Reprinted from Advances in Economic T
edited by Werner Hildenbrand

CHAPTER 2

The theory of incentives: an overview

Jean-Jacques Laffont and Eric Maskin

The theory of incentives is concerned with the problem that a planner (alternatively called a designer, principal, or government, depending on context[1]) faces when his own objectives do not coincide with those of the members of society[2] (whom we shall call agents). This lack of coincidence of goals distinguishes incentives theory from the theory of teams (Marschak and Radner, 1972), which postulates identical objectives, but which otherwise shares many features with our subject. In turn, the assumption that the planner, often the surrogate for society itself, has well-defined objectives separates incentives theory from most of social choice theory, which, since Arrow (1951), examines the possibility of *deriving* social objectives from those of individual preferences.

For an incentive problem to arise, noncoincidence of goals is not enough; the planner must care about either what agents know or what they do. That is, his objective function must depend either on agents' *information* or on their *behavior*.

An example of pure informational dependence is provided by the literature on resource allocation mechanisms. There, the planner's objective – social welfare – is a function of consumers' (agents') preferences and endowments. The incentive problem is, typically, that of eliciting this information.

Pure behavioral dependence is exhibited by an employee-employer relationship in which the employer is interested only in the employee's output. In this case, incentives pertain not to revealing what the employee knows but to inducing him to work hard. Of course, incentive problems typically involve both kinds of dependence.

The planner pursues his objectives by the choice of an incentive

14.147
Spring 1983

scheme,[3] a rule that specifies, in advance, the planner's behavior on the basis of his perceptions of agents' information and actions.[4]

This choice is nontrivial if either (1) some of the agents' payoff-relevant information is not known, a priori, to the planner or else (2) the planner cannot observe agents' actions perfectly. (If the planner both knew all relevant information and could precisely monitor actions, he could presumably force agents to take the optimal action based on this information by promising dire consequences otherwise.)

The first difficulty is frequently called the problem of adverse selection. It is not necessary that the planner's own objective function depend on agents' information - as in the allocation literature - for adverse selection to arise; it is enough that agents' payoffs should. Consider, for example, a monopolist wishing to maximize his (expected) revenue using a nonlinear price schedule. Although his revenue function does not depend directly on consumers' taste parameters - only on their demand - such information is obviously relevant to his choice of schedule. The monopolist's problem is, indeed, a prime example of pure adverse selection.

At the other extreme is the employer-employee relationship mentioned before. Imagine that the employee possesses no information not known to the employer. Suppose, furthermore, that the employee's output depends stochastically on his effort, which is unobservable by the employer. Then, the employer faces a problem of the second kind - a *moral hazard* problem. His failure to induce the "optimal" effort level by the agent derives solely from his limited ability to monitor this effort.

The theories of adverse selection and moral hazard are similar, but there are some important differences between them. It is useful, therefore, to keep them separate conceptually, as we do in Section 2.

The planner's choice of an incentive scheme entails a double maximization: He chooses that scheme which maximizes his (typically, expected) payoff subject to the constraint that, given this scheme, agents will maximize their own objective functions: In many contexts agents must be guaranteed a minimal expected payoff to induce them to participate in the scheme at all. In such cases the planner must maximize subject to the additional constraints that agents obtain these minimal levels. The planner is, therefore, the "leader" in a two-move game; his move consists of selecting a scheme.

What it means for an agent to "maximize his objective function" may be complex if there are other agents, for his payoff then depends on their responses to the planner's scheme as well his own. Thus, with more than one agent, an incentive scheme induces a *game* among the agents, and the planner optimizes subject to the agents' being in *equilibrium*. Of course, to say what an equilibrium is, one must specify an equilibrium concept. Even restricting to noncooperative (noncollusive) behavior by agents, at least four solution concepts (not including their refinements) figure prominently in the incentives literature. We shall have more to say about them later.

We have informally indicated the subject matter of incentives theory in general terms. Of course, questions of incentives are rarely analyzed at this level of generality. Nonetheless, it may be helpful to have a framework within which to relate the disparate pieces of the large incentives literature. To this purpose, we present a formal model in Section 1 and discuss certain modeling difficulties. In Section 2 we show how the literature consists of studies of special cases of this model. Finally, in Section 3, we analyze as an illustrative example a particular, but much studied, special case - a model of public project selection.

1 A general framework

1.1 *The model*

We consider a model consisting of a planner and n agents (indexed $i=1, \ldots, n$). Each agent i has private information represented by $\theta^i \in \Theta^i$. On the basis of this information, he sends a message $m^i \in M^i$ to the planner. The planner replies to these messages with response $r \in R$. The agent then chooses action $a^i \in A^i$. The planner cannot, in general, observe a^i directly but observes the outcome $y^i \in Y^i$ of a^i, θ^i, and his own response r, where y^i is, in general, the value of the realization of a random function $\tilde{y}^i(a^i, \theta^i, r)$. Finally, the planner selects decision $d \in D$.

An *incentive scheme* is a choice by the planner of message spaces $M^1, \ldots, M^n$ (the other spaces, A^i, R, Y^i, and D are exogenous), response function $\rho: M \to R$, and decision function $\delta: M \times Y \to D$, where $M = \Pi M^i$ and $Y = \Pi Y^i$. Thus we can represent an incentive scheme by (M, ρ, δ). For reasons discussed in Section 1.4, efficiency will ordinarily be improved if the planner can take ρ and δ to be random functions. Thus, we shall often write an incentive scheme as $(M, \tilde{\rho}, \tilde{\delta})$ where the tildes indicate possible randomness.

This is not the most general model of incentives that one could imagine, but it is sufficiently broad to accommodate virtually all work on incentives to date.[5] To see how the elements of this model fit together, suppose that agents are production units and that the planner wishes to allocate capital efficiently across these units. Each agent i produces output from capital and labor according to the production process θ^i, known, ex ante, only to him. The planner asks each agent to provide data about his process. Thus m^i consists of possible messages that i could send about his production technology. Based on this data, the planner allocates capital across units. Thus r is an allocation, and ρ an allocation rule. Given his capital, agent i then chooses a quantity of labor a^i. Capital, labor, the production process, and, perhaps, nature combine to produce output $\tilde{y}^i(a^i, \theta^i, r)$. Finally, production units are rewarded by the planner according to the rule δ on the basis of their output and the information they provided.

For another, quite different, example that illustrates the model well, suppose that the planner is an insurance company that insures agents against accidents. Based on the message m^i he provides about his accident-proneness θ^i, agent i is offered an insurance policy r. Whether or not he has an accident (the value of y^i) depends (randomly) on his accident-proneness and the level a^i of preventive care he chooses. (Notice that, in this example, y^i does not depend directly on r.) His ultimate compensation, d, depends on y^i and his policy. Because the policy itself depends on the information he provides, we can write δ as a function of m^i directly.

We shall suppose that agent i's payoff depends on his information θ^i, his action a^i, and the planner's decision d. We shall represent his preferences by the von Neumann-Morgenstern utility function

$$u^i : D \times A^i \times \Theta^i \to \mathbf{R}$$

Agent i's behavior presumably depends on θ^i. Thus, given θ^i, we can represent his behavior by the strategy $\sigma^i(\theta^i) = (\mu^i(\theta^i), \alpha^i(\theta^i, \cdot))$, where $\mu^i(\theta^i) \in M^i$ and, for all r, $\alpha^i(\theta^i, r) \in A^i$. (Throughout this chapter, we shall ignore the possibility that agents might use random (mixed) strategies.) The agent's *contingent strategy* or *strategy rule* is given by the function $\sigma^i(\cdot)$. If agent i is the sole agent, he will choose $\mu^i(\theta^i)$ and $\alpha^i(\theta^i, \cdot)$ to maximize the expected value of $u_i(\cdot, \cdot, \theta^i)$, where we speak of *expected* value, because r, d, and y^i may be random. With more than one agent, an incentive scheme is a genuine game; agent i's payoff will, in general, depend, through r and d, on the strategies of others. Thus, his choice of strategy will ordinarily depend on how he believes others behave. In other words, in addition to the "objective" uncertainty associated with the possible randomness of r, d, and y^i, the agent may face "strategic" uncertainty: uncertainty about others' strategies. Notice that he would face this uncertainty even if he knew the values of others' parameters θ^{-i}. The fact that he might not know these values merely compounds his problem.

There are several alternative hypotheses (drawn from game theory) in the incentives literature about how an agent might act under strategic uncertainty. These hypotheses are embodied in alternative solution concepts. That is, a solution concept implicitly prescribes a way of resolving strategic uncertainty.

We shall discuss four often-studied solution concepts later. For the time being, we observe that for a specified solution concept, the planner's problem is to choose an incentive scheme whose equilibrium maximizes his expected payoff. We assume that the planner's preferences depend on his decision d, the vector of outcomes y, and agents' information $\theta = (\theta^1, \ldots, \theta^n)$.[6] His preferences are represented by the von Neumann-Morgenstern utility function

$$v : D \times Y \times \Theta \rightarrow \mathbf{R}$$

We can think of the planner as the Stackelberg leader in a two-move game. First, he moves by choosing an incentive scheme; then, everyone else reacts to that scheme.

Unfortunately, the preceding phrase, "choose an incentive scheme whose equilibrium maximizes," may not have a well-defined meaning. For a given incentive scheme, and relative to a specific solution concept, there may be no equilibrium, or there may be several. The former possibility poses no great conceptual difficulty; the planner can simply confine his attention to those schemes that have an equilibrium. Moreover, the latter may not be especially troubling in the case of a single agent. If an agent has multiple optimal strategies, it may not be overly heroic to suppose that he chooses the one (or among the ones) that the planner prefers. At any rate, that is what the literature, for the most part, assumes. With more than one agent, however, agents will not, in general, be indifferent among multiple equilibria. Therefore, for the planner to count on a particular equilibrium arising may be unwarranted; agents who prefer another may thwart him. The issue of multiple equilibria has not been uniformly satisfactorily resolved throughout the incentives literature for more than one agent. As we shall see below, it remains, for certain solution concepts in particular, an important difficulty.

In contrast with the optimal incentive scheme – the incentive scheme that maximizes the planner's (expected) payoff – is the *full optimum*, which consists, in addition to response and decision functions, of the strategy rules that the planner, were he permitted, would *impose* on the agents. Because in the full optimum there is no informational problem, we can take the joint message space $M^1 \times \cdots \times M^n$ to be $\Theta^1 \times \cdots \times \Theta^n$. Formally, (θ, r^*, d^*), together with functions

$$a^{*i} : \Theta^i \times R \rightarrow A^i, \quad i = 1, \ldots, n$$

is a full optimum with respect to the prior distribution $F(\theta)$ if it solves the problem

$$\max_{r(\cdot),d(\cdot),a^i(\cdot)} E_\theta v(d,y,\theta)$$

subject to

$$Eu^i(d(\theta,\tilde{y}),a^i(\theta^i,r(\theta)),\theta^i) \geqslant \bar{\mathbf{u}}^i, \quad i=1,\ldots,n$$

for some choice of $\bar{u}^1,\ldots,\bar{u}^n$. The $\bar{u}^i$'s can be interpreted as the "minimal expected payoffs" mentioned in the introduction.

There are two reasons, mentioned in the introduction, why an optimal incentive scheme may not be a full optimum. One is that the value of θ may not be known to the planner a priori. This is the problem of adverse selection. The other is that y^i may depend on a^i. This is moral hazard. We shall have more to say about these two problems later.

1.2 *Solution concepts*

In this subsection we discuss some of the more widely used solution concepts in the incentives literature. Several others will be mentioned in Sections 2 and 3. We divide solution concepts into three categories: those that can be defined without reference to the information that agents possess about one another (e.g., equilibrium in dominant and maximin strategies); those that require the vector θ of informational parameters to be drawn from a joint probability distribution (viz., Bayesian equilibrium); and those that, in effect, assume complete information (e.g., Nash equilibrium).

By far the strongest, but in several ways the least controversial, solution concept is that of equilibrium in dominant strategies. A dominant strategy is a strategy that an agent, given his information, is willing to use regardless of what he believes others know and the way he believes others behave. Formally, $(\mu^i(\theta^i),\ \alpha^i(\theta^i,\cdot))$ is a *dominant strategy* for agent i with information θ^i in incentive scheme $(M,\tilde{\rho},\tilde{\delta})$ if for any choice $(m^{-i},a^{-i}(\cdot))$ of strategies by other agents (where a^{-i} is such that $a^{-i}:R\rightarrow A^{-i}$), $(m^i,a^i(\cdot))=(\mu^i(\theta^i),\ \alpha^i(\theta^i,\cdot))$ maximizes

$$(1.1) \qquad Eu^i(\tilde{d}(m,\tilde{y}(a(\tilde{r}(m)),\theta)),\quad a^i(\tilde{r}(m)),\theta^i)$$

for all θ^{-i}, where the expectation is taken with respect to $\tilde{d}$, $\tilde{y}$, and $\tilde{r}$. The strategy rules $\sigma=(\sigma^1,\ldots,\sigma^n)$ are in dominant strategy equilibrium if for all i and all θ^i, $\sigma^i(\theta^i)=(\mu^i(\theta^i),\ \alpha^i(\theta^i,\cdot))$ is a dominant strategy for agent i with information θ^i.

To suppose that if agents have dominant strategies they will play them is an appealing behavioral postulate, because it assumes very little about agents. It is a weak assumption in three senses. First, it does not specify what beliefs an agent has about others' information. Second, it does not ascribe to an agent any particular theory of how others behave (i.e., how they choose *their* strategies). Third, it does dictate how the agent resolves his strategic uncertainty: The maximaxer, the maximiner, and the Bayesian will all play their dominant strategies, if they have them.

The principal limitation of the dominant strategy solution concept for the planner (apart from its neglect of possible collusion by agents) is the difficulty of designing incentive schemes whose dominant strategy equilibria generate a satisfactory payoff for the designer. The Gibbard-Satterthwaite theorem (Gibbard, 1973; Satterthwaite, 1975) gives an indication of this difficulty. It asserts that, in the case where A and Y are null (so that an incentive scheme is given by $\delta : M \rightarrow D$), if for each i and each ordering of D there exists $\theta^i \in \Theta^i$ such that $u^i(\cdot, \theta^i)$ corresponds to that ordering, then the only incentive schemes $\delta : M \rightarrow D$ for which $\delta(M)$ contains at least three elements and a dominant strategy equilibrium exists are dictatorships, that is, schemes in which there exists an agent j such that for all $\bar{d} \in \delta(M)$ there exists $\bar{m}^j \in M^j$ with $\delta(\bar{m}^j, m^{-j}) = \bar{d}$ for all m^{-j}. Such an agent j is called a dictator because of his complete power to have his own way. If the planner's objective function v reflects the preferences of agents at all democratically, it is clear that a dictatorial incentive scheme will not go very far toward the maximization of the planner's expected payoff.

Despite the negativism of the Gibbard-Satterthwaite result, satisfactory dominant strategy incentive schemes do exist in some models. In Section 3, we shall study one such model in considerable detail.

Maximin strategies, like dominant strategies, implicitly ascribe to an agent neither a theory of what others know nor a theory of how they behave. Maximin equilibrium, however, imposes a very strong method for resolving strategic uncertainty: namely, extreme pessimism. Formally, $(\mu^i(\theta^i), \alpha^i(\theta^i, \cdot))$ is a maximin strategy for agent i with information θ^i if $(m^i, a^i(\cdot)) = (\mu^i(\theta^i), \alpha^i(\theta^i, \cdot))$ maximizes

$$(1.2) \qquad \min_{m^{-i}, a^{-i}, \theta^{-i}} Eu^i(\tilde{d}(m, \tilde{y}(a(\tilde{r}(m)), \theta)), a^i(\tilde{r}(m)), \theta^i)$$

where the expectation is taken with respect to $\tilde{d}$, $\tilde{y}$, and $\tilde{r}$. The strategy rule vector $\sigma(\cdot) = (\sigma^1(\cdot), \ldots, \sigma^n(\cdot))$ is a maximin equilibrium if for all i and θ^i, $\sigma^i(\theta^i)$ is a maximin strategy for agent i with information θ^i.

The shortcomings (except in two-person zero-sum games) of maximin strategies as a plausible hypothesis for strategic behavior are well known. Even accepting the maximin hypothesis, moreover, one cannot typically make very accurate predictions about the outcome of a game. This is because, in many games, virtually *any* strategy maximizes equation (1.2), so that almost anything can be an equilibrium.

