements in games of incomplete information. The work of Moore and Repullo [1988] and Abreu and Sen [1989] for subgame perfect equilibrium with complete information indicates that sequential rationality refinements do expand the set of implementable social choice functions, but to a lesser degree than ruling out weakly dominated strategies. However, some of the "signaling" refinements of sequential equilibrium combine dominance-based refinements with sequential rationality, and so may permit the implementation of even more social choice functions.

Abreu and Matsushima [1990] have taken a slightly different approach, examining implementation in iteratively undominated strategies. They obtain very strong results with finite types and conflicting preferences. In order to obtain these strong results, they relax the implementability definition. According to this relaxed definition, a social choice function is *virtually implementable* if there exists an "exactly" implementable (possibly random) social choice function that approximates it arbitrarily closely. This follows the earlier work of Matsushima [1988] and Abreu and Sen [1989b]. In addition to incentive compatibility, they identify a very weak necessary condition for implementation called *measurability.* The details of their constructive proof are substaintally different from the ones presented earlier in this survey. An interesting discussion (and criticism) of the construction appears in a subsequent note by Glazer and Rosenthal [1990].

B. The committment problem

In the standard mechanism design problem, an *uninformed planner* or perhaps a collection of *uninformed players* commit to a mechanism at the ex ante stage. In particular, they commit to an outcome function and a message space. These completely determine the rules according to which they exchange information and decide on outcomes.

A large can of worms is opened when one attempts to relax the assumption that there are no constraints, other than feasibility, on the outcome function. Considerable attention has been focussed on the implications of impossing various kinds of restrictions on the kinds of outcome functions that may be commited to. For convenience, we divide the "limited committment" approaches into four broad categories:

(B1) Constraints on ex post allocations

(B2) Durability

(B3) Dynamic Contracting

(B4) Information Leakage

B1. Constraints on ex post allocations

The two most common restrictions of the first type are *individual rationality* and *ex post efficiency.* In the Bayesian implementation literature, the analysis of Ma, Moore and Turnbull [1988] carefully addresses issues of individual rationality. They construct a unique implementation of the Demski-Sappington optimal contract with two agents and two types. The individual rationality problem arises there because either agent may choose "non participation" at the message sending stage. One may view the individual rationality constraint as reinterpreting the outcome function $g(m)$ as an "offer function," where the offer may be refused by any of the players, resulting in a status quo payoff G_0^i, for some or all of the players. This is equivalant to adding an extra stage at the end of the mechanism in which each agent chooses between G_0^i and $g(m)$. In general, this may lead to feasibility problems, although not in the Ma, Moore, and Turnbull [1988] setting.

Individual rationality is frequently imposed in applications of the revelation principle to contracting problems. However, with the exception of the work by Ma, Moore, and Turnbull and realted work by Rajan [1989], most of the work on the multiple equilibrium problems in Bayesian mechanism design has paid little attention to the problem of individual rationality.

The issue of ex post efficiency has been addresed in the complete information literature by Maskin and Moore [1989], in "renegotiation-proof" implementaton.[33] Their work is motivated by the observation that in many of the constructive proofs, multiple equilibria are selectively eliminated by allowing the planner to "enforce" relatively arbitrary outcomes. For example, the test allocations that are used to break unwanted equilibria may be very inefficient. They explore the implications of requiring the outcome function to always be efficient.

In order to impose this requirement, one has to assume that there is some sort of ex post bargaining process that follows the play of a mechanism and that *always* produces ex post efficient allocations. This implies that we may write a function $E(g(m), s)$ which maps each realization of the outcome function to another outcome[34] if the true state is s. This E–function, or bargaining technology,[35] is then taken as a transformation on the final outcomes that will arise from any mechanism (M, g). It is important to observe that E incorporates information that is not in the planner's information set – in particular, E depends on s. This implies, among other things, that any efficient social choice function, f, is implementable with renegotiation relative to some E, namely $E(\cdot, s) = f(s)$. Thus it is not at all clear that we should view "renegotiation-proofness" of this sort as a constraint, since it may (almost by assumption) make some non-monotonic social choice functions implementable by embedding their implementability in the *technology.*

This approach is clearly intended to capture, in reduced form, the observation that the reporting of messages followed by a (temporary) outcome generated by the planner's outcome function is not the end of the game. Either the planner cannot reasonably commit to producing an outcome that he knows is inefficient, or the planner cannot prevent the players themselves from negotiating to a Pareto improvement. In either case, the implicit assumption is that there is a continuation game that follows the mechanism, and this continuation game always ends up selecting an efficient allocation. This raises some interesting new issues: What are examples of such continuation games that always lead to efficiency? What renegotiation functions $E(\cdot, s)$, can arise as unique equilibrium outcomes of such continuation games? The recent work by Rubenstein and Wolinsky [1989] offers one approach to this in a bargaining setting where a specific renegotiation process is proposed. Aghion, Dewatripont and Rey [1989] also suggest an approach in which the renegotiation functions themselves are part of the designed mechanisms.

