Review of Economic Studies (1990) 57, 453-475 0034-6527/90/00280453$02.00

Implementation via Augmented Revelation Mechanisms

DILIP MOOKHERJEE
Indian Statistical Institute
and
STEFAN REICHELSTEIN
Stanford University

First version received May 1988; *final version accepted January* 1990 (*Eds.*)

Consider the problem of Bayesian implementation, i.e., of constructing mechanisms with the property that all Bayesian equilibrium outcomes agree with a given choice rule. We show that a general procedure is to start with an incentive-compatible revelation mechanism, and then augment agents' message spaces in order to eliminate undesired equilibria. Specifically, we present an Augmented Revelation Principle, which states that if there exists any mechanism that implements a given choice rule, then an augmented revelation mechanism will also implement it. This principle enables us to obtain necessary conditions for implementation. For a large class of environments these conditions are also sufficient.

1. INTRODUCTION

A central problem in the design of organizations is that agents have private information and that they may use this information strategically to advance their own interests. A major part of the mechanism design literature is concerned with incentive mechanisms that seek to align agents' interests with the goals of the organization. The Revelation Principle has played a prominent role in the study of incentive mechanisms. Roughly, this principle says that the performance attainable by some abstract incentive mechanism can also be attained by a revelation mechanism in which agents have an incentive to report their private information truthfully to the mechanism designer.[1] The Revelation Principle thus provides a simple method for representing the constraints imposed by private information and strategic behaviour. The usefulness of this approach has been demonstrated by diverse applications of the theory, for example to auctions, bilateral contracts, internal organization of large firms, and the microfoundation of macroeconomics.

It is important, however, to realize the exact scope of the Revelation Principle. Given an abstract mechanism and a non-cooperative equilibrium for that mechanism, there

1. See, for example, Dasgupta, Hammond, and Maskin (1979), Myerson (1979) and Harris and Townsend (1981). The Revelation Principle has been stated for a number of equilibrium concepts including dominant strategy equilibrium, Bayesian equilibrium, and maximin equilibrium. In this paper we focus on Bayesian equilibrium.

exists a revelation mechanism for which truthful reporting is an equilibrium. Further, the outcomes induced by the truthful reports coincide with the outcomes induced by the non-cooperative equilibrium of the original mechanism. This leaves open the possibility that the revelation mechanism may possess other (untruthful) equilibria whose outcomes differ from the equilibrium outcomes of the original mechanism.

The severity of this multiple equilibrium problem has been illustrated by Demski and Sappington (1984), Postlewaite and Schmeidler (1986) and Repullo (1986). Demski and Sappington consider a setting where two agents separately undertake production with private, correlated information regarding their costs. The optimal revelation mechanism has the property that aside from truthful reporting, there is an untruthful equilibrium where both agents exaggerate their costs. This equilibrium leaves both agents better-off and the principal worse-off relative to the truthful equilibrium. In a social choice context, Postlewaite and Schmeidler (1986) and Repullo (1986) have produced examples of abstract mechanisms with a unique equilibrium yielding desired outcomes while corresponding revelation mechanisms necessarily possess undesired equilibria.

These examples raise the question of whether there is a general way of amending the Revelation Principle to address multiple equilibrium problems. We show in this paper that if the mechanism designer seeks to ensure that *all* noncooperative equilibria of a mechanism achieve a given performance standard, then attention can be focused on a class of augmented revelation mechanisms. In these mechanisms agents can either report their private information, or send some auxiliary "non-type" message. We present an *Augmented Revelation Principle*, which states that any performance standard which can be implemented (in the sense of being achieved by *all* equilibria) by some mechanism, can be implemented by an augmented revelation mechanism for which truthful reporting is one equilibrium.

This result thus shows that the approach adopted by Ma, Moore and Turnbull (1988) exemplifies a general principle. These authors propose dealing with the multiple equilibrium problem in the Demski-Sappington model by allowing some auxiliary "non-type" messages to one of the producing agents. The role of these additional messages is to destroy all untruthful equilibria in the revelation mechanism, without upsetting the truthful equilibrium, or introducing new equilibria.

One of the main uses of the Revelation Principle has been to represent the constraints imposed by incentive compatibility. We use the Augmented Revelation Principle to obtain a correspondingly stronger condition that needs to be satisfied in order to solve the multiple equilibrium problem. This condition will be referred to as the selective elimination condition; it requires that any undesired equilibrium in an incentive compatible revelation mechanism can be eliminated by offering an auxiliary message option to some agent. At the same time, this agent should not have an incentive to deviate from the truth-telling equilibrium to this new message option. Mathematically, this condition reduces to the existence of a solution to a set of inequalities given by the choice rule.

It turns out that the selective elimination condition (SE) is also sufficient (in combination with incentive compatibility (IC)) for the implementability of a given choice rule under very weak economic assumptions. In particular, if the designer is able to make side-payments to different agents in terms of a transferable and divisible good, then IC and SE are sufficient for implementation for a wide variety of allocation problems. We construct implementing mechanisms through a sequence of augmentations of the original revelation mechanism. At each stage, the current mechanism is augmented in order to destroy at least one undesired equilibrium without introducing new equilibria. If there is a finite number of possible types for each agent, only a finite number of augmentations will be required. This enables us to derive upper bounds on the size of the message space needed for implementation.[2]

Our results complement and extend earlier work on implementation in incomplete information environments. Postlewaite and Schmeidler (1986, 1987), and Palfrey and Srivastava (1989) have identified a necessary condition for implementation, called Bayesian monotonicity. This condition reduces to Maskin's (1977, 1986) monotonicity condition in contexts of complete information (where Bayesian equilibrium reduces to Nash equilibrium). These authors have shown that in combination with incentive compatibility, Bayesian monotonicity is also sufficient for a choice rule to be implementable in exchange economies with at least three agents. Given the central role of both the selective elimination and the Bayesian monotonicity condition, one would expect a close relationship between the two conditions. We show that given incentive compatibility, SE is a stronger condition than Bayesian monotonicity, and thereby provides a useful refinement of the latter. In some environments, however, such as those where agents possess exclusive information, the two concepts happen to coincide. Nevertheless, SE appears to be a simpler and more intuitive condition in contexts of incomplete information. Part of the reason is that SE can be used directly in constructing implementing mechanisms, and can be verbally expressed in this fashion. In contrast, monotonicity is a property of responsiveness of social decisions to individual preferences, one which is easily expressed in complete information contexts, but not when information is incomplete.

Our sufficiency results apply to a variety of allocation problems including those considered by Palfrey-Srivastava (1989) and Postlewaite-Schmeidler (1986).[3] Our mechanisms work for the case of two agents exactly as in the case of three or more agents. Further, our setting allows for the presence of public decisions. Thus, a range of contexts including bilateral contracting and regulation are included by our approach.

The mechanisms we construct satisfy global budget balance requirements. In particular, they avoid the undesirable feature that for certain off-equilibrium plays, the mechanism designer imposes a "bad outcome", i.e. appropriates the entire resource endowment, leaving every agent with zero consumption. Also, as mentioned before, our iterative augmentation procedure allows for implementation with finite message spaces. Specifically, we avoid the "tail-chasing" features of previous mechanisms, where agents can announce any non-negative integer and the entire resource endowment is given to the agent announcing the largest integer.

The usefulness of our approach is demonstrated further by the ease with which it extends to alternative implementation requirements. First, we consider unique implementation, where implementing mechanisms are required to possess a unique equilibrium. Such mechanisms avoid the problem of agents having to coordinate on the choice of equilibrium. We show that our necessity and sufficiency results extend straightforwardly, provided the SE condition is suitably strengthened. Secondly, we consider implementation in undominated strategies, which involves a refinement of Bayesian equilibrium as the solution concept. Palfrey and Srivastava (1987b) have shown that incentive compatibility is by itself necessary and sufficient for implementation if one disregards Bayesian equilibria involving dominated strategies. This remarkably strong

2. The finiteness of the message space owes to the fact that we focus only on pure strategy equilibrium in contexts where each agent's type can be one of a finite number of alternatives. In contrast, the mechanisms employed by Ma, Moore and Turnbull (1988) require a continuum message space in order to eliminate all mixed strategy equilibria. We briefly discuss the subject of mixed strategy equilibria in Section 3.

3. Jackson (1988) has recently extended their approach to a larger class of environments, including public goods and the absence of transfer payments. Jackson does, however, assume the existence of at least three agents, and a "bad outcome".

result is readily explained by our approach of augmenting revelation mechanisms. Rather than eliminate a suboptimal equilibrium, it is sufficient to introduce new message options which cause this equilibrium to involve a dominated strategy. Under very mild assumptions on preferences this can be achieved without affecting the truth-telling equilibrium or creating new equilibria. Finally, the mechanisms we construct improve upon those of Palfrey-Srivastava (1987b) in a number of respects.

The paper is organized as follows. Section 2 introduces the model and some basic definitions. Our main results are presented in Section 3 beginning with the Augmented Revelation Principle (Theorem 3.2), the necessity of conditions SE and IC (Theorem 3.5), and their sufficiency (Theorems 3.8 and 3.9). Extensions to unique implementation are also presented. Section 4 then discusses the relation of our approach to previous work. In particular, we explore the connection between our selective elimination condition and Bayesian monotonicity. Finally, Section 5 extends our approach to implementation in undominated strategies.

2. THE MODEL

The set of agents is denoted by $N=\{1,\ldots,n\}$. All relevant information about the economic organization is embodied in the state of the world θ. We assume that the set of all states Θ is finite and has a product structure, i.e. $\Theta=\times_{i=1}^{n}\Theta_i$. Agent $i\in N$ learns θ_i when the state of the world is $\theta=(\theta_1,\ldots,\theta_n)$. Subsequently, we shall refer to θ_i as Agent i's *type*. The set of possible social outcomes is denoted by X. The specification of the set X incorporates all relevant feasibility restrictions such as individual and aggregate resource limitations.

Agent i's preferences are represented by a von Neumann-Morgenstern utility function $U_i\colon X\times\Theta\to\mathbb{R}$. Thus, $U_i(x\mid\theta)$ denotes agent i's utility from the social outcome x, when the state is θ. Given his private information θ_i, each agent has beliefs regarding the underlying state. We assume that these beliefs arise from application of Bayes rule to a common prior (probability) distribution $p(\theta)$ on Θ, satisfying $p(\theta)\geqq 0$ for all $\theta\in\Theta$, and $\sum_{\theta\in\Theta}p(\theta)=1$. By allowing for the possibility that some states have zero prior probability we are able to accomodate a variety of information structures.[4] For instance, previous work on implementation has paid considerable attention to complete information environments, where the prevailing state is common knowledge among all agents. In that case $p(\theta)>0$ if and only if $\theta_1=\theta_2=\cdots=\theta_n$. It will be notationally convenient to denote the set of possible states by

$$\Theta^*=\{\theta\in\Theta\mid p(\theta)>0\}.$$

Without loss of generality, the marginal probability of any type $\theta_i\colon p_i(\theta_i)=\sum_{\theta_{-i}\in\Theta_{-i}}p(\theta_{-i},\theta_i)$, is strictly positive (where Θ_{-i} denotes $\times_{j\neq i}\Theta_j$, and θ_{-i} is typical element of Θ_{-i}). The beliefs of type θ_i of agent i are then represented by the conditional probability distribution $q_i(\theta_{-i}\mid\theta_i)$ over Θ_{-i}:

$$q_i(\theta_{-i}\mid\theta_i)=\frac{p(\theta_{-i},\theta_i)}{p_i(\theta_i)}. \qquad (1)$$

These beliefs are assumed to be common knowledge between the agents. The mechanism designer knows the prior distribution $p(\theta)$, but not the information θ_i of any agent.