In contrast with dominant and maximin strategies, the Bayesian solution concept (Harsanyi, 1967) is defined explicitly in terms of an agent's beliefs about others. Assume that each agent i believes that θ is drawn from a joint probability distribution G^i (not all agents need have the same G^i). Suppose, furthermore, that agent i believes that agent j uses the strategy rule $\bar{\sigma}^j(\cdot)$. The strategy rule vector $\bar{\sigma}(\cdot) = (\bar{\sigma}^1(\cdot), \ldots, \bar{\sigma}^n(\cdot))$ is a Bayesian equilibrium as long as for all i and θ^i, $\sigma^i(\theta^i) = \bar{\sigma}^i(\theta^i)$ maximizes

$$(1.3) \qquad \int_{\theta^{-i}} Eu^i(\sigma^i(\theta^i), \bar{\sigma}^{-i}(\theta^{-i}))\, dG^i(\theta^i, \theta^{-i})$$

where $Eu^i(\sigma^i(\theta^i), \bar{\sigma}^{-i}(\theta^{-i}))$ is shorthand for agent i's expected utility when strategies are $(\sigma^i(\theta^i), \bar{\sigma}^{-i}(\theta^{-i}))$. Of course, an equilibrium may not exist in general unless agents can play random strategies, but we will not worry about this problem.

One objection to this definition of a Bayesian equilibrium is that it does not explain why agent i believes that others use the strategy rule $\bar{\sigma}^{-i}(\cdot)$. In a conventional Nash equilibrium (in a game where players know all relevant information about each other) a player can calculate other players' equilibrium strategies. In the Bayesian setting, to predict that agent j will use the strategy rule $\bar{\sigma}^j$, one must attribute to him not only probabilistic beliefs about θ but also beliefs about others' beliefs about θ, beliefs about beliefs about beliefs, etc. That is, there is an infinite and increasingly complex sequence of attributions of beliefs. Moreover, besides creating a very complicated problem for agent i, this formulation serves only to push back the unexplained hypotheses one step, from the level of behavior to the level of belief.

One case in which this complexity is avoided is when probabilistic beliefs about θ arise as the result of some common experience – say a public pronouncement about the distribution of θ.[7] In that case, everyone knows the distribution, everyone knows that everyone knows it, and so on. That is, the distribution of θ is common knowledge. Of course, that knowledge of this kind is common is by no means an innocuous assumption, but its enormously simplifying implications have led to its almost universal adoption in the literature on Bayesian incentives.

We turn finally to complete information and Nash equilibrium. The incentives literature employing Nash equilibrium has typically not *formally* modeled the information that agents have about others. That is, θ^i

is generally taken to embody only data about agent i's *own* preferences, endowment, etc. In this case a vector of strategies (not strategy-rules) $(\bar{\sigma}^1(\theta^1), \ldots, \bar{\sigma}^n(\theta^n))$ is an equilibrium for agents with parameters $(\theta^1, \ldots, \theta^n)$ if for all i, $\sigma^i(\theta^i) = \bar{\sigma}^i(\theta^i)$ maximizes

$$Eu^i(\sigma^i(\theta^i), \bar{\sigma}^{-i}(\theta^{-i}))$$

where the expectation is taken with respect to $\tilde{d}$, $\tilde{y}$, and $\tilde{r}$. This approach has the defect that we can no longer speak of an equilibrium vector of strategy *rules*, contrary to our approach so far; if θ^{-i} changes, $\bar{\sigma}^i(\theta^i)$ is no longer, in general, a Nash equilibrium strategy. Thus an approach more in keeping with the rest of this chapter is to let θ^i incorporate *all* of agent i's information. In fact, we can write

$$\theta^i = (\theta^i_i, \theta^i_{-i})$$

where θ^i_i can be interpreted as agent i's information about his own preferences, etc., whereas θ^i_{-i} represents his information about others. The assumption of complete information can then be stated formally as

$$(1.4) \qquad \theta^i_k = \theta^j_k \quad \text{for all } i, j, \text{ and } k$$

On the basis of equation (1.4) we can define a Nash equilibrium of strategy rules. Indeed, formulated in this way, a Nash equilibrium is merely a special case of a Bayesian equilibrium, where the "specialness" is embodied in (1.4). The reason why the literatures on Bayesian and Nash incentives have evolved separately is that work in the former area has typically assumed that θ^i's are distributed independently, whereas the latter approach, as (1.4) indicates, requires not only that the θ^i's be perfectly correlated but coincident.

It may seem strange to model behavior by a solution concept of complete information when, so often in the incentives literature, the very lack of information is the central problem. For example, in many models of public goods allocation only the absence of information about preferences of consumers for these goods potentially prevents a full optimum from being attained. One might argue that if consumers have complete information about each others' preferences, then the planner should have this knowledge too. But if so, he can simply propose the optimum *ab initio*, avoiding the design of an incentive scheme altogether.

Nonetheless, there are at least two distinct justifications for the Nash equilibrium approach. First, the approach makes sense in many situations in which the "planner" is ficti ious (or a surrogate for the collection of agents themselves) and the m‿thod of making collective decisions (the incentive scheme) must be determined well in advance of the decision making itself. For example, in democratic societies, the allocation of resources to public goods is not imposed by an omniscient planner but is decided by legislative methods fixed long before people's (or their representatives') preferences for any particular public good are known. Nonetheless, by the time that representatives actually decide on a particular allocation, they may well have a good idea about what each others' preferences are or (and what will often suffice) at least what the distribution of preferences is. Thus Nash equilibrium (or one of its refinements) may not be too bad a way to model behavior.

The other justification for Nash equilibrium is quite different and relies on viewing an equilibrium as a stationary point of some kind of (usually implicit) adjustment process. The idea is that at each stage of the process, an agent either responds explicitly to the others' current strategies by modifying his own or, ignorant perhaps of what others are doing, "experiments" with his strategy and modifies it according to his experimental success or failure. In either case, strategy revision ceases (a stationary point is reached) when the current strategies form a Nash equilibrium, because only then will agents find further (unilateral) deviation undesirable. Thus Nash equilibrium is the appropriate concept to predict the outcome, even though agents may not have complete information.

Neither of these justifications is entirely satisfactory. The first rationale has the virtue of being consistent game-theoretically. Indeed, if agents do, in fact, have complete information, Nash equilibrium seems virtually the only way to model (noncooperative) behavior. It has, however, the drawback of limited applicability. There are simply many situations where supposing that agents have complete information does not make sense.

The second rationale would appear to apply to situations regardless of agents' information, but it does not cohere so well formally. On the one hand, if agents react merely to the current strategies of others, then there is a strong element of myopia in their behavior. Why do they not foresee the reactions that their own deviations induce in others? Moreover, if they are ignorant of others' preferences, they should presumably attempt to draw inferences about these from others' behavior. On the other hand, if agents merely "experiment" without directly observing the behavior of others, they may have trouble disentangling the effects of their own experiments from those of others.[8]

1.3 *Direct revelation*

The four solution concepts we have considered share the property that, with some qualification, the only incentive schemes a planner need consider are those where the message spaces are of the form

$$M^i = \Theta^i$$

and each agent's equilibrium message is his true parameter. That is, an agent's message is equivalent to an element θ^i of his information space, and in equilibrium he truthfully reveals his information. That one can restrict attention to message spaces of this form has been observed by many, including Harris and Townsend (1981), Gibbard (1973), and Green and Laffont (1977) and has been called the idea of direct revelation (Dasgupta, Hammond, and Maskin, 1979) or the revelation principle (Myerson, 1979).

Its explanation is straightforward. If (M, ρ, δ) is an incentive scheme and $\sigma(\cdot) = (\sigma^1(\cdot), \ldots, \sigma^n(\cdot))$, where $\sigma^j(\cdot) = (\mu^j(\cdot), \alpha^j(\cdot))$ is a corresponding equilibrium, define

$$\bar{M}^j = \Theta^j$$

$$\bar{\delta} : \bar{M} \times Y \to D$$

$$\bar{\rho} : \bar{M} \to R$$

$$\bar{\mu}^j : \Theta^j \to \bar{M}^j$$

$$\bar{\alpha}^j : \Theta^j \times R \to A^j$$

so that

$$\bar{\delta}(\theta, y) = \delta(\mu^1(\theta^1), \ldots, \mu^n(\theta^n), y)$$

$$\bar{\rho}(\theta) = \rho(\mu^1(\theta^1), \ldots, \mu^n(\theta^n))$$

$$\bar{\mu}^j(\theta^j) = \theta^j$$

$$\bar{\alpha}^j(\theta^j, r) = \alpha^j(\mu^j(\theta^j), r)$$

Then, it is immediate to verify that for dominant Bayesian and Nash strategies the strategy rule vector $(\bar{\sigma}^1, \ldots, \bar{\sigma}^n)$ is an equilibrium for the incentive scheme $(\Theta, \bar{\rho}, \bar{\delta})$. Thus, if (M, ρ, δ) is an "optimal" incentive scheme, so is $(\Theta, \bar{\rho}, \bar{\delta})$. For maximin equilibrium, this argument does not quite work, because in moving from (M, ρ, δ) to $(\Theta, \bar{\rho}, \bar{\delta})$ we change the joint strategy space and change the domain over which the minimization (1.2) is performed. Thus, although $\sigma^j(\cdot)$ may be a maximin strategy rule in the former scheme, $\bar{\sigma}^j(\cdot)$ may not be maximin in the latter. This difficulty can, to some extent, be avoided by changing the definition of a maximin equilibrium so that agents' minimizations are performed only over those strategies that could be maximin strategies for others. (For the details on this kind of *restricted* maximin equilibrium, see the work of Dasgupta, Hammond, and Maskin, 1979, pp. 207–9.)

There may also be a problem with direct revelation schemes in the case of dominant strategies. Here the problem is not that $\bar{\sigma}^j(\cdot)$ might fail to be a dominant strategy rule in $(\Theta, \bar{\rho}, \bar{\delta})$ but that even if (M, ρ, δ) has a unique equilibrium (or, alternatively, all equilibria generate the same expected payoff for the planner), $(\Theta, \bar{\rho}, \bar{\delta})$ may have multiple equilibria, some of which give the planner a lower payoff.[9]

Still, one feels intuitively that, at least in incentive problems that are sufficiently nondegenerate, the problem of multiple equilibria with dominant strategies should not be terribly severe in direct revelation schemes. This rather vague intuition can be expressed formally in several different ways. First, it can easily be shown that when agents' preference orderings are strict (i.e., when for all θ^i the indifference sets corresponding to $u^i(\cdot, \cdot, \theta^i)$ are singletons), there exists at most one dominant strategy equilibrium outcome (the equilibrium outcome consists of the planner's decision and the agents' actions) (Dasgupta, Hammond, and Maskin, 1979, p. 196) for each profile $(\theta^1, \ldots, \theta^n)$. Second, it is obvious that if, as in some problems, agents' preferences are strictly convex in their own strategies, they cannot have more than one dominant strategy. Third, suppose that changing θ^i or a^i changes agent i's preferences over D. One way of capturing the idea that the incentive scheme is not degenerate is to suppose that by varying m^i (holding $a^i(\cdot)$ and other players' strategies fixed) agent i can make the incentive scheme trace out a subset of D with the same dimension as the hyperplane tangent to a point of agent i's indifference surface in D and, furthermore, that as *other* agents' strategies vary (holding $a^i(\cdot)$ and m^i fixed) the incentive scheme traces out all of D. Formally, suppose that all spaces have suitable topologies and that the u^j's are analytic functions such that, for all θ^j, $\theta^{j\prime}$, a^j, and $a^{j\prime}$

$$u^j(\cdot, a^j, \theta^j) = u^j(\cdot, a^{j\prime}, \theta^{j\prime})$$

if and only if $a^j = a^{j\prime}$ and $\theta^j = \theta^{j\prime}$. Suppose that (Θ, ρ, δ) is a (differentiable) direct revelation scheme with a truthful dominant strategy equilibrium $\sigma(\cdot) = (\sigma^1(\cdot), \ldots, \sigma^n(\cdot))$. (By truthful, we mean that $\sigma^j(\theta^j) = (\mu^j(\theta^j), \alpha^j(\theta^j))$, where $\mu^j(\theta^j) = \theta^j$ for all j and θ^j.) Assume that y^j (differentiable) does not depend on a^j and that R is null, so that we can ignore ρ. Because δ depends on θ through ρ, σ, and y, we may write δ as a function of θ directly, where $\delta(\theta)$ is the decision if all agents use their truthful dominant strategies. For $d \in D$, $a^j \in A^j$, and θ, let $D^j(d, a^j, \theta^j)$ be the hyperplane tangent to agent j's indifference surface (for utility function $u^j(\cdot, \cdot, \theta^j)$) at (d, a^j). (That such a hyperplane exists at every point implies that indifference surfaces are not thick.) Let $C^j(\delta, \theta)$ be the linear space spanned by the derivative of $\delta(\theta)$ with respect to θ^j, and let $C^{-j}(\delta, \theta)$ be the space spanned by the derivatives of δ with respect to all θ^i's other than θ^j. Because $\sigma^j(\theta)$ is a dominant strategy for agent j with parameter θ^j,

$$C^j(\delta, \theta) \subseteq D^j(\delta(\theta), \alpha^j(\theta^j), \theta^j)$$

The result is Proposition 1.1.

Proposition 1.1: In the formulation of the preceding paragraph, if for all j and θ

$$C^j(\delta, \theta) = D^j(\delta(\theta), \alpha^j(\theta^j), \theta^j) \quad \text{and} \quad C^{-j}(\delta, \theta) = D$$

then if $(\sigma^1(\cdot), \ldots, \sigma^n(\cdot))$ is a dominant strategy equilibrium for the direct revelation incentive scheme (Θ, δ), only truthful strategies are dominant.

Remark: The condition $C^j = D^j$ amounts to requiring that, by varying m^j, agent j can trace out a subset of D with the same dimension as that of the hyperplane tanget to his indifference surface. The condition $C^{-j} = D$ means that the set of outcomes obtained by varying m^{-j} locally looks like D.

Proof: Suppose that besides $\sigma^i(\theta^i)$, $\sigma^{i\prime} = (m^{i\prime}, a^{i\prime})$ is dominant for agent i with parameter θ^i. Because the scheme is direct revelation, $m^{i\prime} = \bar{\theta}^i$ for some $\bar{\theta}^i \in \Theta^i$. Then,

$$(1.5) \quad \frac{\partial u^i}{\partial d}(\delta(\bar{\theta}^i, \theta^{-i}), a^{i\prime}, \theta^i) \cdot \frac{\partial \delta(\bar{\theta}^i, \theta^{-i})}{\partial \theta^i} = 0 \quad \text{for all } \theta^{-i}$$

Because $\sigma^i(\bar{\theta}^i)$ is dominant for agent i with parameter $\bar{\theta}^i$,

$$(1.6) \quad \frac{\partial u^i}{\partial d}(\delta(\bar{\theta}^i, \theta^{-i}), \alpha^i(\bar{\theta}^i), \bar{\theta}^i) \cdot \frac{\partial \delta}{\partial \theta^i}(\bar{\theta}^i, \theta^{-i}) = 0 \quad \text{for all } \theta^{-i}$$

Because $C^i(\delta, \bar{\theta}^i, \theta^{-i}) = D^i(\delta(\bar{\theta}^i, \theta^{-i}), \alpha^i(\bar{\theta}^i), \theta^i)$ by hypothesis and because of (1.5) and (1.6), the vector $\partial u^i / \partial d(\delta(\bar{\theta}^i, \theta^{-i}), a^{i\prime}, \theta^i)$ is a scalar multiple of $\partial u^i / \partial d(\delta(\bar{\theta}^i, \theta^{-i}), \alpha^i(\bar{\theta}^i), \bar{\theta}^i)$ for all θ^{-i}. For some $\bar{\theta}^{-i}$, let $\bar{d} = \delta(\bar{\theta}^i, \bar{\theta}^{-i})$. Because $C^{-i}(\delta, \bar{\theta}^i, \bar{\theta}^{-i}) = D$, $\delta(\bar{\theta}^i, \cdot)$ is (locally) onto D in a neighborhood of $\bar{\theta}^{-i}$, there exists a neighborhood N of $\bar{d}$ such that $\partial u^i / \partial d(d, \alpha^i(\bar{\theta}^i), \bar{\theta}^i)$ is a scalar multiple of $\partial u^i / \partial d(d, \alpha^{i\prime}, \theta^i)$ for all $d \in N$. Thus the ordering corresponding to $u^i(\cdot, \alpha^i(\bar{\theta}^i), \bar{\theta}^i)$ coincides with that of $u^i(\cdot, a^{i\prime}, \theta^i)$ when restricted to N. Because u^i is analytic, these orderings coincide for all of D. But then, from hypothesis,

$$(\alpha^i(\bar{\theta}^i), \bar{\theta}^i) = (a^{i\prime}, \theta^i) \quad \text{and so} \quad \bar{\theta}^i = \theta^i \qquad \text{Q.E.D.}$$

With Bayesian equilibrium, multiple equilibria do not create difficulty in converting an incentive scheme to an equivalent direct revelation scheme; the set of equilibria in the original scheme is isomorphic to that in the direct scheme. However, multiple equilibria are a more general problem for Bayesian equilibrium. Very little is known about the circumstances under which a scheme has unique equilibrium or, alternatively, all equilibrium outcomes are equivalent. Indeed, we will show in Section 3 that in a simple public goods model - much studied in the literature - there is a whole continuum of Bayesian equilibria, almost all "bad."