B2. Durability

The problem of durability arises at the interim stage when one agent could propose to replace the mechanism with a new mechanism and all the other agents would agree to do so. There are many ways of formulating this problem,[36] which is very closely related to the problem of mechanism design by an informal principal.[37]

The fact that renegotiation takes place before any "official" messages have been sent distinguishes this class of variations on imperfect committment from the others. As the problem was originally formulated in Holmstrom and Myerson [1983], the multiple equilibrium problem was not addressed. However, as Legros [1990] points out, the possibility of multiple equilibria suggests a reformulation, which he calls "strong durability."

Strong durability is a property of a mechanism-allocation rule *pair*, rather than being a property of allocation rules, as in Holmstrom and Myerson. The idea is that such a pair is strongly durable if the allocation rule is implemented by the mechanism and there does not exist mechanism that could be unanimously approved in a vote against the original mechanism that implements the allocation rule. An allocation rule is then called strongly durable[38] if there exists a mechanism such that the mechanism-allocation rule is strongly durable. Notice that an allocation may fail to be durable relative to its direct mechanism, but an indirect mechanism may exist that makes it strongly durable.

One can then show a close connection between strong durability and interim efficiency. For example, in independent private value models, an allocation rule is strongly durable if and only if it is interim efficient! This contrasts with the results of Holmstrom–Myerson, where the leading example of that paper demonstrates that interim efficient allocation rules may not be durable. The reason for the difference is that Holmstrom–Myerson restrict attention to direct mechanisms. One can generalize the result of Legros [1990] in an interesting way. He (and Holmstrom and Myerson [1983]) only considers *unanimous* voting procedures. In fact, one can reformulate much of the model to allow for arbitrary (finite) renegotiation processes, as extensive form games that are played prior to the mechanisim, the continution games of which are mechanisms. As long as side-payments are possible, then one can show that in independent private values models any interim efficient allocation rule is strongly durable with this more general definition.[39]

B3. Renegotiation in Dynamic Environments and (B4) Information Leakage

In many settings, it is optimal for the same mechanism to be applied over and over again in a multiperiod contracting problem (Stokey [1979]). However if player types are correlated across periods then *all* parties may wish to change the mechanism as time passes and the information structure evolves.[40]

At this point there has been very little work on the multiple equilbrium problem with renegotiation. The main reason for this is that sequential rationality is imposed in these models. The difficult issues of Bayesian implementation with sequential rationality are open questions. While characterization results for sequential rationality with complete information are highly developed,[41] there are no analogous results in Bayesian settings. Since dynamic contracting problems seem to be one major direction the mechanism design approach is headed, this is an unfortunate gap in the current state of Bayesian implementation theory.

For similar reasons, little is known about the problem of agents making inferences from the messages of other agents, and thus wishing to revise their message. This problem arises in the posterior implementation work of Green and Laffont [1987a]. Chakravorti [1989] explores some of the multiple equilibrium problems with implementation, using a stronger posterior requirement.

C. The control assumption

The standard approach to mechanism design assumes that the planner has complete control over the communication technology, and complete control over the outcome function. Issues concerning renegotiation involve some limitations on the planner's ability to credibly commit to an outcome function, but other issues may arise quite independently of the commitment abilities of the planner. One possibility, collusion by preplay communication, is illustrated by the following example.

EXAMPLE: Collusion by preplay communication.

There are two agents, who work for a principal. The principal is trying to decide between two different projects to undertake. The agents have information about different aspects of the projects, say its cost and some other dimension that the agents care about a lot, but the principal cares about very little, such as whether it requires much effort on the part of the agents. For simplicity, agent 1 can be type t_1 or type t_2 and agent 2 can be type s_1 and s_2. If agent 1 is t_1 then the principal wants to undertake project A and if agent 1 is t_2 then the principal wants to undertake project B. Agent 2 only has information about the second dimension of the project that the principal doesn't care about. The agents' information is independent, and the joint distribution of types is common knowledge, with $prob(t_1) = .5$ and $prob(s_1) = .4$. The agents have idential state contingent utility functions given in Figure 6.