4. At first glance it might appear that our modelling of a state and the information structure differs from that in Postlewaite-Schmeidler (1987) and Palfrey-Srivastava (1989). As shown in Appendix B, however, the two formulations are equivalent.

The goals of the mechanism designer are represented by a *social choice correspondence* (SCC) $F:\Theta^* \twoheadrightarrow X$ which specifies for any $\theta \in \Theta^*$ the set of desired outcomes $F(\theta)$. A *mechanism* $\Lambda = \langle M, g\rangle$ consists of a set of possible messages M, and an outcome function $g: M \to X$. The message space M is the product of individual message spaces M_i, i.e. $M = \times_{i=1}^n M_i$. Given the mechanism Λ, we define a *strategy for agent i* to be a function $\alpha_i: \Theta_i \to M_i$, i.e. where type θ_i of agent i chooses the message $\alpha_i(\theta_i)$. Let $\alpha = (\alpha_1, \dots, \alpha_n)$ denote an n-tuple of strategies, one for each agent. Also, let α_{-i} denote $(\alpha_1, \dots, \alpha_{i-1}, \alpha_{i+1}, \dots, \alpha_n)$, and $\alpha_{-i}(\theta_{-i})$ denote the message-tuple $(\alpha_1(\theta_1), \dots, \alpha_{i-1}(\theta_{i-1}), \alpha_{i+1}(\theta_{i+1}), \dots, \alpha_n(\theta_n))$. Then given the tuple α_{-i} of strategies used by agents other than i, we define the expected utility of type θ_i of agent i from message $m_i \in M_i$ as

$$V_i(\alpha_{-i}, m_i \mid \theta_i) = \textstyle\sum_{\theta_{-i} \in \Theta_{-i}} q_i(\theta_{-i} \mid \theta_i) U_i(g(\alpha_{-i}(\theta_{-i}), m_i) \mid \theta). \tag{2}$$

Given the mechanism Λ, a *pure-strategy Bayesian equilibrium* is an n-tuple $\alpha = (\alpha_1, \dots, \alpha_n)$ of strategies with the property that for all $i \in N$ and all $\theta_i \in \Theta_i$:

$$\alpha_i(\theta_i) \in \arg\max_{m_i \in M_i} V_i(\alpha_{-i}, m_i \mid \theta_i). \tag{3}$$

Unless otherwise mentioned, we shall henceforth refer to a pure-strategy Bayesian equilibrium simply as an equilibrium. Note that for any equilibrium α of Λ the composition $g \circ \alpha$ induces a function that maps from Θ to X. This function is said to agree with the SCC F, if for all possible states $\theta \in \Theta^*$: $(g \circ \alpha)(\theta) \in F(\theta)$. For brevity, we denote this as $(g \circ \alpha) \in F$.

Definition 2.1. The mechanism $\Lambda = \langle M, g\rangle$ is said to *implement* the social choice correspondence F if

(i) Λ has an equilibrium α such that $(g \circ \alpha) \in F$,
(ii) every equilibrium $\bar{\alpha}$ in Λ satisfies $(g \circ \bar{\alpha}) \in F$.[5]

Earlier work on incentive mechanisms (for example, Harris and Townsend (1981) and Myerson (1979)) has focused attention on revelation mechanisms. In a revelation mechanism $M_i = \Theta_i$, for all $i \in N$. We shall refer to the outcome function of a revelation mechanism as a revelation function. For any revelation function $f: \Theta \to X$ the induced revelation mechanism $\langle \Theta, f\rangle$ will be denoted by Λ_f.

Definition 2.2. A revelation mechanism Λ_f is said to be *incentive compatible* (IC), if truthful reporting is an equilibrium, i.e. for all $i \in N$ and all $\theta_i, \bar{\theta}_i \in \Theta_i$:

$$\textstyle\sum_{\theta_{-i}} q_i(\theta_{-i} \mid \theta_i)[U_i(f(\theta_{-i}, \theta_i) \mid \theta) - U_i(f(\theta_{-i}, \bar{\theta}_i) \mid \theta)] \geqq 0. \tag{4}$$

Accordingly, we say that a revelation function is IC if the induced revelation mechanism is incentive compatible.

3. NECESSARY AND SUFFICIENT CONDITIONS FOR IMPLEMENTATION

3.1. *The augmented revelation principle*

As observed by Repullo (1986) and Postlewaite-Schmeidler (1986), it does not suffice to search over the class of revelation mechanisms when the design task is to construct incentive mechanisms with the property that all equilibria achieve desired allocations. In this section we demonstrate, though, that one may confine attention to a class of mechanisms wherein agents can either report their private information or send some other message.

5. The literature has considered a variety of implementation concepts. In subsequent sections and Appendix B we sketch how our results can be extended to accommodate alternative notions of implementation.

Definition 3.1. An augmented revelation mechanism is a mechanism $\Lambda = \langle M, g\rangle$ such that for all $i \in N$:

$$M_i = \Theta_i \cup T_i, \text{ where } T_i \text{ is an arbitrary set.}$$

In the above definition, T_i represents the set of additional message options made available to Agent i, in addition to messages about his type. We now establish that a designer interested in implementing a given SCC can focus on the class of augmented revelation mechanisms, without loss of generality. At first such a proposition might seem trivial since one can always add "dummy" messages corresponding to agents' types. Any implementing mechanism could then be extended so that no agent will ever wish to utilize one of these "dummy" messages. The essential feature of the following result is that truthful reporting of private information must constitute an equilibrium of the mechanism (and thus yield a desired allocation). Thus, the inclusion of all type messages is not an artificial restriction, as agents must use these messages in at least one equilibrium.

Theorem 3.2. *If $F:\Theta^* \twoheadrightarrow X$ is implementable, then F can be implemented by an augmented revelation mechanism, in which truthful reporting is an equilibrium.*

Proof. Let $\Lambda = \langle M, g\rangle$ implement F. We shall construct an augmented revelation mechanism $\bar{\Lambda} = \langle \bar{M}, \bar{g}\rangle$. Given any equilibrium $(\alpha_1, \dots, \alpha_n)$ in Λ, define $\bar{M}_i = \Theta_i \cup T_i$ with $T_i = \{m_i \in M_i \mid m_i \notin \text{range}(\alpha_i)\}$. Consider the functions $\phi_i : \bar{M}_i \to M_i$ with

$$\phi_i(\bar{m}_i) = \begin{cases} \alpha_i(\theta_i) & \text{if } \bar{m}_i = \theta_i \text{ for } \theta_i \in \Theta_i \\ \bar{m}_i & \text{if } \bar{m}_i \in T_i, \end{cases}$$

and define the outcome function $\bar{g}: \bar{M} \to X$ by $\bar{g} = g \circ \phi$, where $\phi = (\phi_1, \dots, \phi_n)$. It is obvious that truth-telling is an equilibrium in $\bar{\Lambda}$. Furthermore, $\bar{g}(\theta) = g(\alpha(\theta)) \in F(\theta)$ for all $\theta \in \Theta^*$.

To check part (ii) of the implementation requirement suppose that $(\bar{\alpha}_1, \dots, \bar{\alpha}_n)$ is an equilibrium for $\bar{\Lambda}$. Consider the strategies $\alpha_i^* = \phi_i \circ \bar{\alpha}_i$ in Λ. Since the outcome of α^* in Λ coincides with that of $\bar{\alpha}$ in $\bar{\Lambda}$ i.e. $(g \circ \alpha^*) = (\bar{g} \circ \bar{\alpha})$, it suffices to show that α^* is an equilibrium in the original mechanism Λ, because by hypothesis the latter implements F. Since $\bar{\alpha}$ is an equilibrium in $\bar{\Lambda}$, we have for any $i \in N$ and $\theta_i \in \Theta_i$:

$$\begin{aligned} \textstyle\sum_{\theta_{-i}} q_i(\theta_{-i} \mid \theta_i) U_i(g(\alpha^*(\theta)) \mid \theta) &\equiv \textstyle\sum_{\theta_{-i}} q_i(\theta_{-i} \mid \theta_i) U_i(\bar{g}(\bar{\alpha}(\theta)) \mid \theta) \\ &\geqq \textstyle\sum_{\theta_{-i}} q_i(\theta_{-i} \mid \theta_i) U_i(\bar{g}(\bar{\alpha}_{-i}(\theta_{-i}), \bar{m}_i) \mid \theta) \\ &= \textstyle\sum_{\theta_{-i}} q_i(\theta_{-i} \mid \theta_i) U_i(g(\alpha_{-i}^*(\theta_{-i}), \phi_i(\bar{m}_i)) \mid \theta) \qquad (5) \end{aligned}$$

for all $\bar{m}_i \in \bar{M}_i$. The result now follows from the fact that the map ϕ_i is onto its range M_i. ||

This result shows that the approach adopted by Ma, Moore, and Turnbull (1988) exemplifies a general principle. If the problem of multiple equilibria can be solved at all, it can be dealt with by augmenting an incentive compatible revelation mechanism.

3.2. *Necessary conditions for implementation*

The Augmented Revelation Principle immediately suggests a necessary condition for implementation. Consider the revelation mechanism that is part of an augmented revelation mechanism implementing the SCC F. If this revelation mechanism admits an equilibrium which results in undesired outcomes, it must be true that some agent has a non-type message (in the augmented mechanism) which eliminates this suboptimal equilibrium without upsetting the truth-telling equilibrium.

Definition 3.3. An equilibrium $\alpha=(\alpha_1,\dots,\alpha_n)$ in a revelation mechanism Λ_f can be *selectively eliminated*, if there exists $i\in N$ and $h:\Theta_{-i}\to X$ such that:

(i) for some $\bar{\theta}_i\in\Theta_i$

$$\textstyle\sum_{\theta_{-i}} q_i(\theta_{-i}\mid\bar{\theta}_i)[U_i(h(\alpha_{-i}(\theta_{-i}))\mid\theta_{-i},\bar{\theta}_i)-U_i(f(\alpha_{-i}(\theta_{-i}),\alpha_i(\bar{\theta}_i))\mid\theta_{-i},\bar{\theta}_i)]>0 \qquad (6)$$

(ii) for all $\theta_i\in\Theta_i$

$$\textstyle\sum_{\theta_{-i}} q_i(\theta_{-i}\mid\theta_i)[U_i(f(\theta)\mid\theta)-U_i(h(\theta_{-i})\mid\theta)]\geqq 0. \qquad (7)$$

In the above definition, Agent i is offered a new message option, which we refer to as a "flag". When i chooses the flag rather than a type message, and other agents report θ_{-i} the outcome chosen by the mechanism is given by $h(\theta_{-i})$. Condition (6) says that i prefers to deviate from α_i to the flag in some state, thereby destroying α as an equilibrium. However, (7) ensures that Agent i does not deviate from truth-telling to the flag, provided other agents are also truthful. The truthful equilibrium is thus preserved.

Definition 3.4. The revelation mechanism Λ_f satisfies the *Selective Elimination* (SE) condition relative to F, if $f\in F$ and if every equilibrium α in Λ_f satisfying $(f\circ\alpha)\notin F$ can be selectively eliminated.

Theorem 3.5. *Suppose $F:\Theta^*\twoheadrightarrow X$ is implementable. Then there exists an incentive compatible revelation mechanism Λ_f which satisfies the selective elimination condition relative to F.*

The proof of this theorem follows directly from Theorem 3.2. Given an augmented revelation mechanism $\bar{\Lambda}=\langle M,g\rangle$ implementing F, one obtains a revelation mechanism Λ_f by considering the type messages of $\bar{\Lambda}$ and letting $f=g|_\Theta$. Since truthful reporting is an equilibrium in $\bar{\Lambda}$, it follows that $f\in F$. Further, any suboptimal equilibrium α of Λ_f cannot be an equilibrium of $\bar{\Lambda}$. There must, thus, exist a non-type message available to some Agent i such that i prefers to deviate to this message when α is being played, without being tempted to do the same when all agents report truthfully. This establishes that selective elimination is a necessary condition for implementation.