The issue of multiple equilibria is important too for Nash equilibrium as a solution concept. But, in this case, the existing literature has dealt with it. Usually the approach has been not to design schemes for which equilibrium is unique - indeed, with Nash equilibrium that is often impossible - but rather (e.g., Groves and Ledyard, 1977; Maskin, 1977; Hurwicz, 1979*a*; Schmeidler, 1980) to ensure that all Nash equilibria are equally desirable. A characterization of when such insurance is possible (in the case where the A^i's are null) is given by Maskin (1977).

1.4 *Random incentive schemes*

In our general incentives model we allow for the possibility that δ and α may be stochastic. There are two reasons why the planner may wish to make them stochastic.

The first is that the spaces D and R may not be convex. Randomization simply permits the planner to convexify these spaces. A recent example where this kind of randomization figures prominently is the problem of auctioning an indivisible object (Myerson, 1978; Maskin and Riley, 1980*a*). Here a nonconvexity is created by the constraint that the seller (planner) can assign the object to at most one bidder. (That is, D is nonconvex.) The seller may therefore wish to randomize among certain bidders to determine the winner in order to overcome the nonconvexity. Another model in which D is inherently nonconvex is the model of discrete public investment studied in Section 3 of this chapter.

The second rationale for randomization is, formally, that the constraints of the planner's maximization problem will not in general be convex. The planner maximizes subject to agents' maximizing as well. If $\bar{\sigma}^i(\theta^i)$ is a maximizing strategy for agent i, then

$$(1.7)\qquad Eu^i(\bar{\sigma}^i(\theta^i)) \geqslant Eu^i(m^i, a^i(\cdot))$$

for any alternative strategy choice $(m^i, a^i(\cdot))$ (where we have, for convenience, written i's utility as a function of his strategy directly and have omitted other agents' strategies). If u^i is linear in strategies, then (1.7) represents a convex constraint. But if utility is strictly concave (if the agent is risk-averse), then the set generated by (1.7) will in general not be convex (because a concave function appears on both sides of the inequality).

This argument constitutes a prima facie case for randomization when agents are risk-averse. On closer examination (Maskin, 1980*b*) it turns out that this kind of randomization is quite generally useful as a screening device in models of adverse selection, but it is desirable only under rather restrictive (and often implausible) assumptions in models of pure moral hazard.

2 The incentive literature

In this section we quickly review the incentives literature to illustrate how work in this field fits neatly into the framework of Section 1. We do not attempt, however, to survey the literature exhaustively. Our greatest emphasis is on work about implementation and resource allocation.

2.1 *Models of adverse selection*

We begin by discussing models in which adverse selection (the inability of the planner to observe agents' information) prevents the attainment of a full optimum. The simplest variety of adverse selection model is one in which agents do nothing but transmit messages. That is, A^i is null.

Pure message transmission. A substantial part of incentives theory consists of models of pure message transmission. These include the implementation, allocation mechanism, nonlinear-pricing, and auction-design literatures.

In the implementation literature, the planner represents society. His objectives are embodied in a correspondence

$$f:\Theta \rightarrow D$$

where Θ^i typically consists of the possible preference orderings over D that agent i can have. For any profile θ, $f(\theta)$ consists of the "welfare-optimal" or "best" decisions. In the notation of Section 1, the planner's objectives can be expressed as

$$v:D \times \Theta \rightarrow \mathbf{R}$$

where

$$\begin{aligned} v(d,\theta) &= 1, \quad d \in f(\theta) \\ &= 0, \quad \text{otherwise} \end{aligned}$$

The implementation problem is to find an incentive scheme (M, δ) (we can ignore ρ) such that for each θ the set of equilibria (with respect to a given solution concept) coincides with, or is a subset of, $f(\theta)$. (See the work of Dasgupta, Hammond, and Maskin, 1979, for more detail.) If such a scheme exists, f is said to be implementable.

The implementation literature subdivides according to solution concepts. In addition to the four solution concepts discussed in Section 1, there are numerous variants.

The basic (negative) result for dominant strategies (see Section 1) is due to Gibbard (1973) and Satterthwaite (1975). Related results are due to Pattanaik (1975), Barberá (1977*a*), Gärdenfors (1976), Kelly (1977), and others. These results are proved for unrestricted domains. Extensions to restricted domains, showing the connection between implementability of a correspondence and the existence of a social welfare function satisfying Arrow's conditions, have been studied by Maskin (1976) and Kalai and Müller (1977). (For more on restricted domains, see the work of Pattanaik and Sengupta, 1977, Moulin, 1980*b*.) In particular, it is shown (Dasgupta, Hammond, and Maskin, 1979) that if the correspondence f is generated by an Arrow social welfare function (i.e., $f(\theta)$ represents the top-ranked elements in the social ordering), then it can be implemented not only in dominant strategies but in *coalitionally* dominant strategies. That is, the formation of collusive coalitions does not change the set of equilibria.

The implementability of a single-valued correspondence f in dominant strategies is equivalent to f's satisfying "independent person-by-person monotonicity." IPM asserts that if $a \in f(\theta)$ and a is strictly preferred to b under $\theta^{i'}$, then $b \notin f(\theta^{i'}, \theta^{-i})$. The conditions for implementability in coalitionally dominant strategies are stronger (independent weak monotonicity), but the two kinds of implementability are equivalent when preferences are strict (indifference is ruled out) and preferences are sufficiently "rich."[10]

There is a recent literature on dominant strategy implementation when preferences and incentive schemes are differentiable. Contributions include those of Chichilnisky and Heal (1980*a*, 1980*b*) (which also consider Nash equilibrium) and Satterthwaite and Sonnenschein (1981).

Another line consists of studying *random* incentive schemes (i.e., schemes where δ is a stochastic function of m). Gibbard (1977) has shown that by allowing the scheme to be stochastic, but otherwise preserving the hypotheses of the Gibbard-Satterthwaite theorem, one enlarges the set of implementable correspondences (here allowing for *random* correspondences) to include those that are lotteries over dictatorships and those correspondences with a range of at most two elements. For related results, see the work of Barberá (1977*b*).

There is practically no analysis of maximin equilibrium at the most general level of implementation theory. There is also relatively little on Bayesian equilibrium; however, see the work of Myerson (1979) and Rosenthal (1979). On the other hand, the literature on Nash equilibrium and its variants is large. Hurwicz and Schmeidler (1978) have studied the possibility of constructing incentive schemes whose Nash equilibria are Pareto-optimal when message spaces have the cardinality of preference spaces. Maskin (1977) enlarged the message spaces and showed that any correspondence that is monotonic (see Section 3.4) and satisfies a weak nonveto property[11] is Nash-implementable. In particular, the Pareto correspondence (the correspondence that selects all Pareto optima) is implementable for any domain of preferences.

There is an intimate connection between Nash and dominant strategy implementation. K. Roberts (1979*a*) demonstrated that with unrestricted domain, the only single-valued Nash-implementable correspondences are dictatorial (see also Pattanaik, 1976). This corresponds to the Gibbard-Satterthwaite theorem for dominant strategies. Dasgupta, Hammond, and Maskin (1979) extended this result by showing that any single-valued Nash-implementable correspondence is implementable if the domain of strategies is rich. This means that, at least for rich domains, one does not extend the set of implementable correspondences by weakening the solution concept from dominant strategies to Nash, if single-valuedness is maintained.

Much work has been done on implementation with variants of Nash equilibrium. Moulin (1979, 1980*a*) studied implementation by successive elimination of dominated strategies (dominance solvability). This solution concept, due to Farquharson (1969), is closely related to the notion of perfect equilibrium proposed by Selten (1975). In particular, Moulin was able to show that in contrast to the results for Nash equilibrium, a large class of single-valued correspondences are implementable even for an unrestricted domain of preferences.

Another variant of Nash equilibrium is the strong equilibrium, in which equilibrium strategies are compared not just with the deviations of single agents but with those of coalitions. Contributions to the theory of strong implementation include the work of Moulin and Peleg (1980), Maskin (1979*a*), and Kalai, Postlewaite, and Roberts (1977).

One difficulty with incentive schemes that implement correspondences in strong equilibrium is that they typically have many ordinary (i.e., noncooperative) Nash equilibria in addition to their strong equilibria. These Nash equilibria, moreover, may well not be elements of $f(\theta)$. That is, to adopt strong equilibrium as a solution concept is not just to allow for the collusion of coalitions but to *insist* on it; without collusion, equilibrium may not be in $f(\theta)$. To accommodate a planner's uncertainty about the collusiveness of agents, Maskin (1979*b*) proposed the concept of *double implementation*, in which the Nash and strong equilibria of an incentive scheme coincide. Because of this coincidence, it makes no difference which coalitions, if any, form; the set of equilibria is always the same.

A line of research related to implementation derives from the work of Peleg (1978*a*). For a scheme to be *consistent* with f, Peleg, in effect, required that at least one of its strong equilibria be in $f(\theta)$. If $f(\theta)$ is a subset of the strong equilibria of the scheme, then, in the terminology of Sengupta (1979), f is *partially implemented.* In addition to these articles by Peleg and Sengupta, work on consistency and partial implementation includes that of Peleg (1978*b*) and Dutta and Pattanaik (1978).

The literature on incentives in *resource allocation* closely resembles that on implementation, but it deals with more structured models. In particular, D becomes the space of possible allocations of goods across agents rather than just an abstract decision space, and the correspondence f becomes an allocation rule. Moreover, preferences are restricted. As before, the subject subdivides according to solution concept.

In his pioneering article, Hurwicz (1972) showed that in a pure exchange economy of *private* goods, no Pareto-optimal, individually rational[12] allocation rule is implementable in *dominant strategies* when the preference domains include at least the Cobb-Douglas family. This result has been extended by Satterthwaite (1976), Satterthwaite and Sonnenschein (1981), and Dasgupta, Hammond, and Maskin (1979) by dropping the hypothesis of individual rationality and substituting the conclusion that only dictatorial allocation rules are implementable.

Most work on dominant strategy allocation rules, however, concerns public goods. Indeed, most of it assumes that agents' preferences for a public good and private good take the form

$$(2.1) \qquad u^i(x,t) = v^i(x) + t$$

where x is the level of the public good and t is a transfer of the private good. Virtually all research has been concerned with successful allocation rules, rules that, given $v^1(\cdot),\ldots,v^n(\cdot))$, choose the public-good level to maximize

$$\sum_{i=1}^{n} v^i(x)$$

In three seminal articles, Groves (1973), Clarke (1971), and Smets (1972) independently demonstrated the existence of *successful* allocation rules that are implementable in dominant strategies. Groves showed that a successful allocation rule is implementable if its transfers take the form

$$t^i(v^1(\cdot),\ldots,v^n(\cdot)) = \sum_{j\neq i} v^j(d) + h^i(v^{-i}(\cdot)) \tag{2.2}$$

where $x=d$ maximizes $\Sigma_{i=1}^n v^i(x)$,

$$v^{-i}(\cdot) = (v^1(\cdot),\ldots,v^{i-1}(\cdot),v^{i+1}(\cdot),\ldots,v^n(\cdot))$$

and h^i is an arbitrary function of $v^{-i}(\cdot)$. Let us call the set of successful rules whose transfers satisfy (2.2) the Groves class. Clarke (1971) exhibited the particularly interesting member of the class in which

$$h^i(v^{-i}(\cdot)) = -\sum_{j\neq i} v^j(d^i)$$

where $x=d^i$ maximizes $\Sigma_{j\neq i} v^j(x)$. This is often callen the "pivotal" mechanism because only agents who change the public-good level from what it would be without them get transfers (which are, in fact, negative). The second price auction of Vickrey (1961) is, in fact, the private-good analogue of this rule. Smets (1972) examined the Groves rule in which $h^i\equiv 0$.

Green and Laffont (1977) established that the Groves class coincides with the set of all implementable allocation rules when the domain of possible valuation functions $v^i(\cdot)$ is unrestricted. A monograph by Green and Laffont (1979*a*) provides a detailed analysis of the properties of the Groves class. Green and Laffont (1976) and Hurwicz (1975) established, in particular, that no member of the Groves class has transfers that balance (sum to zero identically). Walker (1980) generalized this result. That no member of the Groves class is immune from manipulation by coalitions has been demonstrated by Green and Laffont (1979*b*) and Bennett and Conn (1977).

There is a small literature on interesting restrictions of the domain of valuation functions. Groves and Loeb (1975) examined quadratic valuations and showed that, for this domain, balancing the transfers is possible. Laffont and Maskin (1980*a*) studied successful and implementable allocation rules when valuation functions are differentiable and are parametrized by θ^i ranging in an open interval of the real line. In this framework, the proof that the set of such allocation rules coincides with the Groves class is virtually immediate, amounting merely to integrating a partial differential equation. (Indeed, in the differentiable setting, the whole question of implementability boils down to the integrability of systems of partial differential equations.) Holmstrom (1979*b*) demonstrated, however, that this characterization depends crucially on the domain of θ^i being smoothly connected. He showed that without this assumption, there are successful and implementable rules outside the Groves class. Laffont and Maskin (1980*a*) also showed that, with differentiability, questions about transfer balance and manipulation by coalitions are easy to handle. An illustration of the power of the "differentiable approach" is given in Section 3 of this chapter.

Another line of work concerns nonsuccessful allocation rules. K. Roberts (1979) showed that when the space of valuation functions is unrestricted, any implementable rule must choose the public-good level to maximize

$$\sum_{i=1} \lambda^i v^i(x) + K(x)$$

where $\lambda^i \geqslant 0$ and K is an arbitrary function. Laffont and Maskin (1980*b*) placed the further restrictions on valuation functions of differentiability and concavity and showed that any member of the class of implementable and neutral (treating all public-good levels symmetrically) rules must take the public-good level to satisfy

$$h\left(\frac{dv^1}{dx}(x), \ldots, \frac{dv^n}{dx}(x)\right) = 0$$

where $h : \mathbf{R}^n \rightarrow \mathbf{R}$ is continuous and semi-strictly increasing. In Section 3 of this chapter we characterize all (piecewise differentiable) implementable allocation rules when the public-good level is restricted to the values 0 and 1 (although we allow for randomization as well).

Finally, instead of working with severely restricted preferences, Roberts and Postlewaite (1976), Hammond (1979), and Mas-Colell (1978) examined dominant strategies in economies with many agents. Roberts and Postlewaite showed that in the limit, price-taking behavior becomes a dominant strategy as a pure exchange economy grows. Hammond studied an economy with a continuum of agents and demonstrated that implementable and Pareto-optimal allocation rules must be competitive. Similarly, Mas-Colell showed that an implementable allocation rule satisfying convexity, anonymity, nondegeneracy, and neutrality properties is necessarily competitive.

The literature on maximin equilibrium and resource allocation is considerably smaller than that for dominant strategies. Dubins (1974) exhibited a balanced allocation rule that can be implemented in maximin equilibrium when utility functions take the form of equation (2.1). Green and Laffont (1979*a*, Chapter 7) showed that this rule is not individually rational (does not guarantee agents at least the utility of their initial endowments) and that it encounters difficulties when consumption sets are bounded from below. They constructed a generalized rule that is individually rational on average and argued that this modified Dubins rule is the static analogue of the Malinvaud-Drèze-de la Vallée

Poussin planning procedure (see the discussion of planning procedures that follows). Thomson (1979*a*) characterized all maximin implementable allocation rules when preferences are of the form (2.1) and also in the 0–1 project case.