	$s_1 t_1$	$s_1 t_2$	$s_2 t_1$	$s_2 t_2$
A	0	1	1	0
B	1	0	0	1

Figure 6

Thus we see that the agents would prefer project A if the type profile were (t_1, s_1) or (t_2, s_2) and prefer project B otherwise. One can easily see that the planners social preferred allocation rule, given in Figure 7, is incentive compatable and is uniquely implemented in Bayesian equilibrium.

	s_1	s_2
t_1	*A*	*A*
t_2	*B*	*B*

Figure 7

However, the direct mechanism that implements it can be undermined if the agents are able to communicate prior to playing the mechanism. That is, there exists an extensive form game which consists of a "cheaptalk" stage followed by the play of the direct mechanism, which contains an equilibrium outcome different from the one intended by the planner.

Consider the following very simple form of preplay communication. Agent 2 tells agent 1 whether he is s_1 or s_2. Then the agents play the mechanism. One equilibrium of this two stage game is for agent 1 to ignore what agent 2 tells him and agent 2 always says he is type s_1, and then players act truthfully in the direct mechanisms. This produces the planners choice. However, there is another equilibrium in which agent 2 tells agent 1 his true type, and agent 1 reports truthfully to the planner if and only if agent 2 tells him s_2. Thus, we see that some mechanisms may uniquely implement social choice functions but are not immune to preplay communication.

A number of recent papers have suggested that possibilities of cheaptalk may undermine the ability of a planner to implement some allocation rules, because of the multiplicity of equilibria that often arise in direct games with preplay communication.[42] This suggests a role for indirect mechanisms, and an application of the techniques developed for the existence/characterization theorems of Section 4. Palfrey and Srivastava [1991b] investigate the use of indirect mechanisms for "cheaptalk-proof" implementations of interim efficient allocation rules in economic environments.

Given any mechanism (M, g), define a class of extensive form games, called the set of (finite) *communication extensions of* (M, g).[43] This consists of all games which begin with a finite number of communication stages where players transmit messages to each other, followed by a final stage in which the mechanism (M, g) is played out. Of course, there may be many new equilibrium outcomes in a communication extension of (M, g) because information may be transmitted between the agents, as in the example above.

Let $X(M, g)$ be the set of all Bayesian equilibrium allocation rules of the mechanism (M, g). Let $X^c(M, g)$ be the set of all Bayesian equilibrium allocation rules of all communication extensions of (M, g). We say that (M, g) is *cheaptalk-proof* if $X(M, g) = X^c(M, g)$. Unfortunately, it is hard to prove general theorems about cheaptalk-proof mechanism design, because it may be that $X(M, g)$ and $X^c(M, g)$ only differ in ways that are inconsequential in utility terms.

For this reason, we introduce the concept of *essential implementation*, as a slightly weaker implementation requirement.

Definition 9: Let x be a social choice function. Then (M, g) *essentially implements* x if x is an equilibrium outcome of (M, g) and all other equilibrium allocation rules of (M, g) produce the same (interim) utility allocations to all types of all players.

We can define *essentially cheaptalk-proof* in a similar way. Let $U(M, g)$ be the set of interim utility allocations associated with all allocation rules in $X(M, g)$, and let $U^c(M, g)$ be the set of interim utility allocations associated with all allocation rules in $X^c(M, g)$.

Definition 10: (M, g) is *essentially cheaptalk-proof* if $U(M, g) = U^c(M, g)$.

We may then show that there is a close connection between interim efficiency of an allocation and the possibility for it to be essentially implemented using an essentially cheaptalk-proof mechanism. Specifically, we restrict attention to enviornments of the following sort:

1. Sidepayments.
2. Private values: $U(x(t), t) = U(x(t), t_i)$ for all i, t, x.
3. Independent types: $q(t) = q_1(t_1) \cdots q_N(t_N)$ for all t.

Theorem 6: If x is interim efficient and incentive compatible, then it is essentially implementable using an essentially cheaptalk-proof mechanism.

Proof (from Palfrey and Srivastava [1991b]): It is easiest to prove for the case of transferable utility, so that if we write $x(t) = (a(t), w(t))$ then for all i, t_i, and x we have $U_i(x(t), t_i) = V_i(a(t), t_i) + w_i(t)$. Given the allocation rule x, for each a, i, and t_i, let $P_i(a, t_i)$ be the (interim) probability of outcome a conditional on player i being type t_i and let $W_i(t_i)$ be the expected transfer to player i. The mechanism is defined as follows. The message space for each agent is $T_i \times [0, 1]$. The outcome function is defined below:

(I) If $m_i = (t_i, 0)$ for all i then $g(m) = x(t)$.