Mathematically, condition SE requires the existence of a solution to a set of inequalities given by the SCC F. To check this condition may be rather difficult in complex environments, since one first needs to find the equilibria of the revelation mechanism and then check inequalities (6) and (7). However, as argued in Section 4 below, this task appears to be simpler than checking the validity of Bayesian monotonicity, another necessary condition for implementation which has been the focus of previous literature. Bayesian monotonicity requires that analogous inequalities be satisfied for any reporting strategy, irrespective of whether or not it is an equilibrium.[6] Since in many principal-agent models the structure of preferences and information permits at least a partial characterization of equilibria, SE will be simpler to verify than Bayesian monotonicity, to the extent that SE allows one to discard directly those strategies that cannot possibly be equilibria.

While implementation requires that all equilibria comply with the SCC, it is left unspecified how agents coordinate on the choice of equilibrium. This problem could be avoided by insisting on mechanisms with a unique equilibrium. We now show that the

6. Strictly speaking, Bayesian monotonicity requires these inequalities to be satisfied for all reporting strategies involving compatible reports, i.e. where the reported state is not a zero probability state.

selective elimination approach can be used directly to obtain conditions for unique implementation, i.e. the existence of mechanisms with a unique equilibrium attaining the outcomes prescribed by the SCC. Based on the preceding arguments one might expect that a necessary condition for unique implementation is that any equilibrium other than truth-telling can be selectively eliminated. This turns out to be true except for equivalent strategies. Define two messages θ_i, $\bar{\theta}_i$ in a revelation mechanism Λ_f to be *equivalent* if they give rise to identical outcomes in all situations: $f(\theta_{-i}, \theta_i) = f(\theta_{-i}, \bar{\theta}_i)$ for all $\theta_{-i} \in \Theta_{-i}$. Correspondingly, define two strategy-tuples α and α^* in Λ_f to be equivalent if $\alpha_i(\theta_i)$ is equivalent to $\alpha_i^*(\theta_i)$, for all $i \in N$ and all $\theta_i \in \Theta_i$.

Definition 3.6. The revelation mechanism Λ_f satisfies the Strong Selective Elimination (SSE) condition relative to the SCC F, if $f \in F$, and if any equilibrium α in the revelation mechanism Λ_f which is not equivalent to truthful reporting, can be selectively eliminated.

Thus, SSE strengthens condition SE by requiring the selective elimination of any equilibirum of the revelation mechanism that is not equivalent to truth-telling, even if it does lead to desirable outcomes.

Theorem 3.7. *Suppose $F: \Theta^* \twoheadrightarrow X$ is uniquely implementable. Then there exists an incentive compatible revelation mechanism Λ_f which satisfies the SSE condition relative to F.*

Proof. See Appendix A. ||

3.3. *Sufficient conditions for implementation*

Incentive compatibility and the selective elimination condition ensure the existence of a revelation mechanism with one desirable equilibrium and the property that all suboptimal equilibria can be selectively eliminated. These conditions, however, do not guarantee that F is implementable, since the augmentation of the revelation mechanism may give rise to new undesired equilibria. This possibility is demonstrated by example in a related paper, Mookherjee and Reichelstein (1989). An essential feature of that example is that the set of alternatives X is narrow (there are only two alternative outcomes). In contrast, the following results show that when the mechanism designer can use private transfer payments to reward or punish agents, then IC and SE are essentially sufficient for implementation.

To understand the basic idea underlying our construction, suppose that α is a suboptimal equilibrium in Λ_f. Condition SE says that some agent can be assigned a flag which upsets the equilibrium α. To ensure that the flag is never raised in equilibrium, we use the following augmentation. Some other agent is given a new set of messages, referred to as counterflags. The outcome function is extended in the following manner. Counterflags pay-off for the second agent only if the first agent chooses his flag option. In turn, the first agent does not want to raise the flag, if the other agent plays a counterflag. Table I illustrates the augmented mechanism.

For simplicity, we illustrate the case where a given endowment of private goods is allocated among two agents. In the augmented revelation mechanism, Agent 1 has the option of raising the flag FL in addition to playing his type messages. Agent 2 is given one counterflag corresponding to each of his type messages. Condition SE specifies the outcomes if Agent 1's flag meets type messages from Agent 2. If Agent 2 plays CFL^j against Agent 1's type messages, the resulting outcome is as if Agent 2 had played θ_2^j,

TABLE I

Agent 1'*s Payoffs*

	$\theta_2^1 \cdots \theta_2^{k_2}$	$CFL^1 \cdots CFL^j \cdots CFL^{k_2}$
θ_1^1 $\vdots$ $\theta_1^{k_1}$	$f_1(\theta_1, \theta_2)$	$f_1(\theta_1, \theta_2) + \varepsilon$
FL	$h_1(\theta_2)$	$h_1(\theta_2) - \delta$

Agent 2'*s Payoffs*

	$\theta_2^1 \cdots \theta_1^{k_2}$	$CFL^1 \cdots CFL^j \cdots CFL^{k_2}$
θ_1^1 $\vdots$ $\theta_1^{k_1}$	$f_2(\theta_1, \theta_2)$	$f_2(\theta_1, \theta_2) - \varepsilon$
FL	$h_2(\theta_2)$	$h_2(\theta_2) + \delta$

except that Agent 2 is charged an amount ε which is transferred to Agent 1 (subscript i denotes the allocation to Agent i). If CFL^j meets FL, then the outcome is as if Agent 2 reported θ_2^j, and Agent 1 chose FL, with the exception that Agent 2 is rewarded with δ, which is transferred from Agent 1.

By definition of SE, the strategy tuple α ceases to be an equilibrium in the augmented mechanism. Truthful reporting, however, continues to be an equilibrium. For Agent 1 this follows directly from condition *SE*, while for Agent 2 it is implied by the ε-charge for raising *CFL*'s. Finally, for suitable choices of ε and δ, the augmented mechanism has no equilibrium involving either *FL*, or any CFL^j. Suppose to the contrary that Agent 1 plays *FL* in some instances. If the charge ε is sufficiently small relative to the reward δ, every type of Agent 2 will prefer CFL^j over θ_2^j. Under relatively weak assumptions (developed below) there exists δ such that every type of Agent 1 prefers to play some type-message instead of *FL*. Hence, there exists no equilibrium in which Agent 1 plays *FL*. This implies that there is also no equilibrium in which Agent 2 chooses any *CFL*. In summary, the suboptimal equilibrium α has been eliminated, while truthful reporting remains an equilibrium in the augmented mechanism. Furthermore, there is no equilibrium in the augmented mechanism in which any agent chooses a non-type message. By repeating this procedure, all suboptimal equilibria of Λ_f can be eliminated through successive augmentations.

To state our sufficiency results formally, we introduce the following economic structure into the model. Assume there are l private goods and let $x_i \in \mathscr{R}_+^l$ denote a consumption vector of these private goods for Agent i. In addition, a public decision y has to be made from a set of possible decisions Y. Let $C_i \subset Y \times \mathscr{R}_+^l$ denote the set of allocations that are feasible for Agent i. Note that C_i is assumed to be independent of Agent i's characteristics. By C_0 we denote all allocations $(y, x_1, \ldots, x_n)$ which are socially feasible, i.e. allocations that satisfy the technological and aggregate resource constraints of the organization. We then have

$$X = \{(y, x_1, \ldots, x_n) \in C_0 \mid (y, x_i) \in C_i \text{ for all } i \in N\}.$$

While this setup corresponds to general resource allocation problems with private and public goods, it applies in particular to the environments considered typically in principal-agent models. In the context of a multidivisional firm, the public decision y

may represent production assignments. Alternatively, in a bargaining or auction context y may represent the allocation of a set of goods amongst the participants. We shall use the following assumptions regarding agents' preferences and feasible sets.

A1: $C_0 = \{(y, x_1, \ldots, x_n) \mid y \in Y, x_i \in \mathscr{R}^l_+\}$.

A2: Given any $(y, x_i) \in C_i$: $(y, x_i + \delta) \in C_i$ for all $\delta \geqq 0$. Further, there exists $\varepsilon > 0$ such that $(y, x_i - \varepsilon) \in C_i$.[7]

A3: Agent i's utility function $U_i(y, x_i \mid \theta)$ is continuous and strictly increasing in all components of x_i.

A4: Given any Agent $i \in N$, any $(\bar{y}, \bar{x}_i) \in C_i$ and any other $y \in Y$, we can find $x_i \in \mathscr{R}^l_+$ such that $(y, x_i) \in C_i$ and

$$\textstyle\sum_{\theta_{-1}} q_i(\theta_{-i} \mid \theta_i)[U_i(y, x_i \mid \theta) - U_i(\bar{y}, \bar{x}_i \mid \theta)] \geqq 0 \quad \text{for all } \theta_i \in \Theta_i.$$

Assumption A1 implies that there are no aggregate feasibility constraints. In this respect the model applies primarily to principal-agent situations in which the principal has enough resources not to be constrained by an ex-post budget balance condition. (Note that technological constraints are embodied in the specification of the set Y.) A2 requires the availability of small rewards or penalties in terms of private goods, starting from any feasible outcome. The individually feasible sets must be "open" in this sense. However, an alternative to A2 would be to assume that C_i is closed, and instead make an interiority assumption on the allocations given by the revelation mechanism Λ_f, as well as on the allocations specified by the function $h(\cdot)$ given by SE. Assumption A4 requires that given any feasible allocation and any alternative public decision, there exist private transfers which compensate agents for a switch in the public decision.

Theorem 3.8. *Suppose assumptions* A1–A4 *hold. Then $F: \Theta^* \twoheadrightarrow X$ is implementable, if there exists an incentive compatible revelation mechanism which satisfies SE relative to F. Furthermore, F can be implemented by a finite mechanism, i.e. one with finite message spaces for each agent.*

The proof of this theorem begins with the revelation mechanism Λ_f, and employs a finite sequence of augmentations. At each stage one suboptimal equilibrium of the previous-stage mechanism is selectively eliminated, without introducing any new equilibria. Since the sets Θ_i are finite there can only be a finite number of equilibria in Λ_f, requiring only a finite number of augmentation stages to eliminate all undesired equilibria. At every stage of this procedure, one agent is given a flag while one other agent is given a set of counterflags (as many as the cardinality of the type space of that agent). The outcome function is extended in a manner which differs somewhat from the illustration in Table I, owing to the existence of public decisions in this setting. The main difference is in the manner in which Agent 1 is induced to switch to the type message θ_1^1 from the flag FL whenever Agent 2 chooses some counterflag CFL^j. Assumption A4 ensures the possibility of constructing private good transfers that sufficiently compensate Agent 1 for the switch in the public decision between the message combinations (FL, CFL^j) and (θ_1^1, CFL^j).

As a byproduct of Theorem 3.8 one obtains an upper bound on the size of the message space needed for implementation. If r is the smallest number of suboptimal equilibria amongst all incentive compatible revelation mechanisms Λ_f satisfying SE

7. The notation $x_i > 0$ denotes $x_i \in \mathbb{R}^l_+$ and $x_i \neq 0$. In contrast, $x_i \gg 0$ or $x_i \in \mathbb{R}^l_{++}$ denotes the situation where all components of x_i are positive.

relative to F, the maximum number of messages needed for implementation (when aggregated over all agents) is

$$\sum_{i=1}^{n} k_i + r(k+1), \tag{8}$$

where $k_i = |\Theta_i|$ and $k = \max\{k_1, \dots, k_n\}$. The formula in (8) reflects that corresponding to each suboptimal equilibrium some Agent is given a flag and another Agent j is given k_j counterflags ($k_j = |\Theta_j|$). We note that the number of counterflags could be reduced substantially, if one were to allow for large penalties and rewards. A single *CFL* would then be sufficient to guarantee that the *FL* option does not give rise to new equilibria. The agent raising *CFL* would be "showered with gold" if his *CFL* were to be played against *FL*, and punished sufficiently if *CFL* were to meet type messages only.