Using the Harsanyi (1967) concept of Bayesian equilibrium (see Section 1), d'Aspremont and Gérard-Varet (1979) showed that there exist Bayesian implementable allocation rules that are successful and for which transfers balance when preferences take the form of equation (2.1) and when the joint probability distribution of valuation functions is common knowledge, with $v^i(\cdot)$ distributed independently of $v^{-i}(\cdot)$. Arrow (1979) offered a similar analysis. Laffont and Maskin (1979*a*) characterized the class of Bayesian implementable successful rules and demonstrated its close connection with the Groves class. They also extended the d'Aspremont-Gérard-Varet results to the case where the v^i's are "negatively correlated." Unfortunately, these analyses examined only "truthful" equilibria in which agents, in effect, reveal their true preferences. The possibility of untruthful equilibria was not considered. That untruthful equilibria are likely to exist in profusion is discussed in Section 3. Other work on Bayesian incentives and resource allocation includes that of Harris and Townsend (1981) and Ledyard (1977), both general discussions of the issues and concepts involved in resource allocation with incomplete information.

Groves and Ledyard (1977) inspired much of the literature on incentives and Nash equilibrium. They developed an incentive scheme that, for any number of private and public goods, and for a domain of preferences restricted little more than by the "classical" assumption of convexity, monotonicity, and continuity, has the feature that all its Nash equilibrium[13] outcomes are Pareto-efficient. (There may, however, be difficulties with the existence of equilibrium unless preferences are restricted rather more. See the work of Green and Laffont, 1979*a*, Chapter 7.) Schmeidler (1980) exhibited an incentive scheme for a pure exchange economy of private goods whose Nash equilibria coincide with the Walrasian (competitive) equilibria when preferences are classical. However, the scheme violates both individual and aggregate feasibility constraints out of equilibrium. Hurwicz (1979*a*) devised a scheme whose Nash equilibria coincide with the Lindahl equilibria of a classical economy, but again the scheme may be aggregately infeasible when in disequilibrium. These infeasibilities, moreover, are necessary; feasible implementation in Nash equilibrium of the Lindahl and Walras correspondences is impossible. The difficulty is, as shown by Postlewaite, that these correspondences are not monotonic at the boundary of the feasible set as Nash implementation requires (Maskin, 1977). However, as Hurwicz, Postlewaite, and Maskin (1979) demonstrated, the *constrained* Walrasian and Lindahl correspondences are Nash implementable (the constrained correspondences include, in addition to ordinary Walrasian and Lindahl allocations, the allocations obtained by constraining an individual's demand from exceeding the total endowment of the economy). Moreover, as shown by Hurwicz (1979*a*), the constrained Walrasian and Lindahl correspondences are the *smallest* continuous, Pareto-optimal, and indivually rational correspondences that are Nash implementable if the domain includes all classical preferences. That is, any other such correspondence must include all constrained Walrasian allocations (for a private-good economy) or all constrained Lindahl allocations (for an economy with public goods). Other work on Nash implementation in resource allocation includes that of Hurwicz (1975), Thomson (1980), Walker (1977), and Wilson (1978).

The incentive schemes mentioned so far for allocation of resources have been "one-shot" games: Agents report their messages, on the basis of which the planner chooses an allocation (although, as discussed in Section 1, Nash equilibrium is sometimes viewed as a stationary point in an adjustment process). An alternative approach, pioneered by Drèze and de la Vallée Poussin (1971), Malinvaud (1972), and Tideman (1972), is to allocate through a dynamic incentive scheme. In the three articles cited, each agent consumer reports his marginal rates of substitution between a public good and private good at each instant of time. The planner uses this information to alter the public-good level and to make transfers of private good. Over time, the allocation converges to a Pareto optimum. Moreover, along the way, the utility of each consumer continually increases (i.e., the procedure is individually rational). Champsaur (1976) showed that the class of such "MDP" procedures is "neutral" or "unbiased"; that is, any individually rational Pareto optimum is the limit point of a member of this class (see the work of Champsaur, Drèze, and Henry, 1977, for a comprehensive study of the stability and existence of solutions in these procedures).

One important question about MDP procedures is the incentive for truthful reporting of marginal rates of substitution. Drèze and de la Vallée Poussin (1971) showed that truthful revelation is a local maximin strategy (i.e., maximizes the minimum instantaneous payoff, the instantaneous payoff here being the gradient of utility) and consequently also globally maximin (maximizes the minimum utility of the final allocation). They also observed that at the stopping point of a process, revelation of true marginal rates of substitution forms a Nash equilibrium. Malinvaud (1971) suggested that MDP procedures will converge even if agents "lie" along the way. Indeed, J. Roberts (1979) proved that if, at each instant, consumers report their Nash equilibrium strategies of the local revelation game (by local revelation games we mean that consumers report so as to maximize the instantaneous increase of utility), Nash equilibrium is unique, but the equilibrium strategies are untruthful except at the stopping point.[14] The procedure still converges to a Pareto optimum, although at a slower speed than under truthful revelation (see also Henry, 1977).

As discussed in Section 1, modeling consumers' behavior by Nash equilibrium implicitly entails one of two alternative assumptions. Either one assumes that consumers know one another's preferences and so can directly calculate the Nash equilibrium, or one supposes that equilibrium is itself reached through an iterative adjustment procedure. Both assumptions have unappealing features. In particular, the second leads to a double infinity – an infinity of adjustments in the local Nash game and another in the MDP procedure itself. Schoumaker (1977) and Henry (1977) attempted to disentangle this double infinity by studying discrete-time versions of the MDP procedure.

One way of avoiding both assumptions is to devise a procedure ensuring that at each instant truthful revelation is a dominant strategy for the agent. Green and Laffont (1979*a*) devised procedures with this incentive property, but these are not individually rational, and their transfers do not balance. Fugigaki and Sato (1981), however, exhibited a class of generalized MDP procedures for which truthful reporting is locally dominant. One member of this class, moreover, is individually rational. Laffont and Maskin (1980*b*) exhibited the entire class of such procedures. They also showed the close connection between the theory of dynamic procedures and the static schemes mentioned earlier.

Most of the literature on incentives in resource allocation has taken preferences to be the information that agents transmit to the planner. There is, however, a small literature in which the relevant information is endowments. This includes the work of Postlewaite (1979) (dominant strategy equilibrium), Maskin (1980*a*) (dominant strategy and Nash equilibrium), and Hurwicz, Maskin, and Postlewaite (1979) (Nash equilibrium, both preferences and endowments private information).

The theory of *optimal nonlinear pricing* by a monopolist who does not know the preferences of individual consumers (although he may know the distribution of preferences) is another instance of pure message transmission by agents. Here the monopolist assumes the role of planner and maximizes profit (or expected profit, if he does not know the actual distribution of preferences). An incentive scheme is a rule that on the basis of an agent's professed preferences assigns the agent a quantity of the good and a price he must pay for it. A more familiar, but entirely equivalent, formulation has the monopolist announce a schedule relating prices and quantities, with the agents then choosing their favorite points along the schedule. Contributions to this literature include those of Spence (1977), Goldman, Leland, and Sibley (1977), Harris and Raviv (1981), Maskin and Riley (1980*a*), K. Roberts (1979*a*) among numerous others. An interesting special case is where the monopolist sells a single or several indivisible items. Then the monopolist's selling scheme is an auction. In an auction, an agent's message is his bid, and the incentive scheme is a rule that assigns each agent an amount to pay and a probability of winning the item on the basis of these bids. Optimal auctions (from the monopolist's viewpoint) have been studied by Harris and Raviv (1978), Riley and Samuelson (1981), Myerson (1981), Maskin and Riley (1980*b*), and Holt (1980).

The value of information in models of pure information transmission has been studied by Green (1979) and Green and Stokey (1980*a*).[15] In this work, θ^i, agent i's information, is a signal that is correlated with the payoff-relevant state of nature. On the basis of the message the agent sends him (there is only one agent), the planner takes a decision. The planner's and agent's ultimate payoffs depend on this decision as well as on the state of nature. To place this model within our framework, we must "expect out" the state of nature - which is not observed by anyone until all actions are taken - so that objective functions do not depend on the state.

Adverse selection with "active" agents. Some incentive models of adverse selection involve agents taking actions instead of (or in addition to) sending messages. These remain essentially adverse selection models, however, because the actions are perfectly observable.

In the literature on optimal commodity taxation (Diamond and Mirrlees, 1971), for example, agents choose net trades that are perfectly observable by the planner (tax authority). In the notation of Section 1, a net trade corresponds to a^i, and $y^i = a^i$. Decisions (taxes) are a function of the y^i's alone. However, the only obstacle to the attainment of a full optimum is the planner's imperfect knowledge of agents' preferences over net trades. Indeed, we can reformulate the problem equivalently as a pure message transmission in which agents report their preferences, and the planner thereupon assigns them net trades.

The recent research on implicit contracts with asymmetric information (for example, Green, 1980; Grossman and Hart, 1981) provides another example. In these models, a worker (who may be thought of as the planner) signs a contract with a firm (the agent) that specifies his compensation for each level of employment. It is assumed that when employment decisions are to be made, only the firm knows the worker's productivity, and so he chooses the level of employment unilaterally. This procedure is, of course, equivalent to an incentive scheme in which an employment-compensation pair is assigned to each possible announcement the firm could make about the worker's productivity.

2.2 *Models of moral hazard*

We next turn to models in which the failure to attain a full optimum is due to the inability of the planner to observe agents' actions perfectly. Many of these fall under the rubric of the *principal-agent problem*. For instance, see the work of Ross (1973), Holmstrom (1979*b*), Guenerie and Laffont (1979), Harris and Raviv (1979), Mirrlees (1975), Shavell (1979*a*, 1979*b*), Grossman and Hart (1980), and Stiglitz (1974). In these models, the principal (planner) observes outcome y, which depends randomly on the agent's action. A "decision" often takes the form of a monetary reward. An incentive scheme assigns a reward to the agent for each possible observed outcome. The planner's payoff depends on the reward (negatively) and outcome, whereas the agent's payoff depends on the reward (positively) and his action.

In the principal-agent problem, only moral hazard creates incentive problems. There are a number of models, however, that combine moral hazard and adverse selection.

One example is the capital allocation model described in Section 1. Another is the income tax model of Mirrlees (1971). In this latter model, agents share the same preferences for consumption and leisure. They differ, however, in their (constant) marginal products for producing the consumption good. Adverse selection arises because the planner (tax authority) does not know individual agents' marginal products. There is an additional problem ("moral hazard") created by his inability to observe agents' labor-leisure choices.

A final example combining moral hazard and adverse selection is described in the "bonus" literature (Weitzman, 1976). In these models, a planner attempts to elicit statistical information from an agent by an incentive scheme depending on the agent's message and an observed outcome y that depends stochastically on θ.

3 Indivisible public projects: an extended example of incentive theory

In this section we concentrate on a single, but much-studied, problem in incentive theory: the question whether or not society ought to undertake a given public project. Within the scheme of Section 1, this is a pure informational problem; the answer depends solely on agents' preferences – there is no question of observing their behavior. Nevertheless, the problem is representative of a large chunk of the incentive literature and is thus a useful illustrative example. Our intention is to show that by assuming that the relevant functions are differentiable (or, at least, piecewise differentiable), many of the major theorems, as well as some new results, can be easily derived. For a summary of these new results, see Section 3.2.

Throughout we shall consider a model with n consumers and two goods: one public, one private. The public good can either be produced ($x=1$) or not be produced ($x=0$); that is, it is indivisible. (For a similar analysis when the possible public project levels are continuous, see the work of Laffont and Maskin, 1979*a*, 1979*b*, 1980*a*, 1980*b*; the two theories are qualitatively very similar.) However, because, as pointed out in Section 1.4, randomization may be desirable in incentive problems where the outcome space is not convex, we shall often allow x to assume any value in the interval [0, 1]; x is then to be interpreted as the probability that the project will be carried out. Consumers' preferences for the public good and a vector of transfers $t=(t^1, \ldots, t^n)$ of private good are assumed to be representable by utility functions of the form

$$u^i(x,t) = \theta^i x + t^i \tag{3.1}$$

where θ^i lies in Θ^i, an open interval (containing zero) of the real line. If $x=0$ or 1, then equation (3.1) simply asserts that preferences are additively separable and linear in the private good. θ^i is consumer i's marginal rate of substitution or his "willingness to pay" (in terms of private good) for the public project.[16] If x can assume values strictly between 0 and 1, then (3.1) further implies that the consumer is risk-neutral in his attitude toward gambles on the level of public good. We shall assume that the functional form (3.1) is public knowledge but that the value θ^i is known, a priori, to consumer i alone.

For most of this part of the chapter, we shall work with dominant strategies (or coalitionally dominant strategies) as our solution concept. We do this not only because dominant strategies have figured most prominently in the literature to date but also because this solution concept, for a number of reasons discussed in Section 1.2, is the least controversial and the one making the weakest behavioral assumptions of those in current use.

3.1 *Definitions and summary of results*

In Section 1 we argued that, except for possible problems caused by multiple equilibria, it suffices to consider only direct revelation schemes when working with dominant strategies. Because we can show (see Theorem 3.8, *infra*) that, in our framework, multiple equilibrium outcomes cannot occur, we shall work only with such schemes.

An incentive scheme δ is a mapping,

$$\delta = (x, t^1, \ldots, t^n) : \prod_{i=1}^{n} \Theta^i \to [0,1] \times \mathbf{R}^n$$

which associates with each n-tuple $\hat{\theta} = (\hat{\theta}^1, \ldots, \hat{\theta}^n)$ of announced preference parameters a (possibly random) public decision $x(\hat{\theta})$ and a vector $t(\hat{\theta}) = (t^1(\hat{\theta}), \ldots, t^n(\hat{\theta}))$[17] of private-good transfers. Let $\tilde{\theta}^{-i} = (\tilde{\theta}^1, \ldots, \tilde{\theta}^{i-1}, \tilde{\theta}^{i+1}, \ldots, \tilde{\theta}^n)$ and

$$(\bar{\theta}^i, \tilde{\theta}^{-i}) = (\tilde{\theta}^1, \ldots, \tilde{\theta}^{i-1}, \bar{\theta}^i, \tilde{\theta}^{i+1}, \ldots, \tilde{\theta}^n).$$

An incentive scheme is *individually rational* if and only if

$$\theta^i x(\theta) + t^i(\theta) \geqslant 0 \quad \text{for all } i, \theta$$

That is, agent i, whatever the value of his preference parameter, can guarantee himself at least a zero or "status quo" payoff by announcing the truth.

An incentive scheme is *incentive-compatible in dominant strategies* (DSIC) if and only if

$$\theta^i x(\theta^i, \theta^{-i}) + t^i(\theta^i, \theta^{-i}) \geqslant \theta^i x(\hat{\theta}^i, \theta^{-i}) + t^i(\hat{\theta}^i, \theta^{-i})$$

for all i, θ^i, $\hat{\theta}^i$, and θ^{-i}. That is, the truth is always a dominant strategy.

Let C be a subset of $\{1, \ldots, n\}$. θ^C shall represent a vector of characteristics of members of C, whereas θ^{-C} shall be a vector for the complement of C. An incentive scheme is *incentive-compatible in coalitionally dominant strategies* (CDSIC) if for all $C \subseteq \{1, \ldots, n\}$

$$\sum_{i \in C} [\theta^i x(\theta^C, \theta^{-C}) + t^i(\theta^C, \theta^{-C})] \geqslant \sum_{i \in C} [\theta^i x(\hat{\theta}^C, \theta^{-C}) + t^i(\hat{\theta}^C, \theta^{-C})]$$

for all i, θ, and $\hat{\theta}^C$. That is, the truth is a dominant strategy even for a collusive coalition.