(II) If $m_j = (t_j, 0)$ for all $j \neq i$ and $m_i = (t_i \epsilon)$ with $\epsilon > 0$, then a is chosen randomly by the planner according to the distribution $P(\cdot, t_i)$ and i receives a transfer equal to $W_i(t_i) - \epsilon$. All players $j \neq i$ divide the cost of this transfer equally.

(III) If more than one agent reports a positive number then the agent who reports the smallest positive number plays the role of i in (II). If there is a tie, then the agent with the lowest index plays that role.

It is easy to see that if x is incentive compatible then it is a Bayesian equilibrium for everyone to honestly report their type and for no type of any agent to report a positive number. Second, since x is interim efficient if any other type-reporting strategies coupled with always sending 0 is not interim utility equivalent to x then it cannot be an equilibrium, since some type of some agent would be better off reporting some small enough positive number. Third, there can be no equilibrium with any type of any agent reporting a positive number, since it is always better to report a smaller positive number. ∎

This result, and the proof, extends to common value environments where the dependence of U_i on t_{-i} is additively separable from the dependence of U_i on t_i. This includes what Myerson [1981] calls "revision effects." The extent to which cheaptalk-proof implementation can be achieved with more general preferences or in environments with dependent types is an interesting open quesiton. It is also not known whether it is possible to implement inefficient allocations. However, it would seem to be the case that allocation rules such as the one in the example could not be implemented in a cheaptalk-proof way even with sidepayments.[44]

The construction in Theorem 6 exploits the fact that in independent private values models, interim expected utilities depend only on the "reduced form" of an allocation rule, denoted above by $P_i(\cdot)$ and $W_i(\cdot)$, each of which depends only on player i's type. This is also true in common value models as long as the dependence of player i's utility on the other players' types enters separably.

If players' utilities only depend on the reduced form, the planner simply gives players the opportunity of unilateraly "buying out" of the mechanism in exchange for a small payment. If player i buys out of the mechanism, the planner ignores the other players' messages and simply imposes a (possibly random) allocation, depending only on i's reported type, t_i, that is equivalent in i's reduced form to the allocation, x, that the planner was trying to implement. In exchange, the "buyer" of this option must pay a small price. Clearly, if all players other than i are behaving truthfully, then i will not want to buy out. However, as long as x is interim efficient, if players are not behaving truthfully, generating an allocation x_α, then either x_α is interim equivalent to x, or some type of some player will be better off "buying out" at some small enough price, $\epsilon > 0$.

There are several applications of the above result, as a number of revelation principle applications involve models with independent private values, and/or revision efforts.[45] One such application would be the multiple equilibrium problem in the double auction of example 2 in Section 2. For example, with a uniform distribution of types, the allocation rule corresponding to the linear bidding equilibrium of the split-the-difference double auction is interim efficient. Therefore, augmenting the double auction as in Theorem 6 will uniquely implement that allocation, despite the continuum of equilibrium in the direct game.

Endnotes

1. The committment assumption is being relaxed in the recent line of work on renegotiation in contracting. See, for example, Dewatripont [1989], Green and Laffont [1987b], Baron and Besanko [1987], Maskin and Tirole [1989], and Hart and Tirole [1987]. Work on the durability of mechansims is also relevent here. See Crawford [1985], Holmstrom and Myerson [1983], Legros [1990], and Cramton and Palfrey [1990b]. Macroeconomists have also studied a closely related committment problem for a planner ("time consistency"), albeit from a somewhat different perspective. (See, for example, Kydland and Prescott [1977]).

2. These are sometimes called "nuisance" messages. (Moore and Repullo [1988]).

3. That the basic issues are the same does not imply that all the theorems are the same. Where it is relevent we will discuss the differences.

4. This is true, for example, in domains with strict preferences. See Dasgupta, Hammond and Maskin [1979].

5. For example, in public goods environments, we typically need assumption of no income effects. This produces well-known positive reslts, summarized in Green and Laffont [1979] and Groves [1982]. Under domain restrictions, we can get positive results with private goods as well, as in Vickrey [1961]. More recently Mookherjee and Reichelstein [1989] have some further positive results.

6. Most of the early work on Nash implementation was formulated in general equilibrium settings. See for example Hurwicz [1979], Groves and Ledyard [1977], Schmeidler [1980], and the references cited in the Postlewaite [1985] survey. The solution concept of maximin equilibrium has also been explored. (Thomson [1978]).