We now provide another sufficiency theorem for environments requiring balanced allocations for all possible strategy combinations. For simplicity, we confine attention to exchange economies without any public decisions.[8] In the following assumptions, let $w \in \mathscr{R}^l_+$ denote the organization's aggregate resource endowment.

A5: $C_0 = \{(x_1, \dots, x_n) \in \mathscr{R}^{n \cdot l} | \sum_{i=1}^{n} x_i = w\}$.

A6: $C_i \subset \mathscr{R}^l_+$; if $x_i \in C_i$ then $(x_i + \delta) \in C_i$ for all $\delta > 0$. For all $x_i \in C_i$, there exists $\varepsilon > 0$ such that $(x_i - \varepsilon) \in C_i$.

A7: For all $x_i, \bar{x}_i \in C_i$, there exists $\delta \geqq 0$ such that $(x_i - \delta) \in C_i$ and

$$\sum_{\theta_{-i}} q_i(\theta_{-i} | \theta_i)[U_i(\bar{x}_i | \theta) - U_i(x_i - \delta | \theta)] \geqq 0 \quad \text{for all } \theta_i \in \Theta_i.$$

In the following theorem one could replace A6 and A7 by assumptions saying that the social choice rule F and the outcome function $h(\cdot)$ (given by condition *SE*) provide every agent with positive consumption of every good, and that the consumption sets C_i be equal to $\mathscr{R}^l_+$.

Theorem 3.9. *Suppose* A3 *and* A5–A7 *hold. Then* $F: \Theta^* \twoheadrightarrow X$ *is implementable, if there exists an incentive-compatible revelation mechanism which satisfies SE relative to F. Furthermore, F can be implemented by a finite mechanism.*

Proof. See Appendix A. ||

The proof of this result employs a finite sequence of augmentations, where at each stage one suboptimal equilibrium is eliminated in the manner depicted in Table I. Assumption A7 ensures the existence of penalties large enough to induce Agent 1 to switch from *FL* to a type message when Agent 2 chooses counterflags instead of type messages. The non-existence of new equilibria in the augmented mechanism is thus ensured in a manner different from that employed in the previous sufficiency theorem (where Agent 1 received large rewards if his type message θ_1^1 met some counterflag of Agent 2). In other respects, however, the proofs of the two theorems are similar.

If we seek to implement F uniquely, then Theorems 3.8 and 3.9 extend as long as the SE condition is strengthened to SSE (the condition we identified as necessary for unique implementation in Theorem 3.7).

8. In the presence of public decisions the following analogue of Theorem 3.9 holds.

A5′: if $(y, x_1 \cdots x_n) \in C_0$ then $(y, \bar{x}_1 \cdots \bar{x}_n) \in C_0$ whenever $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \bar{x}_i$.

A6′: $((y, x_i + \delta) \in C_i$ whenever $(y, x_i) \in C_i$ and $\delta \geqq 0$. Furthermore, for any $(y, x_i) \in C_i$, there exists an $\varepsilon > 0$ such that $(y, x_i - \varepsilon) \in C_i$.

A7′: for any $(y, x_i) \in C_i$, $(\bar{y}, \bar{x}_i) \in C_i$, there exists $\delta \geqq 0$ such that

$$\sum_{\theta_{-i}} q_i(\theta_{-i} | \theta_i)[U_i(\bar{y}, \bar{x}_i | \theta) - U_i(y, x_i - \delta | \theta)] > 0$$

for all $\theta_i \in \Theta_i$. If these assumptions are substituted, then the SCC F is implementable under the conditions of Theorem 3.9.

Theorem 3.10 *Suppose the conditions of either Theorem* 3.8 *or* 3.9 *hold with SSE replacing SE. Then* $F: \Theta^* \twoheadrightarrow X$ *can be uniquely implemented by a finite mechanism.*

The proof parallels that of Theorems 3.8 and 3.9. To obtain a unique equilibrium, however, one first has to "shrink" the revelation mechanism Λ_f in order to obtain a mechanism without equivalent messages. Thereafter, any non-truthful equilibria in Λ_f are not equivalent to truth-telling. Condition SSE and the conditions of either Theorem 3.8 or 3.9 then ensure that any non-truthful equilibrium of the "shrunk" revelation mechanism can be eliminated through successive augmentations.

We finish this section with a few remarks concerning mixed strategies. It can be verified that the Augmented Revelation Principle extends to this notion of implementation, and so do the analogous versions of IC and SE. Our sufficiency results may need to be modified, however, One reason is that the initial revelation mechanism may have an infinite number of mixed strategy equilibria, and therefore finite mechanisms may not suffice. Even if the underlying revelation mechanism has a finite number of mixed strategy equilibria, countably infinite message spaces may be required to ensure the nonexistence of any mixed strategy equilibrium involving flags or counterflags with positive probability. One may associate with every flag an infinite sequence of corresponding counterflags, where ε gets progressively smaller relative to δ. If Agent 1 raises the flag with positive probability, no matter how small, there will then exist a corresponding set of counterflags that will be raised by Agent 2, and our previous arguments will apply from that point onward. Thus, it appears that Theorems 3.8 and 3.9 can be extended to implementation in mixed strategies, except that finite mechanisms may not suffice.

4. RELATION TO PREVIOUS LITERATURE

We now relate our results to the work of Postlewaite-Schmeidler (1986) and Palfrey-Srivastava (1989, 1987a). These authors identify the Bayesian monotonicity condition as necessary for implementation. Further, they have shown that subject to certain economic assumptions, Bayesian monotonicity in conjunction with incentive compatibility is sufficient for a choice rule to be implementable.

Before we explore the connection between Bayesian monotonicity and the selective elimination condition, we note a number of differences between the Postlewaite-Schmeidler and Palfrey-Srivastava sufficiency results and the results presented in Theorems 3.8 and 3.9. First, these authors are concerned with exchange economies with three or more agents. Our setting allows for the presence of both private and public goods. In addition, our mechanisms work for two agents in exactly the same manner as they do for three or more agents. Thus our results can be applied to a variety of bilateral contracting and agency models. Secondly, the mechanism presented in Theorem 3.9 achieves balanced allocations even if agents fail to reach an equilibrium. To avoid undesirable equilibria, Postlewaite-Schmeidler and Palfrey-Srivastava assume that the mechanism designer can appropriate the economy's entire resource endowment, leaving every agent with zero consumption. None of our sufficiency results relies on the existence of a "bad outcome", i.e. an outcome that is ranked worst by all agents irrespective of the prevailing state.

Finally, the mechanisms we construct are more economical with regard to the size of agents' message space. In contrast to the previous mechanisms, we find that finite message spaces are sufficient provided agents' type spaces are finite. In particular, we do not eliminate equilibria through infinite "tail-chasing" games, e.g. where agents can announce any non-negative integer and the economy's entire resources are given to that agent who announces the largest integer.

We now turn to a comparison of Bayesian monotonicity and the selective elimination condition. Bayesian Monotonicity is a generalization of Maskin's (1977) monotonicity condition developed for Nash implementation. Though Bayesian monotonicity is usually stated for a SCC F, it will be more convenient here to define this property for a revelation mechanism.[9]

Definition 4.1. The revelation mechanism Λ_f satisfies *Bayesian Monotonicity* (BM) relative to F if $f \in F$, and if for any reporting strategy $\alpha = (\alpha_1, \ldots, \alpha_n)$ with $\alpha(\theta) \in \Theta^*$ whenever $\theta \in \Theta^*$, the following condition holds. If $(f \circ \alpha) \notin F$, there exists Agent i, state $\bar{\theta} \in \Theta^*$ and a function $t: \Theta \to X$ such that

$$\textstyle\sum_{\theta_{-i}} q_i(\theta_{-i} \mid \bar{\theta}_i)[U_i(f(\alpha_{-i}(\theta_{-i}), \alpha_i(\bar{\theta}_i)) \mid \theta_{-i}, \bar{\theta}_i) - U_i(t(\alpha_{-i}(\theta_{-i}), \alpha_i(\bar{\theta}_i)) \mid \theta_{-i}, \bar{\theta}_i)] < 0, \tag{9}$$

and

$$\textstyle\sum_{\theta_{-i}} q_i(\theta_{-i} \mid \theta_i)[U_i(f(\theta_{-i}, \theta_i) \mid \theta) - U_i(t(\theta_{-i}, \alpha_i(\bar{\theta}_i)) \mid \theta)] \geqq 0 \tag{10}$$

for all $\theta_i \in \Theta_i$ such that $(\theta_i, \alpha_{-i}(\bar{\theta}_{-i})) \in \Theta^*$.

This definition adapts that of Palfrey-Srivastava (1989) in order to facilitate comparison with SE. It is straightforward to show that a SCC F satisfies Bayesian monotonicity in the sense of Palfrey-Srivastava (1989) if and only if every revelation mechanism Λ_f with $f \in F$ satisfies Bayesian monotonicity relative to F (according to Definition 4.1).[10] Furthermore, if a SCC F is implementable, then there exists a revelation mechanism Λ_f with $f \in F$ that is incentive compatible and satisfies Bayesian monotonicity relative to F.

Direct comparison of Definitions 4.1 and 3.3 suggests a strong relation between BM and SE. Inequalities (9) and (10) seem to correspond to (6) and (7) respectively, provided $t(\theta_{-i}, \alpha_i(\bar{\theta}_i))$ is identified with the function $h(\theta_{-i})$ specifying the outcomes associated with the flag. Upon closer inspection, however, the following differences between the two conditions become evident. First, the two conditions pertain to different classes of (undesirable) reporting strategies. BM is concerned with any reporting strategy that maps compatible states (i.e., the set Θ^*) into compatible states. The SE condition, on the other hand, applies to any strategy-tuple that is an equilibrium of Λ_f. The essential aspect of Example 4.3 below is that there is a (suboptimal) equilibrium which cannot be selectively eliminated, yet BM has "no bite" because the equilibrium involves incompatible reports. Secondly, the two conditions differ with regard to the quantifiers in inequalities (10) and (7), respectively. SE requires that Agent i should not deviate to the "flag" for any possible $\theta_i \in \Theta_i$. In contrast, BM requires this only for those θ_i that are compatible with $\alpha_{-i}(\bar{\theta}_{-i})$. The following result establishes that if one confines attention to IC revelation mechanisms (a property that is in any case necessary for implementation), then *SE* is always at least as strong a condition as BM.

Theorem 4.2. (i) *If a revelation mechanism satisfies IC and SE relative to F, then it satisfies BM relative to F.*

(ii) *Suppose each agent has exclusive information, i.e.* $\Theta^* = \Theta$. *Then if a revelation mechanism satisfies BM relative to F, it satisfies SE relative to F.*

9. The Palfrey-Srivastava (1989) definition of Bayesian monotonicity is stronger than that of Palfrey-Srivastava (1987a) and Postlewaite-Schmeidler (1986). The two definitions coincide in the case of "non-exclusive" information structures, which are assumed throughout those two papers.

10. Palfrey-Srivastava (1989) require that *every* $f \in F$ satisfies BM, as they are concerned with full implementation rather than implementation in the sense of Definition 2.1. See Appendix B for details.