A DSIC scheme is *feasible* if and only if

$$\sum_{i=1}^{n} t^i(\theta) \leqslant 0 \quad \text{for all } \theta$$

Feasibility ensures that the designer will not run a deficit of private good. A stronger condition still is the requirement that the budget balance. A DSIC scheme is *balanced* if and only if

$$\sum_{i=1}^{n} t^i(\theta) = 0 \quad \text{for all} \quad \theta \in \Theta$$

If consumers' utilities are interpersonally and cardinally comparable, a common welfare objective is the maximization of the utilitarian social welfare function

$$\sum_{i=1}^{n} (\theta^i x + t^i) \tag{3.2}$$

Clearly, the maximizing choice of x is 1 if $\Sigma\theta^i \geqslant 0$ and 0 if $\Sigma\theta^i < 0$. This corresponds to the idea from cost-benefit theory that a project should be undertaken if the sum of the net benefits is positive. It is natural, therefore, to say that an incentive scheme is *successful* if and only if

$$\begin{aligned} x(\theta) &= 1, \quad \textstyle\sum\theta_i \geqslant 0 \\ &= 0, \quad \textstyle\sum\theta_i < 0 \end{aligned} \tag{3.3}$$

A property considerably weaker than success is the stipulation that the project be undertaken at least when everyone derives net positive benefit and that it be rejected (at least) when everyone suffers a net loss. Hence, a DSIC scheme is *weakly efficient* if and only if

$$\begin{aligned} x(\theta) &= 1, \quad \text{if} \quad \theta_i > 0 \quad \text{for all } i \\ &= 0, \quad \text{if} \quad \theta_i < 0 \quad \text{for all } i \end{aligned}$$

Finally, we shall call a scheme *fully optimal* if it is both successful and balanced. (It is fully optimal in that it maximizes (3.2) subject to the constraint of feasibility.)

The study of incentive in public-good provision has, to date, been largely concerned with successful mechanisms. We shall argue here that this emphasis has been, to some extent, misguided. But first, we begin, in Section 3.2, with success. Apart from reviewing some of the major results from the literature, we demonstrate (Theorem 3.3) that no feasible and successful incentive scheme dominates the much-studied Groves-Clarke pivotal mechanism. We then turn, in Section 3.3, to schemes that need not be successful. So that we can use calculus, we study schemes that are piecewise differentiable (more precisely, regular). We also often limit our attention to deterministic schemes (ones for which x takes on only the values 0 and 1). In the corollary to Theorem 3.7 we characterize all DSIC regular schemes. Next (Theorem 3.8), we show that equilibrium is essentially unique in such schemes. In Theorem 3.9 we characterize all deterministic DSIC schemes. Then, in Theorem 3.11, we show that a regular DSIC scheme that is weakly efficient, balanced, and symmetric (treats all agents identically) must be a positional dictatorship (i.e., there exists an integer i such that for each profile of parameters $(\theta^1, \ldots, \theta^n)$ the agent with the ith highest parameter "decides" on the project - if his parameter is nonnegative, the project is undertaken, otherwise not). Theorem 3.12 exhibits the "best" balanced DSIC scheme, the balanced scheme that maximizes the expectation of the utilitarian criterion (3.2). In Theorem 3.13 we show that the only weakly efficient, feasible, and individually rational DSIC scheme is the nth positional dictator. Finally, Theorem 3.14 establishes that no weakly efficient DSIC incentive scheme is immune to manipulation by coalitions.

Turning to Bayesian equilibrium in Section 3.4, we show that in our public project model the scheme proposed by d'Aspremont and Gérard-Varet (1979) and Arrow (1979) has a continuum of equilibria in addition to the one they proposed. We conclude in Theorems 3.16 and 3.17 by characterizing Nash-incentive-compatible schemes.

3.2 *Successful DSIC incentive schemes*

Theorem 3.1: Successful and feasible DSIC incentive schemes exist (Vickrey, 1961; Groves, 1973; Clarke, 1971; Smets, 1972):

Proof: Take x as in (3.3) and define

$$\begin{aligned} \bar{t}^i(\hat{\theta}) &= \sum_{j\neq i} \hat{\theta}^j, \quad \text{if} \quad \sum_{k=1}^{n} \hat{\theta}^k \geqslant 0 \quad \text{and} \quad \sum_{j\neq i} \hat{\theta}^j < 0 \\ &= -\sum_{j\neq i} \hat{\theta}^j, \quad \text{if} \quad \sum_{k=1}^{n} \hat{\theta}^k < 0 \quad \text{and} \quad \sum_{j\neq i} \hat{\theta}^j > 0 \\ &= 0, \quad \text{otherwise} \end{aligned}$$

The incentive scheme so defined is successful, feasible, and DSIC.

Q.E.D.

The scheme defined in the proof of Theorem 3.1 is called the Groves-Clarke pivotal mechanism (the term "pivotal" refers to the fact that only pivotal agents - those whose strategy changes the public decision from what it would be without them - are affected by transfers). Among feasible successful schemes, the pivotal mechanism is optimal in a sense defined by Theorem 3.3 (*infra*).

We first characterize all successful schemes.

Theorem 3.2: An incentive scheme is DSIC and successful if and only if $x(\cdot)$ satisfies (3.3) and

$$(3.4) \qquad t^i(\theta) = \sum_{j \neq i} \theta^j + h^i(\theta^{-i}), \quad \text{if} \quad \sum_{i=1}^{n} \theta^i \geqslant 0$$
$$= h^i(\theta^{-i}), \quad \text{if} \quad \sum_{i=1}^{n} \theta^i < 0$$

where $h^i(\cdot)$ is an arbitrary function of θ^{-i}.

Proof: (See the work of Green and Laffont, 1979*a*, Chapter 3, and Theorem 3.9, *infra.*) The pivotal mechanism is not balanced. Some of the transfers may be strictly negative; that is, there may be a net budget surplus. A natural question, therefore, is whether or not there exist feasible and successful schemes for which the magnitude of the surplus is smaller. Although it is easy to give examples of feasible schemes yielding smaller surpluses for some values of θ, no such scheme dominates the pivotal mechanism uniformly. That is:

Theorem 3.3: There exists no feasible and successful incentive scheme (x, t) such that for all θ

$$|\sum t^i(\theta)| \leqslant |\sum \bar{t}^i(\theta)|$$

with strict inequality for some θ, where, as before, $\bar{t}^i$ is the pivotal mechanism's transfer to agent i.

Proof: Because the proof is long and messy, we relegate it to the Appendix.

Corollary: (See the work of Green and Laffont, 1976, 1979*a*, Chapter 5, and Hurwicz, 1975.) There exists no fully optimal DSIC scheme.

Proof: The Groves-Clarke mechanism is successful and feasible but not balanced; it sometimes generates a strictly positive surplus. From Theorem 3.3, there exists no successful and feasible scheme that dominates Groves-Clarke. That is, no such scheme is balanced. Q.E.D.

The Groves-Clarke mechanism shows that success and feasibility are mutually consistent. Similarly, by taking $h^i(\theta^{-i}) \equiv 0$ in (3.4), we obtain a successful and individually rational scheme. However, feasibility, individual rationality, and success cannot be satisfied simultaneously.

Theorem 3.4: No feasible, individually rational, successful DSIC scheme exists.

Proof: Choose θ such that $\sum_{j=1}^{n} \theta^j > 0$ and such that for all i there exists $\bar{\theta}^i \in \Theta^i$ with $\bar{\theta}^i + \sum_{j \neq i} \theta^j < 0$. Suppose that $\delta = (x, t)$ is a feasible, individually rational, successful scheme. From Theorem 3.2, t^i satisfies (3.4) for some function h^i. Thus, from success, the payoff to agent i with parameter $\bar{\theta}^i$, if other agents have parameters θ^{-i}, is $h^i(\theta^{-i})$. From individual rationality,

$$(3.15) \quad h^i(\theta^{-i}) \geqslant 0$$

But from feasibility and success,

$$(n-1) \sum_{i=1}^{n} \theta^i + \sum_{i=1}^{n} h^i(\theta^{-i}) \leqslant 0$$

and so

$$\sum_{i=1}^{n} h^i(\theta^{-i}) < 0$$

in contradiction to (3.15). Q.E.D.

Theorem 3.5: No successful CDSIC scheme exists.

Proof: (See the work of Bennett and Conn, 1977, and Green and Laffont, 1979*a*, Chapter 5, and Theorem 3.14, *infra.*)

Because fully optimal DSIC incentive schemes do not exist, the requirement that schemes nonetheless be successful is arbitrary. After all, success pertains only to the public decision, and so to require success alone in a scheme is to ignore the welfare implications of its private transfers. A more general approach consists of characterizing the class of all DSIC incentive schemes and then optimizing whatever welfare function one might have (e.g., function (3.2)) subject to the scheme's being in this or a narrower class (e.g., the class of balanced schemes or the class of successful schemes). This is the approach we now briefly pursue. We establish a number of results characterizing DSIC incentive schemes, and, in particular (see Theorem 3.12), we consider the optimization of (3.2) subject to the scheme's being balanced.

3.3 General DSIC incentive schemes

An incentive scheme is differentiable if $x(\cdot)$ and $t(\cdot)$ are differentiable. For analytical simplicity, we are concerned in this section with *simple* and *regular* incentive schemes. A *simple* incentive scheme (x, t) is defined in terms of a closed set A with the property that if $\theta \in A$ and $\theta' \geqslant \theta$, then $\theta' \in A$. The scheme is simple if there exist differentiable functions x^- and x^+ such that

$$\begin{aligned} x(\theta) &= x^-(\theta), \quad \theta \notin A \\ &= x^+(\theta), \quad \theta \in A \end{aligned}$$

A *regular* scheme (x, t) is a straightforward generalization of a simple scheme. Instead of the single set A, there is a collection of closed sets $A_1, \ldots, A_q$, each with the property that if $\theta \in A_j$ and $\theta' \geqslant \theta$, then $\theta \in A_j$. For each j, there is a pair of differentiable functions x_j^- and x_j^+ such that

$$x(\theta) = \sum_{j=1}^{n} x^{z(j)}(\theta)$$

where $z(j)$ is $-$ if $\theta \notin A_j$, and $z(j)$ is $+$ if $\theta \in A_j$.

Some examples of simple incentive schemes are the Groves-Clarke mechanism (see Theorem 3.1), the ith dictatorship, and the ith positional dictatorship.

Example 3.1: The ith dictatorship.

$$\begin{aligned} x(\theta) &= 1, \quad \theta^i \geqslant 0 \\ &= 0, \quad \text{otherwise} \end{aligned}$$

$$t^j(\theta) = h^j(\theta^{-j}) \quad \text{for all } j$$

Example 3.2: The ith positional dictatorship.

$$\begin{aligned} x(\theta) &= 1, \quad \{j \mid \theta^j \geqslant 0\} \quad \text{has at least } i \text{ elements,} \\ &= 0, \quad \text{otherwise} \end{aligned}$$

$$t^j(\theta) \equiv h^j(\theta^{-j}) \quad \text{for all } j$$

An example of a regular scheme that is not simple is the random dictatorship.

Example 3.3: Random dictatorship.

$$x(\theta) = \sum_{j=1}^{n} x^k(\theta), \quad \text{where} \quad x^k(\theta) = 0, \quad \text{if} \quad \theta^k < 0$$

$$= \frac{1}{n}, \quad \text{if} \quad \theta^k \geqslant 0$$

$$t^j(\theta) = h^j(\theta^{-j}) \quad \text{for all } j$$

We shall call a public decision function $x: \Pi_{i=1}^{n} \Theta^i \rightarrow [0,1]$ *implementable* if there exists a transfer rule vector t such that (x, t) is a DSIC scheme.

Lemma: A public decision function $x(\cdot)$ is implementable only if it is weakly increasing.

Proof: Let $\bar{\theta}^i$ and $\tilde{\theta}^i$ be alternative values of agent i's characteristic. If x is implementable, there exists t such that

(3.16) $\quad \bar{\theta}^i x(\bar{\theta}^i, \theta^{-i}) + t^i(\bar{\theta}^i, \theta^{-i}) \geqslant \bar{\theta}^i x(\tilde{\theta}^i, \theta^{-i}) + t^i(\tilde{\theta}^i, \theta^{-i}) \quad$ for all θ^{-i}

and

(3.17) $\quad \tilde{\theta}^i x(\tilde{\theta}^i, \theta^{-i}) + t^i(\tilde{\theta}^i, \theta^{-i}) \geqslant \tilde{\theta}^i x(\bar{\theta}^i, \theta^{-i}) + t^i(\bar{\theta}^i, \theta^{-i}) \quad$ for all θ^{-i}

Adding (3.16) to (3.17) and collecting terms, we obtain

$$(\bar{\theta}^i - \tilde{\theta}^i)x(\bar{\theta}^i, \theta^{-i}) - x(\tilde{\theta}^i, \theta^{-i})) \geqslant 0 \qquad \text{Q.E.D.}$$

We shall begin by characterizing differentiable DSIC schemes.

Theorem 3.6: A differentiable incentive scheme (x, t) is DSIC if and only if (i) x is weakly increasing and (ii)

$$t^i(\theta) = -\int_0^{\theta^i} s \frac{\partial x}{\partial \theta^i} (s, \theta^{-i})\, ds + h^i(\theta^{-i})$$

where h^i is an arbitrary piecewise differentiable function of θ^{-i}, $i = 1, \ldots, n$.

Proof: We begin with necessity. From the lemma, x must be weakly increasing. For each θ^{-i}, agent i chooses $\hat{\theta}^i$ to maximize

$$\theta^i x(\hat{\theta}^i, \theta^{-i}) + t^i(\hat{\theta}^i, \theta^{-i})$$

If the maximum is to occur at $\hat{\theta}^i = \theta^i$, we must have

$$(3.18)\quad \theta^i \frac{\partial x}{\partial \theta^i}(\theta^i,\theta^{-i}) + \frac{\partial t^i}{\partial \theta^i}(\theta^i,\theta^{-i}) = 0$$

Because (3.18) must hold for all θ, (3.18) is an identity. Thus

$$(3.19)\quad t^i(\theta) = -\int_0^{\theta^i} s \frac{\partial x}{\partial \theta^i}(s,\theta^{-i})\, ds + h^i(\theta^{-i})$$

where h^i is an arbitrary differentiable function of θ^{-i}. Thus necessity is established. For sufficiency, observe that, in view of (ii),

$$\theta^i x(\theta^i,\theta^{-i}) + t^i(\theta^i,\theta^{-i}) \geqslant \theta^i x(\hat{\theta}^i,\theta^{-i}) + t^i(\hat{\theta}^i,\theta^{-i}) \quad \text{for all} \quad \hat{\theta}^i,\theta^{-i}$$

if and only if

$$(3.20)\quad (\theta^i - \hat{\theta}^i)x(\hat{\theta}^i,\theta^{-i}) \leqslant \int_{\hat{\theta}^i}^{\theta^i} x(s,\theta^{-i})\, ds \quad \text{for all } \hat{\theta}^i$$

But (3.20) holds because x is weakly increasing. Q.E.D.