7. This nearly sounds like a contradiction. However, the planner is uninformed, so each individual has some private information that the planner does not know. The complete information assumption used in Nash equilibrium implementation means that players have redundant information. For example, if the planner does not know the individuals' preference, then by "complete information" we mean that the entire preference profile is common knowledge among the players. For alternative interpretations of Nash implementation, the reader is referred to Maskin's [1986] survey.

8. For characterizations of incentive compatibility in Nash equilibrium, see Mookherjee and Reichelstein [1990]. With more than two players, the incentive compatibility problem can be

entirely avoided.

9. This point is clearly made in Repullo [1987].

10. Information is called "non-exclusive" (Postlewaite and Schmeidler [1986]) if the pooled information of any collection of $N-1$ of the agents refines the infomation of the N^{th} agent. In many applications, non-exclusive information and complete information structures produce essentially identical results (Palfrey and Srivastava [1986], Blume and Easley [1990]).

11. The reason for this is pretty obvious, since Bayesian equilibrium is just a Nash equilibrium with a different information structure – or, alternatively, complete information Nash equilbrium is a special case of Bayesian Nash equilibrium. These connections are discussed at length in Harsanyi [1967-8]. The distinction of the Bayesian implementation approach is most apparent when individuals have "exclusively private information" so that forcing contracts are not available to the planner.

12. As an historical note, the Laffont and Maskin [1982] survey for the Fourth World Congress ten years ago focussed mainly on the incentive compatibility problem. They comment that "Bayesian incentive schemes are plagued by multiple equilibrium" (p. 77).

13. This is not critical. Agents 2 and 3 could have correlated information, and the same problem arises (see Palfrey and Srivastava [1991c]), but the analysis becomes messier.

14. A similar problem arises in the multiple-agent environment studied by Demski and Sappington [1984]. Ma, Moore and Turnbull [1988] construct on indirect mechanism to resolve the multiple equilibrium problem in that evnironment. They refer to it as the problem of "agents cheating." There may also be other *mixed* strategy equilbria, which would depend parametrically on the cardinal utilities. However, implementation theory usually considers implementation via *pure* strategies only. An exception is in a section of Matsushima [1990b].

15. These are *pure stratgies.* Mixed stratgies are, for the most part, ignored in implementation theory.

16. The possibility of inconsistencies arises when there is some degree of non-exclusivity of information.

17. Extensive analysis of related normative concepts may be found in May [1952] and Fishburn [1973].

18. In this section, we state the necessary conditions for social choice correspondence. The definitions for social choice functions are obtained directly.

19. This is Moore and Repullo's [1988] terminology.

20. Details of the construction with independent types are found in Palfrey and Srivastava (1991c).

21. The construction in Palfrey and Srivastava [1989b] allowed free disposal, in the sense that 0 was a feasible allocation. The free disposal assumption is not needed if information is diffuse.

22. Postlewaite and Schmeidler [1986, 1987] and Palfrey and Srivastava [1987, 1989b] use a different representation of information, in which player types are not explicitly mentioned. Instead, for each state $s \in S$, individual i observes an event, $E^i(s)$, which is the set of all states that he cannot distinguish from S. While these representations are essentially identical (see Mookherjee and Reichelstein [1990, p. 474] and Jackson [1991]), the "type" representation is less cumbersome, especially when assuming a common prior $q(t)$ and diffuse information ($q(t) > 0$ for all $t \in T$). We will stick with the type representation here.

23. They also provide a characterization of the set of Bayesian implementable allocations in exchange economies when information may be non-exclusive, and Jackson [1991] provides more general results for that case. See below. Finally, notice that Theorem 4 does not require $N = 3$, in contrast to most other results.

24. Ties can be resolved in any number of ways. For example, they can be broken in favor of the agent with the lowest index.

25. The terminology "public information" and "publicly predictable information" was used in Palfrey and Srivastava [1986] and Blume and Easley [1983], respectively. Those concepts are equivalent to NEI.

26. This statement requires the caveat that there exists a universally bad outcome. The early results were for pure exchange economies, where the **0** outcome served that purpose.

27. In fact, it is equivalent to complete information when there are only two agents.

28. The term splicing comes from Jackson [1991].

29. Closure is not an issue with diffuse information, and does not play a role in the implementation of social choice functions.

30. A deception α is incompatible if $q(\alpha(t)) = 0$ for some t.

31. Generally, rational expectations equilibria are not incentive compatible. See, for example, Blume and Easley [1990]. NEI environments are quite special in this respect.