Proof. (i) Suppose α is a reporting strategy in Λ_f, with $\alpha(\theta)\in\Theta^*$ for any $\theta\in\Theta^*$, and $(f\circ\alpha)\notin F$. Consider first the case where α is an equilibrium in Λ_f. Then SE implies that there exists Agent i, $\bar{\theta}_i\in\Theta_i$ and $h:\Theta_{-i}\to X$ such that conditions (6) and (7) are satisfied. Define $t:\Theta\to X$ by $t(\theta_{-i},\theta_i)=h(\theta_{-i})$ for all $\theta_i\in\Theta_i$. It then follows that (9) and (10) hold.

Next, consider the case where α is not an equilibrium in Λ_f. Then there exists agent i, and $\bar{\theta}_i$, $\hat{\theta}_i$ in Θ_i such that

$$\textstyle\sum_{\theta_{-i}} q_i(\theta_{-i}\,|\,\bar{\theta}_i)[U_i(f(\alpha_{-i}(\theta_{-i}),\alpha_i(\bar{\theta}_i))\,|\,\theta_{-i},\bar{\theta}_i)-U_i(f(\alpha_{-i}(\theta_{-i}),\hat{\theta}_i)\,|\,\theta_{-i},\bar{\theta}_i)]<0. \tag{11}$$

Further, by hypothesis Λ_f is IC, so

$$\textstyle\sum_{\theta_{-i}} q_i(\theta_{-i}\,|\,\theta_i)[U_i(f(\theta_{-i},\theta_i)\,|\,\theta)-U_i(f(\theta_{-i},\hat{\theta}_i)\,|\,\theta)]\geqq 0 \quad \text{for all } \hat{\theta}_i \text{ in } \Theta_i. \tag{12}$$

Define $t:\Theta\to X$ by $t(\theta)=f(\theta_{-i},\hat{\theta}_i)$. It then follows that (11) implies (9), and (12) implies (10).

(ii) If $\Theta^*=\Theta$, then condition BM considers all reporting strategies α in Λ_f, including equilibrium ones. Further (10) holds for all $\theta_i\in\Theta_i$. Defining $h:\Theta_{-i}\to X$ by $h(\theta_{-i})=t(\theta_{-i},\alpha_i(\bar{\theta}_i))$, it follows that BM implies SE. ||

Part (ii) of the above theorem says that subject to incentive compatibility the two conditions are equivalent in the case of completely exclusive information, an assumption that appears reasonable in a variety of incomplete information models. The following example shows that with complete information (the polar opposite of completely exclusive information) SE may be more demanding than BM.

Example 4.3. Suppose there are two agents with complete information about the actual environment. Bayesian equilibrium thus reduces to Nash equilibrium. Let $\Theta=\{\theta,\bar{\theta}\}$ and $X=\{a,b,c\}$.

Agents' preferences in state θ are as follows:

$$U_1(b\,|\,\theta)>U_1(c\,|\,\theta)=U_1(a\,|\,\theta)$$
$$U_2(a\,|\,\theta)=U_2(c\,|\,\theta)>U_2(b\,|\,\theta).$$

In state $\bar{\theta}$, the two agents' preferences are:

$$U_1(b\,|\,\bar{\theta})>U_1(c\,|\,\bar{\theta})>U_1(a\,|\,\bar{\theta})$$
$$U_2(a\,|\,\bar{\theta})>U_2(b\,|\,\bar{\theta})>U_2(c\,|\,\bar{\theta}).$$

The social choice rule F is single valued and selects: $F(\theta)=\{a\}$ and $F(\bar{\theta})=\{b\}$. We first argue that any revelation mechanism Λ_f with $f\in F$ (with $f(\theta,\theta)=a$ and $f(\bar{\theta},\bar{\theta})=b$) does satisfy Bayesian monotonicity. Let $\alpha=(\alpha_1,\alpha_2)$ be a reporting rule such that $(f\circ\alpha)\notin F$. BM only requires us to consider compatible reporting rules, i.e. satisfying $\alpha(\theta)\in\Theta^*$ for all $\theta\in\Theta^*$. Hence, either $\alpha_1(\theta)=\alpha_2(\theta)=\bar{\theta}$, or $\alpha_1(\bar{\theta})=\alpha_2(\bar{\theta})=\theta$, or both. If $\alpha_1(\theta)=\alpha_2(\theta)=\bar{\theta}$, then $(f\circ\alpha)(\theta)=b\notin F(\theta)$. We may then choose the function $t(\cdot)$ to satisfy $t(\bar{\theta},\bar{\theta})=c$. The fact that Agent 2 prefers $t(\alpha(\theta))=c$ to $f(\alpha(\theta))=b$ in state θ ensures that (9) is met. Furthermore inequality (10) holds since $U_2(b\,|\,\bar{\theta})>U_2(c\,|\,\bar{\theta})$. A similar argument can be made for the case where $\alpha_1(\bar{\theta})=\alpha_2(\bar{\theta})=\theta$.

Consider now the revelation mechanism Λ_f with $f\in F$ selecting $f(\theta,\bar{\theta})=b$ and $f(\bar{\theta},\theta)=c$. It is readily checked that the revelation mechanism Λ_f is IC. Further, Λ_f has a suboptimal equilibrium $\alpha(\cdot)$, where $\alpha_1(\theta)=\bar{\theta}$, $\alpha_2(\theta)=\theta$ and $\alpha_1(\bar{\theta})=\alpha_2(\bar{\theta})=\bar{\theta}$. Note that this reporting rule involves incompatible reports and therefore BM does not apply. We claim that this suboptimal equilibrium cannot be selectively eliminated. Agent 2 cannot be given a flag since he is getting his top ranked alternative in state θ. If Agent 1 is given a flag then the outcome corresponding to (FL, θ) would have to be alternative b, since this is the only alternative that Agent 1 ranks above c in state θ. As a consequence, however, truthful reporting ceases to be an equilibrium for Agent 1 in θ. Hence, we have constructed an example of an (incentive compatible) revelation mechanism satisfying BM but not SE.

The work of Mookherjee and Reichelstein (1989) shows that two features of Example 4.3 are crucial: there are only two agents, and the mechanism designer cannot punish the agents severely for incompatible reports, i.e. there is no bad outcome.[11] In fact, one can show (see Mookherjee and Reichelstein (1989)) that if agents have complete information, BM implies SE provided either (i) there are at least three agents, or (ii) there exists a bad outcome. The following example, though, shows that with intermediate information structures (information is neither complete nor completely exclusive) SE may be a stronger condition than BM even when there are three or more agents.

Example 4.4. The construction employed here is similar to that in Example 4.2. We begin with two agents and include additional agents later on. The set of social outcomes is again $X = \{a, b, c\}$. Let $\Theta_i = \{\theta_i, \bar{\theta}_i\}$ and suppose that Θ^* consists of three possible states with equal prior probability: $s_1 \equiv (\theta_1, \theta_2)$, $s_2 \equiv (\theta_1, \bar{\theta}_2)$ and $s_3 = (\bar{\theta}_1, \bar{\theta}_2)$. In contrast, the state $s_4 \equiv (\bar{\theta}_1, \theta_2)$ has prior probability zero. The two agents are assumed to have the following prefence orderings:

$$U_1(b \mid s_1) > U_1(c \mid s_1) = U_1(a \mid s_1), \qquad U_2(a \mid s_1) > U_2(c \mid s_1) > U_2(b \mid s_1)$$

$$U_1(b \mid s_2) > U_1(c \mid s_2) = U_1(a \mid s_2), \qquad U_2(c \mid s_2) > U_2(a \mid s_2) > U_2(b \mid s_2)$$

$$U_1(b \mid s_3) > U_1(c \mid s_3) > U_1(a \mid s_3), \qquad U_2(a \mid s_3) > U_2(b \mid s_3) > U_2(c \mid s_3).$$

The SCC F is single valued and selects $F(s_1) = \{a\}$, $F(s_2) = F(s_3) = \{b\}$. We impose further conditions on preferences so that the revelation function f with $f(s_1) = a, f(s_2) = f(s_3) = b$ and $f(s_4) = c$ is the only incentive compatible revelation function that agrees with F. The reader may check that this will be case if

$$U_2(b \mid s_3) - U_2(c \mid s_3) \geqq U_2(a \mid s_2) - U_2(b \mid s_2).$$

Further, if $U_2(c \mid s_2) - U_2(b \mid s_2) \geqq U_2(b \mid s_3) - U_2(c \mid s_3)$, then the corresponding revelation mechanism Λ_f has a suboptimal equilibrium in which Agent 1 always reports $\bar{\theta}_1$ and Agent 2 always reports θ_2 (with c as the resulting outcome). By an argument similar to that in Example 4.2 it can be verified that this suboptimal equilibrium cannot be selectively eliminated provided that the value $U_1(b \mid s_1)$ and $U_2(a \mid s_3)$ are sufficiently large relative to the other utility values.

It remains to demonstrate that Λ_f does satisfy BM. As shown in the proof of part (i) in Theorem 4.2 IC implies that the requirements of BM will be satisfied for any compatible reporting strategies that do not constitute an equilibrium. Hence, it is sufficient to show that there is no equilibrium α in Λ_f satisfying $(f \circ \alpha) \not\subseteq F$ and $\alpha(\theta) \in \Theta^*$ for all $\theta \in \Theta^*$. The reader may check that this is indeed the case by going through the following steps. First, there is no compatible equilibrium with $\alpha_1(\theta_1) = \bar{\theta}_1$. The compatibility requirement would imply that $\alpha_2(\theta_2) = \alpha_2(\bar{\theta}_2) = \bar{\theta}_2$ whenever $\alpha_1(\theta_1) = \bar{\theta}_1$. However, there

11. Moore and Repullo (1989) have developed a strengthening of the monotonicity condition that, in conjunction with a weak version of the no-veto-power condition, is necessary and sufficient for Nash implementation with two agents. In Mookherjee and Reichelstein (1989) we show that IC and SE are exactly equivalent to this strengthened condition.

cannot be such an equilibrium since Agent 2 would prefer message θ_2 (resulting in either outcome c or a), whenever his information is θ_2. Secondly, since $\alpha_1(\theta_1)=\theta_1$ in any compatible equilibrium, it follows directly that $\alpha_1(\bar{\theta}_1)=\bar{\theta}_1$. Finally, if Agent 1 reports truthfully, the only possible equilibrium is for Agent 2 to report truthfully as well, proving our claim.

To extend this example to situations involving more than two agents, we may simply add "dummy" agents. These agents have no information, i.e. their Θ_i's are singletons, and they are indifferent among all outcomes regardless of the prevailing state. Clearly, Bayesian monotonicity will continue to hold, while it remains impossible to satisfy SE, since no dummy agent can be used to destroy the suboptimal equilibrium. In summary, this example shows that for general information structures, where information is neither complete nor exclusive, SE will be a stronger condition than BM.

5. IMPLEMENTATION IN UNDOMINATED STRATEGIES

Earlier sections have shown that incentive compatibility, by itself, is not sufficient for implementation. In addition, the selective elimination condition has to be satisfied. Examples constructed by Palfrey-Srivastava (1987a) suggest that many economically interesting social choice rules do not satisfy Bayesian monotonicity (and, a fortiori, the SE condition). These negative results naturally lead to the question of whether a larger class of social choice rules become implementable if one were to use a refinement of Bayesian equilibrium as the solution concept. This seems plausible since, for any given mechanism, the set of admissible equilibria can only shrink if a stronger equilibrium concept is used.