We are now ready to characterize simple DSIC incentive schemes. For simple schemes, as defined earlier, let

$$a^i(\theta^{-i}) = \min\{\theta^i \mid (\theta^i,\theta^{-i}) \in A\} \quad \text{(if this minimum exists)}$$

Theorem 3.7: A simple scheme is DSIC if and only if (i) x is weakly increasing and (ii)

$$t^i(\theta) = -\int_0^{\theta^i} s \frac{\partial}{\partial \theta^i} x^-(s,\theta^{-i})\, ds + h^i(\theta^{-i}), \quad \theta \notin A$$

$$= -\int_0^{\theta^i} s \frac{\partial}{\partial \theta^i} x^+(s,\theta^{-i})\, ds + h^i(\theta^{-i}) + C^i(\theta^{-i}), \quad \theta \in A$$

where h^i is an arbitrary piecewise differentiable function and

$$C^i(\theta^{-i}) = 0, \quad \text{if} \quad a^i(\theta^{-i}) \quad \text{is not defined}$$

$$= \int_0^{a^i(\theta^{-i})} (x^-(s,\theta^{-i}) - x^+(s,\theta^{-i}))\, ds, \quad \text{otherwise}$$

Proof: Suppose that the simple scheme (x, t) is DSIC. From the lemma, x must be weakly increasing. From Theorem 3.6,

$$(3.21)\quad t^i(\theta) = -\int_0^{\theta^i} s\frac{\partial x^-}{\partial\theta^i}(s,\theta^{-i})\,ds + h^i(\theta^{-i}),\quad \text{if}\quad \theta \notin A$$

$$= -\int_0^{\theta^i} s\frac{\partial x^+}{\partial\theta^i}(s,\theta^{-i})\,ds + k^i(\theta^{-i}),\quad \text{if}\quad \theta \in A$$

Now $\hat{\theta}^i x(\hat{\theta}^i, \theta^{-i}) + t^i(\hat{\theta}^i, \theta^{-i})$ is evidently continuous as a function of $\hat{\theta}^i$ for $\hat{\theta}^i < a^i(\theta^{-i})$ and for $\hat{\theta}^i > a^i(\theta^{-i})$. We claim it is continuous as well at $\hat{\theta}^i = a^i(\theta^{-i})$. If not, then $\gamma(\theta^{-i}) \neq 0$, where

$$\gamma(\theta^{-i}) = \lim_{\hat{\theta}^i \to a^i(\theta^{-i})} (\hat{\theta}^i x^-(\hat{\theta}^i, \theta^{-i}) + t^i(\hat{\theta}^i, \theta^{-i}))$$
$$- (a^i(\theta^{-i})x^+(a^i(\theta^{-i}), \theta^{-i}) + t^i(a^i(\theta^{-i}), \theta^{-i}))$$

Suppose $\gamma(\theta^{-i}) > 0$. Then for $\theta^i = a^i(\theta^{-i})$ and $\hat{\theta}^i$ slightly less than θ^i, agent i's payoff is higher from the strategy $\hat{\theta}^i$, if his parameter is θ^i, than from announcing the truth. Similarly, if $\gamma(\theta^{-i}) < 0$, agent i's payoff is larger from $\hat{\theta}^i = a^i(\theta^{-i})$ than from $\hat{\theta}^i = \theta^i$, if θ^i is slightly less than $a^i(\theta^{-i})$. Hence, $\gamma(\theta^{-i}) = 0$, and continuity at $a^i(\theta^{-i})$ is established. But from (3.21), continuity implies

$$a^i(\theta^{-i})x^-(a^i(\theta^{-i}), \theta^{-i}) - \int_0^{a^i(\theta^{-i})} s\frac{\partial x^-}{\partial\theta^i}(s,\theta^{-i})\,ds + h^i(\theta^{-i})$$

$$= a^i(\theta^{-i})x^+(a^i(\theta^{-i}), \theta^{-i}) - \int_0^{a^i(\theta^{-i})} s\frac{\partial x^+}{\partial\theta^i}(s,\theta^{-i})\,ds + k^i(\theta^{-i})$$

Integrating by parts and rearranging terms, we obtain

$$k^i(\theta^{-i}) = h^i(\theta^{-i}) + \int_0^{a^i(\theta^{-i})} (x^-(s,\theta^{-i}) - x^+(x,\theta^{-i})\,ds$$

thus establishing the necessity of (ii). For sufficiency, consider $\theta \notin A$ and $\hat{\theta}^i \geqslant a^i(\theta^{-i})$. The gain to agent i with parameter θ^i from announcing the truth rather than $\hat{\theta}^i$ is

$$(3.22)\quad \theta^i x^-(\theta) - \int_0^{\theta^i} s\frac{\partial x^-}{\partial\theta^i}(s,\theta^{-i})\,ds$$

$$- \left(\theta^i x^+(\hat\theta^i,\theta^{-i}) - \int_0^{\hat\theta^i} s\frac{\partial x^+}{\partial\theta^i}(s,\theta^{-i})\,ds + C(\theta^{-i})\right)$$

$$= (\hat\theta^i - \theta^i)x^+(\hat\theta^i,\theta^{-i}) - \int_{\theta^i}^{a^i(\theta^{-i})} x^-(s,\theta^{-i})\,ds$$

$$- \int_{a^i(\theta^i)}^{\hat\theta^i} x^+(s,\theta^{-i})\,ds$$

But (3.22) is nonnegative because x is (weakly) increasing. Similarly, the gain from truth telling is nonnegative if $\theta\in A$ and $\hat\theta^i < a^i(\theta^{-i})$. Q.E.D.

The generalization of Theorem 3.7 to regular incentive schemes is immediate. For all i, j, let

$$a_j^i(\theta^{-i}) = \min\{\theta^i \mid (\theta^i,\theta^{-i}) \in A_j\} \quad \text{(if this minimum exists)}$$

Corollary: A regular scheme is DSIC if and only if (i) x is weakly increasing and (ii)

$$t^i(\theta) = \sum_{j=1}^{q}\left[-\int_0^{\theta^i} \frac{\partial}{\partial\theta^i} x_j^{z(j)}(s,\theta^{-i})\,ds + h^i(\theta^{-i}) + y(j)C_j^i(\theta^{-i})\right]$$

where h^i is an arbitrary piecewise differentiable function, $y(j)=1$ for $\theta\in A_j$ and 0 for $\theta\notin A_j$ and

$$C_j^i(\theta^{-i}) = 0, \quad \text{if} \quad a_j^i(\theta^{-i}) \quad \text{is not defined}$$

$$= \int_0^{a_j^i(\theta^{-i})} (x_j^+(s,\theta^{-i}) - x_j^-(s,\theta^{-i}))\,ds, \quad \text{otherwise}$$

We now can demonstrate that a DSIC regular scheme has a unique equilibrium public decision for each choice of θ.

Theorem 3.8: A DSIC regular scheme (x, t) has a unique equilibrium public decision $x(\theta)$ for each choice of θ.

Proof: We shall argue for the case of simple schemes. Suppose that, for some θ and $\bar{\theta}$, $\hat{\theta}=\bar{\theta}$ is a dominant strategy equilibrium if the true parameters are θ. That is, $x(\bar{\theta})$ is an equilibrium public decision in addition to $x(\theta)$. In particular, because $\bar{\theta}^1$ is a dominant strategy for agent 1, we have

$$\theta^1 x(\theta^1,\theta^{-1}) + t^1(\theta^1,\theta^{-1}) = \theta^1 x(\bar{\theta}^1,\theta^{-1}) + t^1(\bar{\theta}^1,\theta^{-1})$$

that is,

$$\theta^1(x(\theta^1,\theta^{-1}) - x(\bar{\theta}^1,\theta^{-1})) = t^1(\bar{\theta}^1,\theta^{-1}) - t^1(\theta^1,\theta^{-1}) \tag{3.23}$$

If either both (θ^1,θ^{-1}) and $(\bar{\theta}^1,\theta^{-1})$ lie in A or do not lie in A, then (3.23) becomes

$$\begin{aligned}\theta^1(x(\theta^1,\theta^{-1}) - x(\bar{\theta}^1,\theta^{-1})) &= -\int_0^{\bar{\theta}^1} s\frac{\partial}{\partial\theta^1}x(s,\theta^{-1})\,ds \\ &\quad + \int_0^{\theta^1} s\frac{\partial}{\partial\theta^1}x(s,\theta^{-1})\,ds \\ &= \theta^1 x(\theta^1,\theta^{-1}) - \bar{\theta}^1 x(\bar{\theta}^1,\theta^{-1}) \\ &\quad - \int_{\bar{\theta}^1}^{\theta^1} x(s,\theta^{-1})\,ds\end{aligned}$$

Rearranging, we obtain

$$(\bar{\theta}^1 - \theta^1)x(\bar{\theta}^1,\theta^{-1}) = \int_{\theta^1}^{\bar{\theta}^1} x(s,\hat{\theta}^{-1})\,ds \tag{3.24}$$

Because x is weakly increasing, we conclude that

$$x(\hat{\theta}^1,\hat{\theta}^{-1}) = x(\theta^1,\theta^{-1}), \quad \text{for all } \hat{\theta}^1 \text{ between } \theta^1 \text{ and } \bar{\theta}^1. \tag{3.25}$$

If $(\theta^1,\theta^{-1}) \notin A$ but $(\bar{\theta}^1,\theta^{-1}) \in A$ (the opposite case can be argued similarly), then (3.23) becomes

$$\theta^1(x^-(\theta^1,\theta^{-1}) - x^+(\bar{\theta}^1,\theta^{-1})) = -\int_0^{\bar{\theta}^1} s\frac{\partial}{\partial\theta^1}x^+(s,\theta^{-1})\,ds$$

$$+ \int_0^{\theta^1} s \frac{\partial}{\partial \theta^1} x^-(s, \theta^{-1})\, ds + C^1(\theta^{-1})$$

$$= -\bar{\theta}^1 x^+(\bar{\theta}^1, \theta^{-1}) + \theta^1 x^-(\theta^1, \theta^{-1})$$

$$+ \int_{a^1(\theta^{-1})}^{\bar{\theta}^1} x^+(s, \theta^{-1})\, ds$$

$$+ \int_{\theta^1}^{a^1(\theta^{-1})} x^-(s, \theta^{-1})\, ds$$

Rearranging, we find

$$(\bar{\theta}^1 - \theta^1) x^+(\bar{\theta}^1, \theta^{-1}) = \int_{a^1(\theta^{-1})}^{\theta^1} x^+(s, \theta^{-1})\, ds + \int_{\theta^1}^{a^1(\theta^{-1})} x^-(s, \theta^{-1})\, ds$$

from which we conclude, because x is weakly increasing and $x^+ \geqslant x^-$, that (3.25) again holds. Thus, in all cases (3.25) holds. Continuing iteratively for $i = 2, \ldots, n$,

$$x(\theta) = x(\bar{\theta}) \qquad \text{Q.E.D.}$$

From Theorem 3.7 we can immediately characterize those regular DSIC schemes that are *deterministic* (x can take on only the values 0 and 1). Such schemes, of course, are automatically simple.

Theorem 3.9: A regular deterministic incentive scheme is DSIC if and only if

(i) $$\begin{aligned} x(\theta) &= 0, \quad \theta \notin A \\ &= 1, \quad \theta \in A \end{aligned}$$

(ii) $$\begin{aligned} t^i(\theta) &= h^i(\theta^{-i}), \quad \theta \notin A \\ &= -a^i(\theta^{-i}) + h^i(\theta^{-i}), \quad \theta \in A \text{ and } a^i(\theta^{-i}) \text{ defined} \\ &= h^i(\theta^{-i}), \quad \text{otherwise} \end{aligned}$$

Note that Theorem 3.2 is a special case of Theorem 3.9, in which

$$\begin{aligned} x(\theta) &= 1, \quad \textstyle\sum \theta^i \geqslant 0 \\ &= 0, \quad \text{otherwise} \end{aligned}$$

Hence $a^i(\theta^{-i}) = -\sum_{j\neq i}\theta^j$. Theorem 3.9 also shows that the form of the transfers when the public decision rule is that of a dictatorship or positional dictatorship is

$$t^i(\theta) = h^i(\theta^{-i})$$

We turn next to the issue of balance.

Theorem 3.10: If a regular incentive scheme is balanced, then

$$\frac{\partial^n x_j^+}{\partial\theta^i \dots \partial\theta^n} \equiv 0 \equiv \frac{\partial^n x_j^-}{\partial\theta^1 \dots \partial\theta^n}, \quad j = 1, \dots, q$$

Proof: For simplicity we shall argue for the case of simple schemes. Choose $\theta \notin A$. From balance,

$$\sum_{i=1}^{n} t^i(\theta) \equiv 0 \tag{3.26}$$

From Theorem 3.7,

$$\sum_{i=1}^{n}\left[\int_0^{\theta^i} s\frac{\partial x^-}{\partial\theta^i}(s,\theta^{-i})\,ds + h^i(\theta^{-i})\right] \equiv 0$$

and so

$$\sum_{i=1}^{n}\theta^i\frac{\partial^n x^-}{\partial\theta^1 \dots \partial\theta^n} \equiv 0$$

We have, therefore,

$$\frac{\partial^n x^-}{\partial\theta^1 \dots \partial\theta^n} \equiv 0$$

Similarly for x^+. Q.E.D.

We observed earlier that when the public decision rule corresponds to a generalized dictatorship, the transfers take the form $t^i(\theta) = h^i(\theta^{-i})$. In particular, if $h^i(\theta^{-i}) \equiv 0$, the scheme is automatically balanced. One may ask whether or not there exist balanced deterministic DSIC schemes that are nondictatorial. The following example answers the question affirmatively.

Example 3.4: Let $n=3$ and let $x(\theta)=1$ for $\theta^2+\theta^3 \geqslant 0$ and 0 otherwise,

$$t^1(\theta) = -\theta^2 - \theta^3, \quad \theta^2 + \theta^3 \geqslant 0$$
$$= 0, \quad \text{otherwise}$$
$$t^2(\theta) = \theta^3, \quad \theta^2 + \theta^3 \geqslant 0$$
$$= 0, \quad \text{otherwise}$$
$$t^3(\theta) = \theta^2, \quad \theta^2 + \theta^3 \geqslant 0$$
$$= 0, \quad \text{otherwise}$$

The scheme of Example 3.4 treats agent 1 asymmetrically, and with good reason: Nondictatorial, deterministic balanced schemes cannot be symmetric. We shall call an incentive scheme (x, t) symmetric if x is a symmetric function.

Theorem 3.11: A weakly efficient, symmetric, deterministic, balanced, and regular DSIC scheme must be a positional dictatorship.

Proof: See Appendix.

Theorem 3.11 applies only to deterministic schemes, and so it is natural to ask what "good," *nondeterministic* schemes that are balanced look like. By a good scheme, we mean one that maximizes the utilitarian criterion (3.2) in an expected sense, where the expectation is performed with respect to a prior distribution $F(\theta^1, \ldots, \theta^n)$. That is, we seek a balanced scheme (x, t) that maximizes

(3.44) $$\int \sum_{i=1}^{n} (\theta^i x(\theta) + t^i(\theta)) \, dF(\theta)$$

Because transfers sum to zero in a balanced scheme, (3.44) becomes

(3.45) $$\int \sum_{i=1}^{n} \theta^i x(\theta) \, dF(\theta)$$

From Theorem 3.10, balance implies that if x is n times differentiable (or the pointwise limit of a sequence of n-times differentiable functions), then

(3.46) $$x(\theta) = \sum_{i=1}^{n} x_i(\theta^{-i})$$

It is easy to see, moreover, that (3.46) is a sufficient condition for the existence of transfers that balance. Thus, if we restrict our attention to x's that are pointwise limits of n-times differentiable functions, we seek $x_1, \ldots, x_n$ to maximize

(3.47) $$\int\left(\sum_{i=1}^{n}\theta^i\right)\left(\sum_{j=1}^{n}x_j(\theta^{-j})\right)dF(\theta)$$

such that

(3.48) $$0 \leqslant \sum x_j(\theta^{-j}) \leqslant 1$$

and

(3.49) x_j's are weakly increasing

Suppose that $x_2, \ldots, x_n$ have already been chosen optimally and that $E(\theta^1 \mid \theta^2, \ldots, \theta^n)$ (the expectation of θ^1 conditional on $\theta^2, \ldots, \theta^n$) is independent of $(\theta^2, \ldots, \theta^n)$. We must choose x_1 to maximize

(3.50) $$\int(E\theta^1 + \theta^2 + \cdots + \theta^n)x_1(\theta^{-1})\,dF^1(\theta^{-1})$$

subject to constraints (3.48) and (3.49), where F^1 is the marginal distribution of θ^{-1}. From (3.48) and (3.49), for any θ^{-1},

(3.51) $$x_1(\theta^{-1}) \leqslant \lim_{\tilde{\theta}^{-1}\to\infty} x_1(\tilde{\theta}^{-1}) \leqslant 1 - \sum_{i=2}^{n}\lim_{\tilde{\theta}^{-i}\to\infty} x_i(\tilde{\theta}^{-i})$$

Take

$$\mu_i = \lim_{\tilde{\theta}^{-i}\to\infty} x_i(\tilde{\theta}^{-i})$$

From (3.50) and (3.51), the optimal choice of $x_1(\theta^{-1})$ is

$$\begin{aligned} x_1(\theta^{-1}) &= 0, \quad \sum_{j=2}^{n}\theta^j < E\theta^1 \\ &= 1 - \sum_{i=2}^{n}\mu_i, \quad \sum_{j=2}^{n}\theta^j \geqslant E\theta^1 \end{aligned}$$

Thus (3.47) becomes

(3.52) $$\sum_{i=1}^{n}\mu_i \operatorname{Prob}\left(\sum_{j\neq i}\theta^j \geqslant -E\theta^i\right)$$

We can therefore state Theorem 3.12.