32. With complete information, this assumption is very weak. With incomplete information, this assumption rules out a large class of interesting environments when a player's type indexes what he knows about the other players' types.

33. There has been no analogous work in Bayesian implementation.

34. One could define $E(\cdot)$ arbitrarily. If E is efficient as s then we will have $E(g(m), s) = g(m)$ at s.

35. The outcomes determined by this bargaining technology may themselves not be attainable as Nash equilibrium to some mechanism.

36. See Crawford [1985], Holmstrom and Myerson [1983], Cramton and Palfrey [1990b], Legros [1990] and the references therein.

37. See Myerson [1983] and Maskin and Tirole [1989, 1990].

38. Perhaps *durably implementable* would be a better term.

39. This follows from the same logic as the proof of Theorem 2 in Legros [1990] or the proof of Theorem 3 in Palfrey and Srivastava [1991b]. A general formulation of interim renegotiation-proofness is in Palfrey and Srivastava [1991c].

40. See footnote 2. These problems can also arise with symmetrically informed agents if some state contingent contracts are not possible.

41. Abreu and Sen [1989a], Herrero and Srivastava [1989], and Moore and Repullo [1988].

42. See for example, Matthews and Postlewaite [1989], Chakravorti [1989], and Zou [1990].

43. Farrell [1983] calls this a communication version of (M, g).

44. But, it may be possible with the addition of an uninformed player.

45. This would include, among others, Myerson [1981], Harris and Raviv [1981], Cramton and Palfrey [1990a] Ledyard and Palfrey [1990], Mailath and Postlewaite [1990], Myerson and Satterthwaite [1983], and others.

References

Abreu, D. and H. Matsushima. (1990). "Virtual Implementation in Iteratively Undominated Strategies: Incomplete Information," mimeo, Princeton University.

Abreu, D. and A. Sen. (1989a). "Subgame Perfect Implementation: A Necessary and Almost Sufficient Condition," *Journal of Economic Theory*, 50:285-99.

Abreu, D. and A. Sen. (1989b). "Virtual Implementation in Nash Equilibrium," mimeo.

Aghion, P., M. Dewatripont and P. Rey. (1989). "Renegotiation Design Under Symmetric Information," mimeo.

Baron, D. and D. Besanko. (1987). "Commitment and Fairness in a Continuing Relationship," *Review of Economic Studies* 54:413-36.

Blume, L. and D. Easley. (1983). "Implementation of Rational Expectations Equilibrium with Strategic Behavior," manuscript, Cornell University.

__________. (1990). "Implementation of Walsarian Expectations Equilibria," *Journal of Economic Theory*, 51:207-27.

Chakravorti, B. (1988). "Mechanisms with no Regret: Welfare Economics and Information Reconsidered," BEBR Faculty Working Paper No. 1438, University of Illinois.

__________. (1989). "Sequential Rationality, Implementation, and Communication in Games," mimeo, University of Illinois.

Chatterjee, K. and W. Samuelson. (1983). "Bargaining under Incomplete Information," *Operations Research*, 31:835-51.

Cramton, P. and T. Palfrey. (1990a). "Cartel Enforcement with Uncertainty about Costs," *International Economic Review*, 31:17-47.

__________. (1990b). "Ratifiable Mechanisms: Learning from Disagreement." Working Paper #731, California Institute of Technology.

Crawford, V. (1985). "Efficient and Durable Decision Rules: A Reformulation," *Econometrica*, 53:817-36.

d'Aspremont, C. and L. Gerard-Varet. (1979). "Incentives and Incomplete Information," *Journal* *of Public Economics*, 11:25-45.

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.

Demski, J. and D. Sappington. (1984). "Optimal Incentive Contracts with Multiple Agents," *Journal of Economic Theory*, 33:152-71.

Dewatripont, M. (1989). "Renegotiation and Information Revelation Over Time: The Case of Optimal Labor Contracts," *Quarterly Journal of Economics*, 104:589-620.

Farrell, J. (1983). "Communication in Games, I: Mechanism Design Without a Mediator," mimeo, MIT.

Fishburn. P. (1973). *The Theory of Social Choice*, Princeton: Princeton University Press.

Fudenberg, D. and J. Tirole. (1990). "Moral Hazard and Renegotiation in Agency Contracts," *Econometrica*, 58:1279-1319.

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

Glazer, J. and R. Rosenthal. (1990). "A Note on Abreu-Matsushima Mechanisms," mimeo, Boston University.