A first step in this direction was taken by Moore-Repullo (1988) who analyzed implementation in subgame perfect equilibria for complete information environments. In a recent paper Palfrey-Srivastava (1987b) have studied implementation in undominated Bayesian equilibrium strategies, i.e. Bayesian equilibria in which no agent uses weakly dominated messages. They consider a model with completely exclusive information, private values and a mild restriction on preferences (termed value distinguished types). Palfrey-Srivastava show that in such cases any incentive compatible choice rule can be implemented in undominated equilibrium strategies. Thus, the restriction to undominated equilibria makes a substantial difference since Bayesian monotonicity (or the selective elimination condition) is no longer required.

We show in this section that our approach of augmenting revelation mechanisms can be adapted to implementation in undominated equilibrium strategies. As before, some agents will be given flags and counterflags in addition to their type messages. The effect of this augmentation is that the suboptimal equilibria of the original revelation mechanism involve dominated strategies and are therefore excluded from consideration.

Definition 5.1. Given a mechanism $\Lambda=\langle M, g\rangle$, a message $m_i \in M_i$ is *dominated* by message $\bar{m}_i \in M_i$ for type θ_i of Agent i, if $V_i(\alpha_{-i}, m_i \mid \theta_i) \leqq V_i(\alpha_{-i}, \bar{m}_i \mid \theta_i)$ for all possible strategies α_{-i} of other agents, with strict inequality holding for at least one α_{-i}.

A message is said to be undominated for any given agent, if it is not dominated by any other message, for any type of the agent. A strategy α_i is undominated for Agent i if $\alpha_i(\theta_i)$ is undominated for all $\theta_i \in \Theta_i$.

Definition 5.2. Given a mechanism $\Lambda=\langle M, g\rangle$, an undominated equilibrium $(\alpha_1, \ldots, \alpha_n)$ in Λ is a Bayesian equilibrium such that α_i is undominated for all $i \in N$.

Definition 5.3. The mechanism Λ is said to implement $F:\Theta^* \twoheadrightarrow X$ in undominated equilibrium strategies, if there exist undominated equilibria in Λ, and for every undominated equilibrium $\alpha = (\alpha_1, \ldots, \alpha_n) : (g \circ \alpha) \in F$.

The first question to ask is whether the Revelation Principle and the Augmented Revelation Principle extend to implementation in undominated equilibrium strategies. The Revelation Principle would require in this context that if there exists a mechanism Λ with one undominated equilibrium resulting in F (i.e., $(g \circ \alpha) \in F$), then there exists a revelation mechanism Λ_f where truthful reporting is an undominated equilibrium and $f \in F$. It can be shown that this is indeed true in the case of completely exclusive information but not in general.[12] However, the next result shows that the Augmented Revelation Principle always holds, irrespective of the information structure.

Theorem 5.4. *Suppose $F:\Theta^* \twoheadrightarrow X$ is implementable in undominated equilibrium strategies. Then there exists an augmented revelation mechanism that implements F in undominated equilibrium strategies. Furthermore, truth-telling is an undominated equilibrium in this mechanism.*

The proof of this result is left as an exercise to the reader. To simplify the subsequent analysis we assume from here on that information is completely exclusive so that $\Theta^* = \Theta$. We also assume private values, i.e. $U_i(\cdot)$ depends on θ_i only.

Before stating our general result, we illustrate why Bayesian incentive compatibility is essentially sufficient for implementation in undominated equilibrium strategies. For simplicity, consider again the economic model of Section 3 where each agent is allocated a vector $x_i \in \mathscr{R}^l_+$ of privately consumed resources. For the moment we disregard any feasibility constraints. Suppose now that α is a suboptimal equilibrium of the IC revelation mechanism Λ_f with $f \in F$. Without loss of generality, assume that $\alpha_1(\theta_1^1) = \theta_1^{k_1}$, i.e. under the equilibrium α, type 1 of agent 1 reports his last type k_1.

To "eliminate" the suboptimal equilibrium α we augment Λ_f by giving Agent 1 an additional message, again denoted Flag. In contrast to the construction of Section 3, however, type θ_1^1 will not have an incentive to deviate to the Flag. The only purpose of the Flag option is to make $\theta_1^{k_1}$ a dominated message for type θ_1^l of Agent 1. As a consequence, α ceases to be an undominated equilibrium. Note that for the original revelation mechanism Λ_f, incentive compatibility itself implies that the truth is undominated. To ensure that the truth is undominated in the augmented mechanism as well, Agent 2 is also given a masterflag (MFL). Our construction assumes "value distinguished" types (an assumption introduced by Palfrey–Srivastava (1987b)). In the present context this assumption says that there exist $\bar{x}_1, \underline{x}_1 \in \mathscr{R}^l_+$ such that $U_1(\bar{x}_1 \mid \theta_1^1) > U_1(\underline{x}_1 \mid \theta_1^1)$ and $U_1(\underline{x}_1 \mid \theta_1^{k_1}) > U_1(\bar{x}_1 \mid \theta_1^{k_1})$. Table II depicts the payoffs in the augmented mechanism for the special case of two agents.

The vectors δ, ε, η and $\eta - 2\delta$ are (small) vectors in $\mathscr{R}^l_+$ which are all non-zero. By construction of $\underline{x}_1$ and $\bar{x}_1$, FL dominates $\theta_1^{k_1}$, for type θ_1^1. Hence, α is an equilibrium involving dominated strategies. In contrast, truthful reporting remains an undominated strategy for every type of Agent 1 (for type $\theta_1^{k_1}$ this follows from the definition of $\underline{x}_1$ and

12. With exclusive information truthful reporting must be undominated whenever it is an equilibrium. This is because all states that can be reported in the revelation mechanism have positive prior probability. It is straightforward, though, to construct examples where information is complete and truthful reporting will be dominated whenever it constitutes an equilibrium. At the same time the outcomes corresponding to truthful reporting can be implemented by an undominated equilibrium of some indirect mechanism.

TABLE II

Agent 1's Payoffs

	$\theta_2^1 \cdots \theta_2^{k_2}$	$CFL^1 \cdots CFL^{k_2}$	MFL
θ_1^1 $\vdots$ $\theta_1^{k_1}$	$f_1(\theta_1, \theta_2)$	η $\eta-2\delta$	0 $\underline{x}_1$
FL	$f_1(\theta_1^{k_1}, \theta_2)$	$\eta-\delta$	$\bar{x}^1$

Agent 2's Payoffs

	$\theta_2^1 \cdots \theta_2^{k_2}$	$CFL^1 \cdots CFL^{k_2}$	MFL
θ_1^1 $\vdots$ $\theta_1^{k_1}$	$f_2(\theta_1, \theta_2)$	$f_2(\theta_1, \theta_2)-\varepsilon$	0
FL	$f_2(\theta_1^{k_1}, \theta_2)$	$f_2(\theta_1^{k_1}, \theta_2)+\delta$	0

$\bar{x}_1$, and for all other types it follows from the possible use of *CFL*'s). Truthful reporting also remains an equilibrium. Finally, for a proper choice of the vectors ε and δ, there will be no new equilibrium in the augmented mechanism. Note that Agent 2 will never play *MFL* in equilibrium. Agent 2 will use *CFL*'s only if Agent 1 plays *FL* in some instances. If Agent 1 plays *FL* in some instances, Agent 2 will only play *CFL*'s provided the ε-cost is sufficiently small relative to the δ reward. In that case, however, Agent 1 will report θ_1^1 instead of *FL*. Thus, the augmented mechanism possesses no new equilibria.

To state our general result, suppose there are l different private goods to be allocated between n agents in the organization. As before, let $w \in \mathscr{R}^l_{++}$ represent the organization's aggregate resource endowment. For simplicity, we shall not insist on global balance, i.e. the mechanism designer can appropriate resources for off-equilibrium messages.[13]

A8: $C_0 = \{(x_1, \ldots, x_n) \in \mathscr{R}^{n \cdot l} \mid \sum_{i=1}^n x_i \leqq w\}$.

A9: $C_i = \mathscr{R}^l_+$.

Agents' utility functions will again be assumed to be continuous and monotone. In addition, we shall require that agents' types can be "value distinguished".[14]

A10: For all $\theta_i, \bar{\theta}_i \in \Theta_i$, there exist $\bar{x}_i, \underline{x}_i \in \mathscr{R}^l_+$ satisfying $\bar{x}_i \leqq w$, $\underline{x}_i \leqq w$ such that

$$U_i(\bar{x}_i \mid \theta_i) > U_i(\underline{x}_i \mid \theta_i) \quad \text{and} \quad U_i(\underline{x}_i \mid \bar{\theta}_i) > U_i(\bar{x}_i \mid \bar{\theta}_i).$$

Finally, we call a revelation mechanism Λ_f *interior* if $f_i(\theta) \gg 0$ for all $i \in N$, $\theta \in \Theta^*$. Interiority simply ensures that for any allocation given by the function f, the mechanism designer can assign (small) penalties and rewards.

Theorem 5.5. *Suppose* A2 *and* A8–A10 *hold. The SCC* $F: \Theta^* \twoheadrightarrow X$ *can be implemented by a finite mechanism in which truthful reporting is the unique undominated equilibrium, if there exists an incentive compatible, interior revelation mechanism* Λ_f satisfying $f \in F$.

13. The proof of Theorem 5.5 shows that balance can be achieved without any changes in our construction if there are three or more agents.

14. In the absence of BM (or SE), Palfrey and Srivastava (1987b) show this assumption to be a necessary condition for implementation in undominated strategies.

The proof of Theorem 5.5 essentially consists of a repeated application of the construction shown in Table II. At each augmentation stage one non-truthful equilibrium of the original revelation mechanism Λ_f is excluded since it becomes a dominated strategy. After a finite number of augmentations, truthful reporting remains the unique undominated equilibrium. A detailed proof can be found in an earlier version of this paper, Mookherjee and Reichelstein (1988).

Our mechanism differs from that of Palfrey-Srivastava (1987b) in several respects. First, their mechanism requires three or more agents. Second, the message spaces in their mechanism are infinite. An individual agent's message has several components including that agent's type, other agents' types, an allocation and a number belonging to the interval $[0, a)$. Essential to their construction is that this interval is half-open, i.e. the value a is not included. A message involving any positive number is dominated by a message involving a higher number. Their procedure amounts to eliminating a suboptimal equilibrium by letting some equilibrium messages be dominated by other messages which in turn are dominated by other messages and so on. Such procedures lead to "tail-chasing" where agents may have no best undominted responses, a problem avoided by our mechanism. Finally, the mechanism of Palfrey-Srivastava will, in general, possess multiple equilibria all of which result in the same outcome.

In summary, a significantly larger class of choice rules becomes implementable, if one focuses on undominated equilibria. This refinement of Bayesian equilibrium, however, leaves some open questions. Note that the construction in Table II has plausibly eliminated the undesirable equilibrium α through weak dominance. At the same time, we preserved the desired truthful equilibrium in a rather implausible manner. Type $\theta_1^{k_1}$ might well reason that Agent 2 will never choose the dominated message *MFL*. Consequently, type $\theta_1^{k_1}$, might as well choose *FL* instead of reporting truthfully. Thus, the desired truthful equilibrium is vulnerable to a second-order dominance argument. Though this particular problem can be avoided by a modification of our mechanism, we believe that the primary task is to develop a stronger implementation concept. A promising candidate appears to be implementation by equilibria which are immune to iterated dominance.

APPENDIX A

Proof of Theorem 3.7. It suffices to show that if F is uniquely implementable, then there exists an augmented revelation mechanism in which truthful reporting is an equilibrium. In addition, every equilibrium in this augmented revelation mechanism is equivalent to the truth.