Theorem 3.12: The balanced incentive scheme (x, t), where

$$x(\theta) = \sum_{j=1}^{n} x_j(\theta^{-j})$$

$$x_j(\theta^{-j}) = 0, \quad \sum_{i \neq j} \theta^i + E\theta^j < 0$$
$$= \mu_j, \text{ otherwise}$$

$$\sum_{j \in D} \mu_j = 1$$

$$D = \left\{ j \mid \text{Prob}\left\{ \sum_{i \neq j} \tilde{\theta}^i + E\theta^j \geqslant 0 \right\} \text{ is maximal} \right\}$$

maximizes (3.44) among all balanced incentive schemes for which x is the pointwise limit of a sequence of n-times differentiable functions.

Note that if F is symmetric and $E\theta^j = 0$, then x_j in Theorem 3.12 becomes

$$x_j(\theta^{-j}) = 0, \quad \sum_{i \neq j} \theta^i < 0$$
$$= \frac{1}{n}, \quad \sum_{i \neq j} \theta^i \geqslant 0$$

That is, the best balanced scheme has a public decision rule that is utilitarian for each of the n coalitions of $n-1$ agents, where each coalition contributes probability weight $1/n$.

It is natural to compare the welfare properties of the best balanced scheme with those of the Groves-Clarke mechanism. The former has balanced transfers but does not always take the public decision maximizing the utilitarian criterion. The latter takes the correct public decision but does not always balance the budget (although, from Theorem 3.3, no alternative feasible and successful DSIC scheme dominates it). There certainly seems to be no a priori reason to favor the Groves-Clarke mechanism. Therefore, we believe that the emphasis in the literature on successful mechanisms, ignoring welfare losses due to transfer imbalances, is somewhat misguided. Indeed, as the following example shows, the best balanced scheme can do better than the Groves-Clarke mechanism in the expected sense of (3.44).

Example 3.5: Take $n=2$ and let $F(\theta^1, \theta^2)$ be the joint uniform distribution on $[-\frac{1}{2}, \frac{1}{2}]$. That is,

$$F(\theta^1, \theta^2) = (\theta^1 + \tfrac{1}{2})(\theta^2 + \tfrac{1}{2})$$

The expected sum of utilities under the Groves-Clarke mechanism is the expected sum of utilities from the public decision

$$\int_{\theta^1+\theta^2\geqslant 0} (\theta^1 + \theta^2)\, dF(\theta^1,\theta^2) = \frac{1}{6}$$

plus the expected sum of transfers

$$-\int_{\substack{\theta^1+\theta^2\geqslant 0\\ \theta^1\geqslant 0\\ \theta^2\leqslant 0}} \theta^2\, dF(\theta^1,\theta^2) - \int_{\substack{\theta^1+\theta^2\geqslant 0\\ \theta^1\leqslant 0\\ \theta^2\geqslant 0}} \theta^1\, dF(\theta^1,\theta^2) - \int_{\substack{\theta^1+\theta^2< 0\\ \theta^1\geqslant 0\\ \theta^2\leqslant 0}} \theta^1\, dF(\theta^1,\theta^2)$$

$$- \int_{\substack{\theta^1+\theta^2< 0\\ \theta^1\leqslant 0\\ \theta^2\geqslant 0}} \theta^2\, dF(\theta^1,\theta^2) = -\frac{1}{12}$$

That is, the expected sum of utilities is 1/12. The expected sum under the best balanced mechanism, on the other hand, is just the expected public payoff:

$$\int_{\substack{\theta^1\geqslant 0\\ \theta^2\geqslant 0}} (\theta^1 + \theta^2)\, dF + \frac{1}{2}\int_{\substack{\theta^1\geqslant 0\\ \theta^2< 0}} (\theta^1 + \theta^2)\, dF + \frac{1}{2}\int_{\substack{\theta^1< 0\\ \theta^2> 0}} (\theta^1 + \theta^2)\, dF = \frac{7}{48}$$

Therefore, the best balanced mechanism is better than Groves-Clarke. Of course, we have taken the extreme position of treating the budget surplus under Groves-Clarke as a total loss. Nonetheless, the example illustrates that it is unduly restrictive to consider only successful schemes.

We next turn to feasibility and individual rationality. As in the special case when x satisfies (3.3), we can readily choose t so that (x, t) is DSIC and feasible or DSIC and individually rational if x satisfies the conditions for a regular public decision rule. However, feasibility and individual rationality together cannot be satisfied by a weakly efficient DSIC scheme unless it is the nth positional dictator.

Theorem 3.13: If a weakly efficient and feasible regular DSIC scheme is individually rational, then it is the nth positional dictator.

Proof: Suppose that (x, t) satisfies the hypotheses. We shall assume for convenience that (x, t) is simple. Choose $\bar{\theta}\geqslant 0$. From weak efficiency, $\bar{\theta}\in A$. If for some i, $a^i(\bar{\theta}^{-i})$ is not defined, then choose $\bar{\theta}^i<0$. We have

$(\tilde{\theta}^i, \bar{\theta}^{-i}) \in A$. From Theorem 3.7, agent i's payoff when the parameters are $(\tilde{\theta}^i, \bar{\theta}^{-i})$ is

$$\int_0^{\tilde{\theta}^i} x^+(s, \bar{\theta}^{-i})\, ds + h^i(\bar{\theta}^{-i})$$

From individual rationality,

$$h^i(\bar{\theta}^{-i}) \geqslant \int_{\tilde{\theta}^i}^0 x^+(s, \bar{\theta}^{-i})\, ds$$

From choice of $\tilde{\theta}^i$, and because $x^+(0, \bar{\theta}^{-i}) = 1$, $\int_{\tilde{\theta}^i}^0 x^+(s, \bar{\theta}^{-i})\, ds > 0$. Thus

$$h^i(\bar{\theta}^{-i}) > 0 \tag{3.53}$$

If, for given i, $a^i(\bar{\theta}^{-i})$ is defined, then from weak efficiency,

$$a^i(\bar{\theta}^{-i}) \leqslant 0 \tag{3.54}$$

For such i, choose $\hat{\theta}^i < a^i(\bar{\theta}^{-i})$. Then $(\hat{\theta}^i, \bar{\theta}^{-i}) \notin A$. Thus i's payoff when the parameters are $(\hat{\theta}^i, \bar{\theta}^{-i})$ is

$$\int_0^{\hat{\theta}^i} x^-(s, \bar{\theta}^{-i})\, ds + h^i(\bar{\theta}^{-i})$$

From individual rationality,

$$h^i(\bar{\theta}^{-i}) > \int_{\hat{\theta}^i}^0 x^-(s, \bar{\theta}^{-i})\, ds \tag{3.55}$$

Therefore, because, from (3.54), $\hat{\theta}^i < 0$,

$$h^i(\bar{\theta}^{-i}) \geqslant 0 \tag{3.56}$$

From Theorem 3.7, i's transfer, when the parameters are $\bar{\theta}$ and $a^i(\bar{\theta}^{-i})$ is not defined, is

$$-\bar{\theta}^i x^+(\bar{\theta}) + \int_0^{\bar{\theta}^i} x^+(s, \bar{\theta}^{-i})\, ds + h^i(\bar{\theta}^{-i}) \tag{3.57}$$

From weak efficiency, $x^+(s, \bar{\theta}^{-1}) = 1$ for all $s \geqslant 0$. Thus, (3.57) reduces to

$$h^i(\bar{\theta}^{-i})$$

which is positive, from (3.53). Thus, because the sum of the transfers is nonpositive by feasibility, there must exist i for whom $a^i(\bar{\theta}^{-i})$ is defined and whose transfer, when parameters are $\bar{\theta}$, is nonpositive. That is,

$$(3.58) \quad -\bar{\theta}^i x^+(\bar{\theta}) + \int_0^{\bar{\theta}^i} x^+(s, \bar{\theta}^{-i})\, ds + h^i(\bar{\theta}^{-i})$$

$$+ \int_0^{a^i(\bar{\theta}^{-i})} (x^{-(s,\bar{\theta}^{-i})} - x^+(s, \bar{\theta}^{-i}))\, ds + h^i(\bar{\theta}^{-i}) \leqslant 0$$

The first two terms on the left-hand side of (3.58) cancel, because $x^+(s, \theta) = 1$ for $s \geqslant 0$. Therefore, (3.58) becomes

$$(3.59) \quad h^i(\bar{\theta}^{-i}) \leqslant \int_{a^i(\bar{\theta}^{-i})}^0 (x^-(s, \bar{\theta}^{-i}) - x^+(s, \bar{\theta}^{-i}))\, ds$$

Because in (3.55) $\hat{\theta}^i < a^i(\bar{\theta}^{-i})$, (3.59) implies

$$(3.60) \quad h^i(\bar{\theta}^{-i}) \leqslant h^i(\bar{\theta}^{-i}) - \int_{a^i(\bar{\theta}^{-i})}^0 x^+(s, \bar{\theta}^{-i})\, ds = h^i(\bar{\theta}^{-i}) + a^i(\bar{\theta}^{-i})$$

Thus, $a^i(\bar{\theta}^{-i}) \geqslant 0$, and so, from (3.54),

$$(3.61) \quad a^i(\bar{\theta}^{-i}) = 0$$

Furthermore, from feasibility, if there exists j for which $a^j(\bar{\theta}^{-j})$ is not defined, then there exists i for which (3.59) and hence (3.60) hold with strict inequality, an impossibility. Thus, (3.61) holds for all i and all $\bar{\theta} \geqslant 0$. But this implies that the incentive scheme is the nth positional dictatorship. Q.E.D.

We turn finally to the issue of coalitions. The following result generalizes Theorem 3.5.

Theorem 3.14: There exists no weakly efficient, regular CDSIC incentive scheme.

Proof: See Appendix.

3.4 *Other solution concepts*

In Subsections 3.2 and 3.3 we dealt exclusively with the solution concept of dominant strategies. In this subsection, we briefly treat Bayesian and Nash equilibria in our simple public project model.

Turning first to Bayesian equilibrium, we suppose that it is common knowledge that the θ^i's are distributed according to the (cumulative) probability distribution $F(\theta^1, \ldots, \theta^n)$. Knowing the prior F and his own parameter θ^i, agent i has beliefs given by the posterior distribution $F^i(\theta^{-i} \mid \theta^i)$. We shall suppose that the distribution of θ^{-i} is, in fact, independent of θ^i, so that we may write $F^i(\theta^{-i})$. (For a treatment of the dependent case, see the last section of the chapter by Laffont and Maskin, 1979*a*.) An incentive scheme (x, t) is *incentive-compatible in Bayesian strategies* (BSIC) if and only if

$$\int [\theta^i x(\theta^i, \theta^{-i}) + t^i(\theta^i, \theta^{-i})]\, dF^i(\theta^{-i}) \geqslant \int [\theta^i x(\hat{\theta}^i, \theta^{-i}) + t^i(\hat{\theta}^i, \theta^{-i})]\, dF^i(\theta^{-i})$$

for all i, θ^i, and $\hat{\theta}^i$. That is, telling the truth maximizes an agent's expected utility, given that others tell the truth.

In the corollary to Theorem 3.3 we demonstrated that there is no fully optimal DSIC scheme. One advantage that the Bayesian approach to incentive has is that full optimality is attainable.

Theorem 3.15: (See the work of Arrow, 1979, and d'Aspremont and Gérard-Varet, 1979.) There exist fully optimal BSIC incentive schemes.

Proof: The proof is by explicit example. (For a characterization of all such schemes, see the work of Laffont and Maskin, 1979*a*.) Take

$$t^i(\theta) = \int_{\theta^{-i}} \sum_{j \neq i} \theta^j x(\theta)\, dF^i(\theta^{-i}) - \frac{1}{n-1} \sum_{k \neq i} \int_{\theta^{-k}} \sum_{j \neq k} \theta^j x(\theta)\, dF^k(\theta^{-k}) \tag{3.73}$$

where x, of course, satisfies (3.3). By construction,

$$\sum_{i=1}^{n} t^i(\theta) = 0$$

Therefore the scheme is balanced. In maximizing his expected utility, agent i can ignore the second term in (3.73) because it does not depend on θ^i. Thus agent i chooses $\hat{\theta}^i$ to maximize

$$(3.74)\quad \int\left[\theta^i x(\hat{\theta}^i,\theta^{-i}) + \int_{\theta^{-i}} \sum_{j\neq i} \theta^j x(\hat{\theta}^i,\theta^{-i})\, dF^i(\theta^{-i})\right] dF^i(\theta^{-i})$$

After rearrangement, (3.65) becomes

$$(3.75)\quad \int_{\theta^{-i}} \left(\sum_{j=1}^{n} \theta^j x(\hat{\theta}^i,\theta^{-i})\right) dF^i(\theta^{-i})$$

But for each θ^{-i}, $\hat{\theta}^i=\theta^i$ maximizes $\sum_{j=1}^{n} \theta^j x(\hat{\theta}^i,\theta^{-i})$. Therefore, $\hat{\theta}^i=\theta^i$ maximizes (3.66). Q.E.D.

We discussed some of the drawbacks of Bayesian incentive theory in Section 1. Bayesian equilibrium demands both stronger behavioral assumptions and stronger informational assumptions (most notably, the assumption that F is common knowledge) than dominant strategy equilibrium. Furthermore, as we shall now see, Bayesian incentive schemes are plagued by multiple equilibria.

Suppose that $n=2$ and that

$$\begin{aligned} F^i(\theta^i) &= 0, \quad \theta^i \leqslant -1 \\ &= \frac{\theta^i + 1}{2}, \quad -1 \leqslant \theta^i \leqslant 1 \\ &= 1, \quad \theta^i \geqslant 1 \end{aligned}$$

That is, the distribution of θ^i is uniform on the interval $[-1,1]$. From the proof of Theorem 3.13, we know that the strategy rules $(\bar{\mu}^1(\cdot),\bar{\mu}^2(\cdot))$, where

$$\bar{\mu}^i(\theta^i) = \theta^i \quad \text{for all } \theta^i$$

form a Bayesian equilibrium in the incentive scheme defined by (3.3) and (3.64). However, there is a continuum of other equilibria. For any $k \geqslant 1$, define

$$\begin{aligned} \mu_k^1(\theta^1) &= -1, \quad k\theta^1 < -1 \\ &= k\theta^1, \quad -1 \leqslant k\theta^1 \leqslant 1 \\ &= 1, \quad k\theta^1 > 1 \end{aligned}$$

and

$$\mu_k^2(\theta^2) = \frac{1}{k}\theta^2$$

The pair (μ_k^1, μ_k^2) is an equilibrium. To see this, first consider agent 1. Agent 1 chooses $\hat\theta^1$ to maximize

$$\theta^1 \int_{\mu_k^2(\theta^2) \geq -\hat\theta^1} dF^1(\theta^2) + \int_{\theta^2 \geq -\hat\theta^1} \theta^2 dF^2(\theta^2)$$
$$= \theta^1\left(1 - \left(\frac{-k\hat\theta^1 + 1}{2}\right)\right) + \frac{1}{4}(1 - (\hat\theta^1)^2)$$

The first-order condition for an interior maximum is, therefore,

$$\frac{k\theta^1}{2} - \frac{\hat\theta^1}{2} = 0 \quad \text{or} \quad \hat\theta^1 = k\theta^1$$

Thus the optimal choice of $\hat\theta^1$ is $k\theta^1$, if $-1 \leqslant k\theta^1 \leqslant 1$, and one of the two endpoints, otherwise. The argument is similar for agent 2.

This example is symptomatic of Bayesian equilibrium in our public project model. For continuous distributions F^i, there will, in general, be continua of equilibria.

Finally, we turn to Nash equilibrium. When Nash equilibrium is the solution concept, we can no longer take agents' strategy spaces to coincide with their parameter spaces Θ^i. This is because, as discussed in Section 1, an agent's relevant information consists not only of his own parameter but of those of others as well. We therefore define an incentive scheme $\delta = (x, t)$ on the domain

$$S^1 \times \cdots \times S^n$$

For each profile $(\theta^1, \ldots, \theta^n)$, let $NE_d(\theta^1, \ldots, \theta^n)$ be the set of Nash equilibrium outcomes for profile $(\theta^1, \ldots, \theta^n)$ in the scheme d. That is,

$$NE_d(\theta^1, \ldots, \theta^n) = \{(x(s), t(s)) \mid s \in \Pi S^i \text{ is a Nash equilibrium for } (\theta^1, \ldots, \theta^n) \text{ in the scheme } d\}$$

$NE_d(\cdot)$ is a correspondence from Θ to $[0,1] \times \mathbf{R}^n$. We will call a correspondence $f: \Theta \to [0,1] \times \mathbf{R}^n$ *monotonic* if, for all $(\theta^1, \ldots, \theta^n)$, $(\theta^{1\prime}, \ldots, \theta^{n\prime}) \in \Theta$, and all $(x, t) \in f(\theta^1, \ldots, \theta^n)$

$$[\forall (x', t') \forall i \quad \theta^i x + t^i \geqslant \theta^i x' + t^{i\prime} \to \theta^{i\prime} x + t^i \geqslant \theta^{i\prime} x' + t^{i\prime}] \to (x, t) \in f(\theta^{1\prime}, \ldots, \theta^{n\prime})$$

The following result is drawn from the work of Maskin (1977).