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

__________. (1987a). "Posterior Implementation in a Two-Period Decision Problem," *Econometrica*, 55:69-94.

__________. (1987b). "Renegotiation and the Form of Efficient Contracts," HIER DP no. 1338.

Groves, T. (1982). "On Theories of Incentive Compatible Choice with Compensation," in *Advances in Economic Theory* (W. Hildenbrand ed.) Cambridge: Cambridge University Press.

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

Harris, M. and A. Raviv. (1981). "Allocation Mechanisms and the Design of Auctions," *Econometrica*, 49:1477-99.

Harris, M. and R. Townsend. (1981). "Resource Allocation with Asymmetric Information," *Econometrica*, 49:33-64.

Harsanyi, J. (1967, 1968). "Games with Incomplete Information Played by Bayesian Players," *Management Science*, 14:159-82, 320-34, 486-502.

Hart, O. and J. Tirole. (1987). "Contract Renegotiation and Coasian Dynamics," *Review of Economic Studies*, 55:509-40.

Herrero, M. and S. Srivastava. (1989). "Implementation via Backward Induction," mimeo, Carnegie-Mellon University.

Holmstrom, B. and R. Myerson. (1983). "Efficient and Durable Decision Rules with Incomplete Information," *Econometrica*, 51:1799-1819.

Hurwicz, L. (1972). "On Informationally Decentralized Systems," in *Decision and Organization* (*Volume in Honor of J. Marschak*), R. Radner and C. B. McGuire eds., pp. 297-336, Amsterdam: North Holland.

Hurwicz, L. (1979). "On Allocations Attainable through Nash Equilibria," *Journal of Economic Theory*, 21:140-65.

Jackson, M. (1991). "Bayesian Implementation," *Econometrica*, 59 (in press).

Johnson, S., J. Pratt and R. Zeckhauser. (1990). "Efficiency Despite Mutually Payoff-Relevant Information: The Finite Case," *Econometrica*, 58:873-900.

Kohlberg, E. and J. F. Mertens. (1986). "On the Strategic Stability of Equilibrium," *Econometrica*, 54:1003-38.

Kydland, F. and E. Prescott. (1977). "Rules Rather than Discretion: The Inconsistency of Optimal Plans," *Journal of Political Economy*, 85:473-91.

Laffont, J. J., and E. Maskin. (1982). "The Theory of Incentives: An Overview.," In *Advances in Economic Theory: Invited Papers for the Fourth World Congress of the Econometric Society at Aix-en Provence*, September 1980, editor: W. Hildenbrand, Cambridge University Press, Cambridge, 31-94.

Laffont, J. J. and J. Tirole. (1987). "Comparative Statics of the Optimal Dynamic Incentive

Contract," *European Economic Review*, 31:901-26.

__________. (1988). "The Dynamics of Incentive Contracts," *Econometrica*, 56:1153-75.

__________. (1990). "Adverse Selection and Renegotiation in Procurement," *Review of Economic Studies*, forthcoming.

Ledyard, J. and T. Palfrey. (1990). "On the Optimality of Lottery Drafts: Characterization of Interim Efficiency in a Public Goods Problem," Social Science Working Paper #717, California Institute of Technology.

Legros, P. (1990). "Strongly Durable Allocations," CAE Working Paper #90-05. Cornell University.

Leininger, W., P. Linhart, and R. Radner. (1989). "Equilibria of the Sealed-Bid Mechanism for Bargaining with Incomplete Information," *Journal of Economic Theory*, 48:107-33.

Ma, C., J. Moore, and S. Turnbull. (1988). "Stopping Agents from Cheating," *Journal of Economic Theory*, 46:355-72.

Mailath, G. and A. Postlewaite. (1990). "Asymmetric Information Bargaining Procedures with Many Agents," *Review of Economic Studies*, 57:351-67.

Maskin, E. (1977). "Nash Equilibrium and Welfare Optimality," mimeo, MIT.

__________. (1985). "The Theory of Implementation in Nash Equilibrium: A Survey." In *Social Goals and Social Organization: Volume in Memory of Elisha Pazner*. Cambridge University Press.

Maskin, E. and J. Moore. (1989). "Implementation with Renegotiation," mimeo.

Maskin, E. and J. Tirole. (1989). "The Principal-Agent Relationship with an Informed Principal, II: Common Values," mimeo, MIT.

__________. (1990). "The Principal-Agent Relationship with an Informed Principal, I: Private Values," *Econometrica*, 58:379-410.