Let $\Lambda = \langle M, g\rangle$ uniquely implement F and let $\alpha = (\alpha_1, \ldots, \alpha_n)$ be the unique equilibrium of Λ. Define $f \equiv g \circ \alpha$. Then $f \in F$ and Λ_f is IC. Consider the following augmentation $\bar{\Lambda} = \langle \bar{M}, \bar{g}\rangle$ of Λ_f:

$$\bar{M}_i = \Theta_i \cup T_i \quad \text{with } T_i = \{m_i \in M_i \mid m_i \notin \text{range}(\alpha_i)\}.$$

Define $\bar{g}(\bar{m}) = g(\phi_1(\bar{m}_1) \cdots \phi_n(\bar{m}_n))$ where $\phi_i : \bar{M}_i \to M_i$ is given by:

$$\phi_i(\bar{m}_i) = \begin{cases} \alpha_i(\theta_i) & \text{if } \bar{m}_i = \theta_i \in \Theta_i \\ \bar{m}_i & \text{otherwise,} \end{cases}$$

so that $\bar{m}_i \notin \Theta_i$ implies $\phi_i(\bar{m}_i) \in T_i$.

Suppose $(\bar{\alpha}_1 \cdots \bar{\alpha}_n)$ is an equilibrium of $\bar{\Lambda}$. Then $\bar{\alpha}_i(\theta_i) \in \Theta_i$ for all $i \in N$, $\theta_i \in \Theta_i$. This follows from the fact that $(\phi_1 \circ \bar{\alpha}_1 \cdots \phi_n \circ \bar{\alpha}_n)$ is an equilibrium of the original mechanism Λ. However, Λ has a unique equilibrium and, therefore, $(\phi_i \circ \bar{\alpha}_i)(\theta_i) = \alpha_i(\theta_i) \notin T_i$. By construction of ϕ_i it follows that $\bar{\alpha}_i(\theta_i) \in \Theta_i$.

Furthermore, $(\phi_i \circ \bar{\alpha}_i)(\theta_i) = (\alpha_i \circ \bar{\alpha}_i)(\theta_i)$ or $\alpha_i(\theta_i) = (\alpha_i \circ \bar{\alpha}_i)(\theta_i)$, and

$$\begin{aligned} f(\theta_{-i}, \theta_i) &= g(\alpha_{-i}(\theta_{-i}), \alpha_i(\theta_i)) \\ &= g(\alpha_{-i}(\theta_{-i}), \alpha_i(\bar{\alpha}_i(\theta_i))) \\ &= f(\theta_{-i}, \bar{\alpha}_i(\theta_i)). \end{aligned}$$

This shows that θ_i and $\bar{\alpha}_i(\theta_i)$ are equivalent messages. Hence, if $(\bar{\alpha}_1, \ldots, \bar{\alpha}_n)$ is an equilibrium of $\bar{\Lambda}$, it must be an equilibrium of Λ_f; further $\bar{\alpha}$ is equivalent to the truth-telling equilibrium.||

Proof of Theorem 3.8. The proof proceeds inductively. We start with the revelation mechanism Λ_f, and selectively eliminate one of its suboptimal equilibria by augmenting Λ_f. The resulting augmented mechanism will have no new equilibria. If there are any suboptimal equilibria remaining, we construct another augmentation. At any given stage we begin with an augmented mechanism Λ in which truth-telling is an equilibrium, and all equilibria involve "type" messages only. At the next stage we obtain $\bar{\Lambda}$, an augmentation of Λ, which possesses the above properties and has at least one less suboptimal equilibrium. Since Λ_f can have at most a finite number of equilibria, a finite number of augmentations will achieve the desired mechanism.

We describe a representative stage of this iterative procedure. Suppose we start with $\Lambda = \langle M, g\rangle$, an augmentation of Λ_f with the following properties: $M_i = \Theta_i \cup T_i$, $g(m) \in X$ for all $m \in M$, and for every equilibrium α of Λ, $\alpha_i(\theta_i) \in \Theta_i$, for all $\theta_i \in \Theta_i$. Let α be a suboptimal equilibrium of Λ and suppose again that Agent 1 is the agent designated by condition *SE*. Consider the mechanism $\bar{\Lambda} = \langle \bar{M}, \bar{g}\rangle$ with:

$$\begin{aligned} \bar{M}_1 &= M_1 \cup FL \\ \bar{M}_2 &= M_2 \cup \{CFL^1 \cdots CFL^{k_2}\} \\ \bar{M}_i &= M_i \quad \text{for } i > 2. \end{aligned} \tag{1}$$

Thus, Agent 1 is given a new flag, denoted *FL*, and Agent 2 is given a set of corresponding counterflags. Naturally, outcomes associated with messages in the previous stage are left unchanged. Outcomes associated with the new messages satisfy the following conditions:

(i) If Agent 1 chooses *FL* and all other agents choose type messages, then the outcome is given by $h(\theta_{-1})$, provided by the SE condition corresponding to the equilibrium α. This gives rise to:[15]

$$\begin{aligned} \bar{g}(m) &= g(m) \quad \text{for all } m \in M, \\ \bar{g}(FL, \theta_{-1}) &= h(\theta_{-1}) \quad \text{for all } \theta_{-1} \in \Theta_{-1}. \end{aligned} \tag{2}$$

(ii) If agent 1 chooses the current flag, Agent 2 does not raise any of his current counterflags but some agent is choosing a previous flag or counterflag, then every agent is treated as if Agent 1 had reported θ_1^1 instead of *FL*. The only exception is that Agent 1 is charged ε for choosing *FL*, if Agent 2 is choosing a previous flag or counterflag.

$$\begin{aligned} \bar{g}(FL, \theta_2, m_{-12}) &= g(\theta_1^1, \theta_2, m_{-12}) \text{ for } m_{-12} \notin \Theta_{-12}, \\ \bar{g}_1(FL, t_2, m_{-12}) &= (g_y(\theta_1^1, t_2, m_{-12}), g_{x_1}(\theta_1^1, t_2, m_{-12}) - \varepsilon) \quad \text{for all } t_2 \in T_2, m_{-12} \in M_{-12} \\ \bar{g}_i(FL, t_2, m_{-12}) &= g_i(\theta_1^1, t_2, m_{-12}) \quad \text{for } i > 1. \end{aligned} \tag{3}$$

(iii) If Agent 2 raises a new CFL^j, but Agent 1 does not choose *FL*, we have two cases: (a) If Agent 1 chooses a message $t_1 \in T_1$, then the outcome is as if Agent 2 had announced θ_2^j instead of CFL^j; (b) If Agent 1 chooses $\theta_1 \in \Theta_1$, then Agents 2 through n are treated as if Agent 2 had announced θ_2^j, except that Agent 2 is charged ε. The same principle applies to Agent 1, except that he gets the private transfer $x_1^*(\theta_2^j, m_{-12})$ (to be determined below) if he announces θ_1^1.

$$\begin{aligned} \bar{g}(t_1, CFL^j, m_{-12}) &= g(t_1, \theta_2^j, m_{-12}) \\ \bar{g}_1(\theta_1^1, CFL^j, m_{-12}) &= (g_y(\theta_1^1, \theta_2^j, m_{-12}), x_1^*(\theta_2^j, m_{-12})) \\ \bar{g}_1(\theta_1^k, CFL^j, m_{-12}) &= g_1(\theta_1^k, \theta_2^j, m_{-12}) \quad \text{for } k \neq 1, \\ \bar{g}_2(\theta_1, CFL^j, m_{-12}) &= (g_y(\theta_1, \theta_2^j, m_{-12}), g_{x_2}(\theta_1, \theta_2^j, m_{-12}) - \varepsilon) \\ \bar{g}_i(\theta_1, CFL^j, m_{-12}) &= g_i(\theta_1, \theta_2^j, m_{-12}) \quad \text{for } i > 2. \end{aligned} \tag{4}$$

(iv) Finally, if *FL* meets CFL^j, we apply the same rules as in (ii) except that Agent 2 is now given a reward δ:

$$\begin{aligned} \bar{g}_1(FL, CFL^j, m_{-12}) &= \bar{g}_1(FL, \theta_2^j, m_{-12}) \\ \bar{g}_2(FL, CFL^j, m_{-12}) &= (\bar{g}_y(FL, \theta_2^j, m_{-12}), \bar{g}_{x_2}(FL, \theta_2^j, m_{-12}) + \delta) \\ \bar{g}_i(FL, CFL^j, m_{-12}) &= \bar{g}_i(FL, \theta_2^j, m_{-12}) \quad \text{for } i > 2. \end{aligned} \tag{5}$$

15. We shall use the notation $g(m) = (g_1(m) \cdots g_n(m))$ and $g_i(m) = (g_y(m), g_{x_i}(m))$.

In step (iii), we use assumption A4 to choose $x_1^*(\theta_2^j, m_{-12})$ such that for all $\theta_1 \in \Theta_1$:

$$\textstyle\sum_{\theta_{-1}} q_1(\theta_{-1}) \mid \theta_1)[U_1(g_y(\theta_1^1, \theta_2^j, m_{-12}), x_1^*(\theta_2^j, m_{-12}) \mid \theta) - U_1(g_1(FL, \theta_2^j, m_{-12}) \mid \theta)] > 0 \tag{6a}$$

and

$$x_1^*(\theta_2^j, m_{-12}) > g_1(\theta_1^1, \theta_2^j, m_{-12}). \tag{6b}$$

This completes the specification of the augmented mechanism $\bar{\Lambda}$. The proof is established by the following claims, which state that α is not an equilibrium in $\bar{\Lambda}$, and that there is no equilibrium in $\bar{\Lambda}$ where Agents 1 and 2 choose any of their new message options.

Claim 1. *α is not a equilibrium in $\bar{\Lambda}$.*

Proof. This follows from the fact that $\alpha_i(\theta_i) \in \Theta_i$ for all θ_i and all $i \in N$, and upon applying (2) in conjunction with SE. ||

To establish that there are no new equilibria we introduce the following notation. Let $Q_j(\theta_i \mid \theta_j) \equiv \sum_{\theta_{-i-j}} q_j(\theta_i, \theta_{-i-j} \mid \theta_j)$ denote the probability assigned to type θ_i of Agent i by type θ_j of Agent j. Also, let $\tau \equiv \min\{Q_j(\theta_i \mid \theta_j) \mid i \neq j, Q_j(\theta_i \mid \theta_j) > 0\}$ denote the minimum positive probability assigned by a type of one agent to a type of another. Since the set of types is finite, it is clear that τ is well defined and strictly positive. Since the conditional beliefs of agents are obtained by updating from a common prior distribution, and the marginal probability of any type is strictly positive, it follows that

(a) For any $\theta_1 \in \Theta_1$, there exists $\theta_2 \in \Theta_2$ such that $Q_2(\theta_1 \mid \theta_2) > 0$.

(b) $Q_1(\theta_2 \mid \theta_1) > 0$ if and only if $Q_2(\theta_1 \mid \theta_2) > 0$, for any $\theta_1 \in \Theta_1$ and any $\theta_2 \in \Theta_2$.

Claim 2. *There does not exist an equilibrium in $\bar{\Lambda}$ where any type of Agent 1 chooses FL.*

Proof. Suppose otherwise and let $\Theta_1^* \subseteq \Theta_1$ denote the set of types θ_1 of Agent 1 that choose *FL* in some equilibrium. Let Θ_2^* denote the set of types θ_2 of Agent 2 that assign positive probability to the event that Agent 1's type is in Θ_1^*. Since types in Θ_2^* assign at least probability τ that Agent 1 is choosing *FL*, every type in Θ_2^* will prefer CFL^j to θ_2^j, provided ε is sufficiently small relative to δ. Hence in the given equilibrium, every type $\theta_2 \in \Theta_2^*$ will choose either some CFL^j, or messages in T_2.