Matsushima, H. (1988). "A New Approach to the Implementation Problem," *Journal of Economic Theory*, 45:128-44.

__________. (1990a). "Unique Bayesian Implementation with Budget Balancing," manuscript, Stanford Univesity.

__________. (1990b). "Characterization of Full Bayesian Implementation," manuscript, Stanford University.

Matthews and Postlewaite. (1989). "Pre-play Communication in Two-Person Sealed-bid Double Auctions," *Journal of Economic Theory*, 48:238-63.

May, K. O. (1952). "A Set of Independent Necessary and Sufficient Conditions for Simple Majority Rule," *Econometrica*, 20:680-84.

Mookherjee, D. and S. Reichelstein. (1990). "Implementation Via Augmented Revelation Mechanisms," *Review of Economic Studies*, 57:453-75.

__________. (1989). "Dominant Strategy Implementation of Bayesian Incentive Compatible Allocation Rules," mimeo, Stanford University.

Moore, J., and R. Repullo. (1988). "Subgame Perfect Implementation," *Econometrica*, 46:1191-220.

Myerson, R. (1979). "Incentive Compatibility and the Bargaining Problem," *Econometrica* 47:61-74.

__________. (1981). "Optimal Auction Design," *Mathematics of Operations Research*, 6:58-73.

__________. (1983). "Mechanism Design by an Informed Principal," *Econometrica*, 51:1767-98.

Myerson, R. and M. Satterthwaite. (1983). "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory*, 29:265-81.

Palfrey, T., and S. Srivastava. (1986). "Private Information in Large Economies," *Journal of Economic Theory*, 39:34-58.

__________. (1987). "On Bayesian Implementable Allocations," *Review of Economic Studies*, 54:193-208.

__________. (1989a). "Mechanism Design with Incomplete Information: A Solution to the Implementation Problem," *Journal of Political Economy*, 97:668-91.

__________. (1989b). "Implementation with Incomplete Information in Exchange Economies," *Econometrica*, 57:115-34.

__________. (1991a). "Nash Implementation Using Undominated Strategies," *Econometrica*, 59 (in press).

__________. (1991b). "Efficient Trading Mechanisms with Preplay Communication," *Journal of Economic Theory*, (in press).

__________. (1991c). *Implementation Theory with Incomplete Information*, monograph.

Postlewaite, A. (1985). "Bayesian Implementation via Nash equilibria in economic environments," in *Social goals and social organization (Essays in memory of Elisha Pazner)*, L. Hurwicz, D. Schmeidler, and H. Sonnenschein eds., pp. 205-28, Press Syndicate of the University of Cambridge: New York.

Postlewaite, A. and D. Schmeidler. (1986). "Implementation in Differential Information Economies," *Journal of Economic Theory*, 39:14-33.

__________. (1987). "Differential Information and Strategic Behavior in Economic Environments: A General Equilibrium Approach." In T. Groves, R. Radner and S. Reiter, eds., *Information, Incentives and Economic Mechanisms – Essays in Honor of Leonid Hurwicz*. University of Minnesota Press.

Radner, R. (1979). "Rational Expectations Equilibrium: Generic Existence and the Information Revealed by Prices," *Econometrica*, 47:655-78.

Rajan, M. (1989). "Cost Allocation in Multi-Agent Settings." Mimeo, Carnegie-Mellon University.

Rubenstein, A. and A. Wolinsky. (1989). "Renegotiation-proof Implementation and Time Preferences," mimeo.

Repullo, R. (1987). "On the Revelation Principle with Complete and Incomplete Information," in *The Design of Organizations*, pp. 179-95.

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.

Satterthwaite, M. and S. Williams (1989). "Bilateral Trade with the Sealed Bid Double Auction: Existence and Efficiency." *Journal of Economic Theory*, 48:107-33.

Schmeidler, D. (1980). "Walrasian Analysis via Strategic Outcome Functions," *Econometrica*, 48:1585-94.

Stokey, N. (1979). "Intertemporal Price Discrimination." *Quarterly Journal of Economics*, 93:355-71.

Thomson, W. (1978). "Maximin Strategies and the Elicitation of Preferences," mimeo.

Vickrey, W. (1961). "Counterspeculation, Auctions, and Competitive Sealed Tenders," *Journal of Finance*, 16:1-17.

Wettstein, D. (1986). "Implementation Theory in Economies with Incomplete Information," Working Paper #26-86, Foerder Institute for Economic Research, Tel Aviv University.

Zou, L. (1990). "The Revelation Principle, Multiple Equilibrium, and Communication Regimes," mimeo.