Since $Q_1(\theta_2 \mid \theta_1) > 0$ if and only if $Q_2(\theta_1 \mid \theta_2) > 0$, any type $\theta_1 \in \Theta_1^*$ assigns probability one to the event that Agent 2's type is in Θ_2^*, and therefore that Agent 2 is choosing either some CFL^j or messages in T_2. It then follows that every type $\theta_1 \in \Theta_1^*$ will choose θ_1^1 instead of *FL*. For messages in T_2, the argument follows from (3); for CFL^j this is a consequence of (5), and (4) in conjunction with (6a). ||

Claim 3. *There is no equilibrium in $\bar{\Lambda}$ where some type of Agent 2 chooses some CFL^j.*

Proof. Using Claim 2, it suffices to consider any equilibrium $\bar{\beta}$ in $\bar{\Lambda}$ where no type of Agent 1 ever uses *FL*, but there is a non-empty set Θ_2^* of types of Agent 2 choosing some CFL^j. It then follows that given any $\theta_2 \in \Theta_2^*$, every type θ_1 of Agent 1 who is assigned positive probability by θ_2 (i.e. $Q_2(\theta_1 \mid \theta_2) > 0$) must be choosing a message $t_1 \in T_1$. Otherwise, the first and fourth equations in (4) imply that type θ_2 of Agent 2 would be better off reporting θ_2^j instead of CFL^j, and thereby avoid the ε-charge imposed whenever Agent 1 chooses a type message.

We now argue that the strategy-tuple β obtained from $\bar{\beta}$ by replacing the message CFL^j with θ_2^j is an equilibrium in the previous stage mechanism Λ. Since any type of Agent 2 choosing CFL^j in $\bar{\beta}$ assigns probability one to the event that Agent 1 is choosing some "non-type" message in T_1, the first equations of (2) and (4) imply that all types of Agent 2 are playing best responses at β in Λ. The same is true for all agents $i \geqq 3$.

It remains to show that there is no type of Agent 1 that can profitably deviate from β in Λ. The first and third equations of (4) show that announcing any message other than θ_1^1 induces the same payoffs for Agent 1 in β (in Λ) as in $\bar{\beta}$ (in $\bar{\Lambda}$). Equation (4), in conjunction with (6b), and the first equation in (2) show that the payoffs associated with the message θ_1^1 are uniformly lower for Agent 1 in the strategy tuple β, compared to $\bar{\beta}$. For any type of Agent 1 that does not choose θ_1^1 in $\bar{\beta}$ (in $\bar{\Lambda}$), it therefore does not pay to deviate to θ_1^1 in β (in Λ). Finally, consider a type θ_1 of Agent 1 that does choose the message θ_1^1 in $\bar{\beta}$ (in $\bar{\Lambda}$). By the above reasoning, that type θ_1 will assign zero probability to the event that Agent 2 chooses some CFL^j. Hence, the payoffs in β associated with all possible messages in M_1, including θ_1^1, are exactly what they were in $\bar{\beta}$.

In summary, we have established that an equilibrium $\bar{\beta}$ involving some CFL^j's but not FL would give rise to a corresponding equilibrium β in the previous stage mechanism Λ in which θ_2^j is substituted for CFL^j. However, in this equilibrium, some types of Agent 1 must be choosing messages in T_1, a contradiction of our initial hypothesis that every equilibrium in Λ involves type messages only. ||

Proof of Theorem 3.9. The proof is similar to the one of Theorem 3.8. To achieve balance, however, we cannot reward the flag raiser for returning to type messages whenever counterflags are raised. Instead the flag raiser will be punished if his flag meets corresponding counterflags.

A representative stage of the iterative augmentation process is as follows. Suppose $\Lambda = \langle M, g\rangle$ is an augmentation of Λ_f with the properties that $M_i = \Theta_i \cup T_i$, $g_i(m) \in C_i$ for all $m \in M$ and $\sum_{i=1}^n g_i(m) = w$. Further, every equilibrium $(\alpha_1, \dots, \alpha_n)$ in Λ satisfies $\alpha_i(\theta_i) \in \Theta_i$, for all $\theta_i \in \Theta_1$. As in the proof of Theorem 3.8, one agent (say, 1) is given a flag FL while another (say, 2) is given a set of counterflags $(CFL^1, \dots, CFL^{k_2})$. The new augmented mechanism $\bar{\Lambda}$ has an outcome function $\bar{g}$ defined by the following rules:

$$\bar{g}(m) = g(m) \quad \text{for all } m \in M \tag{7}$$

$$\bar{g}(FL, \theta_{-1}) = h(\theta_{-1}) \quad \text{where } h(\cdot) \text{ is given by SE.}$$

$$\bar{g}(FL, \theta_2, m_{-12}) = g(\theta_1^1, \theta_2, m_{-12}) \quad \text{for all } m_{-12} \notin \Theta_{-12} \tag{8}$$

$$\bar{g}_1(FL, t_2, m_{-12}) = g_1(\theta_1^1, t_2, m_{-12}) - \varepsilon \quad \text{for all } m_{-12} \in M_{-12}, t_2 \in T_2$$

$$\bar{g}_2(FL, t_2, m_{-12}) = g_2(\theta_1^1, t_2, m_{-12}) + \varepsilon$$

$$\bar{g}_i(FL, t_2, m_{-12}) = g_i(\theta_1^1, t_2, m_{-12}) \quad \text{for } i > 2.$$

$$\bar{g}(t_1, CFL^j, m_{-12}) = g(t_1, \theta_2^j, m_{-12}) \tag{9}$$

$$\bar{g}_1(\theta_1, CFL^j, m_{-12}) = g_1(\theta_1, \theta_2^j, m_{-12}) + \varepsilon$$

$$\bar{g}_2(\theta_1, CFL^j, m_{-12}) = g_2(\theta_1, \theta_2^j, m_{-12}) - \varepsilon$$

$$\bar{g}_i(\theta_1, CFL^j, m_{-12}) = g_i(\theta_1, \theta_2^j, m_{-12}) \quad \text{for } i > 2.$$

$$\bar{g}_1(FL, CFL^j, m_{-12}) = x_1^*(\theta_2^j, m_{-12}) \tag{10}$$

$$\bar{g}_2(FL, CFL^j, m_{-12}) = g_2(\theta_2^j, m_{-12}) + [g_1(FL, \theta_2^j, m_{-12}) - x_1^*(\theta_2^j, m_{-12})]$$

$$\bar{g}_i(FL, CFL^j, m_{-12}) = g_i(\theta_2^j, m_{-12}) \quad \text{for } i > 2.$$

The outcome function $\bar{g}(\cdot)$ is balanced. The payoff $x_1^*(\theta_2^j, m_{-12})$ is chosen such that $x_1^*(\theta_2^j, m_{-12}) < g_1(\overline{FL}, \theta_2^j, m_{-12})$ and every type of Agent 1 prefers $g_1(\theta_1^1, \theta_2^j, m_{-12})$ to $x_1^*(\theta_2^j, m_{-12})$. The rest of the proof proceeds in a fashion similar to that of Theorem 3.8. ||

APPENDIX B

This appendix lays out the differences between our formulation and that of Postlewaite-Schmeidler (1986, 1987) and Palfrey-Srivastava (1987*a*, 1989). First, these authors model information available to agents as partitions over a set of states of nature, while we assume each agent observes the realization of a random variable. The two formulations turn out to be equivalent. Following Postlewaite-Schmeidler and Palfrey-Srivastava, suppose S denotes the set of possible states of nature and Agent i's information is represented by a partition Π_i of S. If the state of nature is $s \in S$, Agent i observes $E_i(s) \in \Pi_i$. Further, it is assumed that for all $s \in S$, $\bigcap_{i=1}^n E_i(s) = \{s\}$, i.e. if agents were to pool their information, they could unambiguously identify the prevailing state . The equivalence of the two formulations becomes transparent if one identifies Θ^* with S, Π_i with Θ_i and $E_i(s)$ with θ_i, respectively.

Another difference between the two models is that Postlewaite-Schmeidler and Palfrey-Srivastava take the SCC F to be a collection of social choice functions, i.e. $F = \{f_j\}_{j \in J}$ for some arbitrary index set J.[16] The notation $f \in F$ then means that there exists $j \in J$ such that $f_j(\theta) = f(\theta)$ for all $\theta \in \Theta^*$. Accordingly, the implementation requirement is that for every equilibrium α of the mechanism $\Lambda = \langle M, g\rangle$ it is true that $(g \circ \alpha) = f_j$ for some $j \in J$. Obvously, this requirement is far more demanding than the one in Definition 2.1, where $(g \circ \alpha) \in F$ if $(g \circ \alpha)(\theta) \in F(\theta)$ for all $\theta \in \Theta^*$. Nonetheless, our results carry over to this more demanding requirement without modification, if the meaning of $f \in F$ in the selective elimination condition is strengthened correspondingly.

16. We are grateful to Thomas Palfrey for pointing out this difference.

A final difference concerns the implementation concept. In the tradition of earlier work on implementation, Postlewaite-Schmeidler and Palfrey-Srivastava are concerned with full implementation, i.e. for any $f \in F$ there exists an equilibrium α such that $g \circ \alpha = f$. It is readily verified that the Augmented Revelation Principle applies to full implementation as well. The necessary conditions for full implementation become stronger, as IC and SE now have to hold for any $f \in F$. It remains to be seen, however, whether the constructions employed in our sufficiency results can be extended to obtain full implementation.

Acknowledgement. We thank Faruk Gul, Leonid Hurwicz, John Moore, Thomas Palfrey, Sanjay Srivastava and two anonymous referees for their useful comments. The second author gratefully acknowledges financial support from NSF grant IRI-870244.

REFERENCES

DASGUPTA, P., HAMMOND, P. and MASKIN, E. (1979), "The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility", *Review of Economic Studies*, **46**, 185-216.

DEMSKI, J. and SAPPINGTON, D. (1984), "Optimal Incentive Contracts with Multiple Agents", *Journal of Economic Theory*, **44**, 156-167.

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

JACKSON, M. (1988), "Full Bayesian Implementation" (Discussion Paper No. 791, Center for Mathematical Studies, Northwestern University).

MA, C., MOORE, J. and TURNBULL, S. (1988), "Stopping Agents from Cheating", *Journal of Economic Theory*, **46**, 355-372.

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

MASKIN, E. (1986), "The Theory of Implementation in Nash Equilibrium: A Survey", in Hurwicz, L. *et al.* (Eds.), *Social Goals and Social Organization*: *Volume in Memory of Elisha Pazner* (Cambridge: Cambridge University Press).

MOOKHERJEE, D. and REICHELSTEIN S. (1988), "Implementation via Augmented Revelation Mechanisms" (Discussion Paper No. 985, Graduate School of Business, Stanford University).

MOOKHERJEE, D. and REICHELSTEIN, S. (1989), "The Revelation Approach to Nash Implementation", (mimeo).

MOORE, J. and REPULLO, R. (1988a), "Nash Implementation—A Full Characterization" (ICERD Discussion Paper, London School of Economics).

MOORE, J. and REPULLO, R. (1988b), "Subgame Perfect Implementation", *Econometrica*, **46**, 1191-1220.

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

PALFREY, T. AND SRIVASTAVA, S. (1987a), "On Bayesian Implementable Allocations", *Review of Economic Studies*, **54**, 193-208.

PALFREY, T. and SRIVASTAVA, S. (1987b), "Mechanism Design with Incomplete Information: A Solution to the Implementation Problem" (mimeo).

PALFREY, T. and SRIVASTAVA, S. (1989), "Implementation with Incomplete Information in Exchange Economies", *Econometrica*, **57**, 115-134.

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

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

REPULLO, R. (1986), "The Revelation Principle under Complete and Incomplete Information", in Binmore, K. and Dasgupta, P. (eds.) *Economic Organization as Games* (Oxford: Basil Blackwell).