Theorem 3.16: For correspondence $f: \Theta \to [0,1] \times \mathbf{R}^n$ there exists an incentive scheme d such that

(3.76) $\forall\theta \quad NE_d(\theta) = f(\theta)$

if and only if f is monotonic.

An incentive scheme d satisfying (3.76) is said to *implement* f. In our public project framework it is simple to characterize those correspondences that are implementable (monotonic):

Theorem 3.17: f is monotonic, and hence implementable, if and only if

$$(0,t) \in f(\theta^1,\ldots,\theta^n) \quad \text{implies} \quad (0,t) \in f(\theta^{1\prime},\ldots,\theta^{n\prime})$$
$$\text{for all} \quad (\theta^{1\prime},\ldots,\theta^{n\prime}) \leqslant (\theta^1,\ldots,\theta^n)$$

and

$$(1,t) \in f(\theta^1,\ldots,\theta^n) \quad \text{implies} \quad (1,t) \in f(\theta^{1\prime},\ldots,\theta^{n\prime})$$
$$\text{for all} \quad (\theta^{1\prime},\ldots,\theta^{n\prime}) \geqslant (\theta^1,\ldots,\theta^n)$$

Proof: Immediate verification. We can consider f as a welfare criterion. Theorem 3.16 shows that implementability places little restriction on welfare criteria in this framework. In particular,

$$f(\theta^1,\ldots,\theta^n) = \left\{(x,t) \mid t = 0 \quad \text{and} \quad x = \begin{cases} 0, & \sum\theta^i < 0 \\ 1, & \sum\theta^i \geqslant 0 \end{cases}\right\}$$

is implementable. Thus, with Nash equilibrium as the solution concept, we can use the utilitarian public decision rule incentive compatibly without making any transfers at all in equilibrium.

APPENDIX

We collect here the proofs of those theorems left unproved in the text.

Theorem 3.3: There exists no feasible and successful incentive scheme (x,t) such that for all θ

$$\left|\sum t^i(\theta)\right| \leqslant \left|\sum \bar{t}^i(\theta)\right|$$

with strict inequality for some θ, where $\bar{t}^i$ is the pivotal mechanism's transfer to agent i.

Proof: We confine our attention to the case $n=3$ and to schemes whose transfer functions are piecewise differentiable. Suppose that the scheme (x,t) uniformly dominates the Groves-Clarke mechanism, $(x,\bar{t})$. From Theorem 3.2, t^i satisfies (3.4) from some choice of $h^i(\theta^{-i})$. If θ is such that $\sum_{j\neq i}\theta^j<0$ for all i, then because $\sum_{i=1}^3 \bar{t}^i(\theta)=0$, we have

$$(3.5)\qquad \sum_{i=1}^{3} h^i(\theta^{-i}) = 0$$

Thus, if the h^i's are differentiable at θ, we have

$$\frac{\partial h^1}{\partial \theta^2} = -\frac{\partial h^3}{\partial \theta^2}$$

$$(3.6)\qquad \frac{\partial h^1}{\partial \theta^3} = -\frac{\partial h^2}{\partial \theta^3}$$

$$\frac{\partial h^2}{\partial \theta^1} = -\frac{\partial h^3}{\partial \theta^1}$$

and

$$(3.7)\qquad \frac{\partial^2 h^1}{\partial \theta^2 \partial \theta^3} = \frac{\partial^2 h^2}{\partial \theta^1 \partial \theta^3} = \frac{\partial^2 h^3}{\partial \theta^1 \partial \theta^2} = 0$$

From (3.5)–(3.7), we conclude that

$$(3.8)\qquad \begin{aligned} h^1(\theta^2,\theta^3) &= f(\theta^2) + g(\theta^3), && \text{for } \theta^2 + \theta^3 < 0 \\ h^2(\theta^1,\theta^3) &= e(\theta^1) - g(\theta^3), && \text{for } \theta^1 + \theta^3 < 0 \\ h^3(\theta^1,\theta^2) &= -e(\theta^1) - f(\theta^2), && \text{for } \theta^1 + \theta^2 < 0 \end{aligned}$$

where e, f, and g are piecewise differentiable. Similarly, because $\sum_{i=1}^{3} \bar{t}^i(\theta^{-i}) = 0$ when $\sum_{j\neq i} \theta^j \geqslant 0$ for all i, we have

$$2(\theta^1 + \theta^2 + \theta^3) + \sum_{i=1}^{3} h^i(\theta^{-i}) = 0$$

and so

$$(3.9)\qquad \begin{aligned} h^1(\theta^2,\theta^3) &= \bar{f}(\theta^2) + \bar{g}(\theta^3) - 2\theta^2, && \text{for } \theta^2 + \theta^3 \geqslant 0 \\ h^2(\theta^1,\theta^3) &= \bar{e}(\theta^1) - \bar{g}(\theta^3) - 2\theta^3, && \text{for } \theta^1 + \theta^3 \geqslant 0 \\ h^3(\theta^1,\theta^2) &= -\bar{e}^2(\theta^1) - \bar{f}(\theta^2) - 2\theta^1, && \text{for } \theta^1 + \theta^2 \geqslant 0 \end{aligned}$$

Now suppose that, for some $\hat{\theta}$,

$$\left|\sum t^i(\hat{\theta})\right| < \left|\sum \bar{t}^i(\hat{\theta})\right|$$

Then

$$(3.10)\qquad \sum_{i=1}^{3} \bar{h}^i(\hat{\theta}^{-i}) < \sum_{i=1}^{3} h^i(\hat{\theta}^{-i})$$

where $\bar{h}^i(\theta^{-i}) = -\sum_{j\neq i} \theta^j$ for $\sum_{j\neq i} \theta^j \geqslant 0$ and 0 otherwise. Clearly, there exist i and k such that $\sum_{j\neq i} \hat{\theta}^j > 0$ and $\sum_{j\neq k} \hat{\theta}^j < 0$. In particular, assume that $\hat{\theta}^1+\hat{\theta}^2>0$, $\hat{\theta}^2+\hat{\theta}^3<0$, and $\hat{\theta}^1+\hat{\theta}^3<0$ (the other cases can be argued similarly). Then (3.10) becomes

$$(3.11)\quad -\hat{\theta}^1+\hat{\theta}^2<f(\hat{\theta}^2)+g(\hat{\theta}^3)+e(\hat{\theta}^1) - g(\hat{\theta}^3)-\bar{e}(\hat{\theta}^1)-\bar{f}(\hat{\theta}^2)-2\hat{\theta}^2-2\hat{\theta}^1$$

If we take $\tilde{\theta}^1<\hat{\theta}^1$ such that $\tilde{\theta}^1+\hat{\theta}^2=0$, we have

$$(3.12)\quad \bar{h}^1(\hat{\theta}^2,\hat{\theta}^3)+\bar{h}^2(\tilde{\theta}^1,\hat{\theta}^3)+\bar{h}^3(\tilde{\theta}^1,\hat{\theta}^2) = h^1(\hat{\theta}^2,\hat{\theta}^3)+h^2(\tilde{\theta}^1,\hat{\theta}^3)+h^3(\tilde{\theta}^1,\hat{\theta}^2)$$

Thus from (3.11) and (3.12), there exists $\bar{\theta}^1$ with $\tilde{\theta}^1\leqslant\bar{\theta}^1\leqslant\hat{\theta}^1$ such that

$$\frac{\partial\bar{h}^2}{\partial\theta^1}(\bar{\theta}^1,\hat{\theta}^3)+\frac{\partial\bar{h}^3}{\partial\theta^1}(\bar{\theta}^1,\hat{\theta}^2)<\frac{\partial h^2}{\partial\theta^1}(\bar{\theta}^1,\hat{\theta}^3)+\frac{\partial h^3}{\partial\theta^1}(\bar{\theta}^1,\hat{\theta}^2)$$

that is,

$$(3.13)\quad 1<e'(\bar{\theta}^1)-\bar{e}'(\bar{\theta}^1)$$

where primes denote derivatives. Choose $\bar{\theta}^2$ and $\bar{\theta}^3$ such that

$$\bar{\theta}^2+\bar{\theta}^3<0,\quad \bar{\theta}^1+\bar{\theta}^2<0,\quad \text{and}\quad \bar{\theta}^1+\bar{\theta}^3=0$$

Then

$$-\bar{\theta}^1-\bar{\theta}^3=f(\bar{\theta}^2)+g(\bar{\theta}^3)-e(\bar{\theta}^1)-\bar{f}(\bar{\theta}^2)+\bar{e}(\bar{\theta}^1)-\bar{g}(\bar{\theta}^3)-2\bar{\theta}^3$$

so that

$$(3.14)\quad -1\leqslant -e'(\bar{\theta}^1)+\bar{e}'(\bar{\theta}^1)$$

Adding (3.13) to (3.14), we have $0<0$. Thus $\hat{\theta}$ cannot exist after all. Q.E.D.

Theorem 3.11: A weakly efficient, symmetric, deterministic, balanced, and regular DSIC scheme must be a positional dictatorship.

Proof: We shall argue the case $n=3$ and suppose that the boundary of A (a deterministic scheme is simple) is piecewise differentiable. From Theorem 3.9, if (x,t) satisfies the hypotheses of the theorem, then t^i is of the form

$$(3.27)\quad \begin{aligned} t^i(\theta) &= h^i(\theta^{-i}),\quad \theta\in A \quad\text{or}\quad a^i(\theta^{-i})\quad\text{undefined}\\ &= -a^i(\theta^{-i})+h^i(\theta^{-i}),\quad \theta\in A\quad\text{and}\quad a^i(\theta^{-i})\quad\text{defined}\end{aligned}$$

We shall suppose, for convenience, that the h^i's are continuous and piecewise differentiable. It will suffice to show that, for every θ in the boundary of A and every i, $a^i(\theta^{-i})$ is either zero or undefined (if

$a^i(\theta^{-i})=0$ and the number of nonnegative components of θ^{-i} is k, then, from symmetry, the incentive scheme is the $(k+1)$th positional dictator).

Suppose not. Choose $\bar{\theta}$ from the boundary of A so that a^1 is differentiable in a neighborhood of $\bar{\theta}^{-1}$ and $a^1(\bar{\theta}^{-1})\neq 0$. We may as well assume that $a^1(\bar{\theta}^{-1})<0$. Consider a sequence $\theta^1(n)$ converging from below to $\bar{\theta}^1$. From balance and (3.27),

$$h^1(\theta^2,\bar{\theta}^3) + h^2(\theta^1(n),\bar{\theta}^3) + h^3(\theta^1(n),\bar{\theta}^2) = 0 \quad \text{for all } n$$

Hence, by continuity,

$$\sum h^i(\bar{\theta}^{-i}) = 0$$

Thus, from balance,

$$\sum b^i(\bar{\theta}^{-i}) = 0 \tag{3.28}$$

where

$$\begin{aligned} b^i(\bar{\theta}^{-i}) &= a^i(\bar{\theta}^{-i}), \quad \text{if} \quad a^i(\bar{\theta}^{-i}) \quad \text{is defined} \\ &= 0, \quad \text{otherwise} \end{aligned}$$

We first show that a^1 must be constant in a neighborhood of $\bar{\theta}^{-1}$(i.e., that $\partial a^1/\partial\theta^2=\partial a^1/\partial\theta^3=0$ in a neighborhood of $\bar{\theta}^{-1}$). We must rule out two cases.

Case I: a^1 strictly decreasing in θ^2 but constant in θ^3 in a neighborhood N of $\bar{\theta}$.

Suppose first that a^3 is not defined in a neighborhood of $\bar{\theta}^{-3}$. Because a^1 is constant as a function of θ^3 in N, we may write a^1 as a function of θ^2 alone in this neighborhood. Hence, for all $\theta\in N$,

$$\theta \in A \quad \text{if and only if} \quad \theta^1 \geqslant a^1(\theta^2)$$

Furthermore, from (3.28),

$$b^3(a^1(\theta^2),\theta^2) = -a^1(\theta^2) - \theta^2$$

since $a^2(a^1(\theta^2),\theta^3)=\theta^2$. From arguments virtually identical with those in the proof of Theorem 3.3, we can conclude that for all $\theta\in N$,

$$h^2(\theta^1,\theta^3) = e(\theta^1) + g(\theta^3)$$

$$h^1(\theta^2,\theta^3) = f(\theta^2) - g(\theta^3)$$

$$\begin{aligned} h^3(\theta^1,\theta^2) &= -e(\theta^1) - f(\theta^2), \quad \theta^1 < a^1(\theta^2) \\ &= -e(\theta^1) - f(\theta^2) + (a^1)^{-1}(\theta^1) + a^1(\theta^2) \\ &\quad + b^3(\theta^1,\theta^2), \quad \theta^1 \geqslant a^1(\theta^2) \end{aligned} \tag{3.29}$$

where e, f, and g are continuous and piecewise differentiable. Choose $\tilde{\theta} \in N$ such that $\tilde{\theta}^1 > a^1(\tilde{\theta}^2)$ and such that e is differentiable at $\tilde{\theta}^1$, f at $\tilde{\theta}^2$, and h^2 and h^1 at $(\tilde{\theta}^1, \tilde{\theta}^2, \tilde{\theta}^1)$. By symmetry,

$$h^1(\tilde{\theta}^2, \tilde{\theta}^1) = h^3(\tilde{\theta}^3, \tilde{\theta}^2)$$

Thus

$$h^1(\tilde{\theta}^2, \tilde{\theta}^1) = h^3(\tilde{\theta}^1, \tilde{\theta}^2) = -e(\tilde{\theta}^1) - f(\tilde{\theta}^2) + (a^1)^{-1}(\tilde{\theta}^1) + a^1(\tilde{\theta}^2) + b^3(\tilde{\theta}^1, \tilde{\theta}^2) \tag{3.30}$$

If $(\tilde{\theta}^1, \tilde{\theta}^2, \tilde{\theta}^1) \in A$ (because $(\tilde{\theta}^1, \tilde{\theta}^2, \tilde{\theta}^1)$ need not be in N, we cannot infer from $\tilde{\theta}^1 > a^2(\tilde{\theta}^1)$ that $(\tilde{\theta}^1, \tilde{\theta}^2, \tilde{\theta}^1)$ is in A)

$$\begin{aligned} 0 &= \sum_i t^i(\tilde{\theta}^1, \tilde{\theta}^2, \tilde{\theta}^1) = -\sum a^i + \sum h^i \\ &= -a^2(\tilde{\theta}^1, \tilde{\theta}^1) - 2b^3(\tilde{\theta}^1, \tilde{\theta}^2) - 2e(\tilde{\theta}^1) - 2f(\tilde{\theta}^2) \\ &\quad + 2(a^1)^{-1}(\tilde{\theta}^1) + 2a^1(\tilde{\theta}^2) + 2b^3(\tilde{\theta}^1, \tilde{\theta}^2) + h^2(\tilde{\theta}^1, \tilde{\theta}^1) \\ &= -a^2(\tilde{\theta}^1, \tilde{\theta}^1) - 2e(\tilde{\theta}^1) - 2f(\tilde{\theta}^2) + 2(a^1)^{-1}(\tilde{\theta}^1) \\ &\quad + 2a^1(\tilde{\theta}^2) + h^2(\tilde{\theta}^1, \tilde{\theta}^1) \end{aligned} \tag{3.31}$$

where we have used the fact, from symmetry, that $a^1(\tilde{\theta}^2, \tilde{\theta}^1) = b^3(\tilde{\theta}^1, \tilde{\theta}^2)$. Differentiating (3.31) by θ^2, we obtain

$$0 = -2\frac{df}{d\theta^2}(\tilde{\theta}^2) + 2\frac{da^1}{d\theta^2}(\tilde{\theta}^2)$$

Because a^1 is strictly decreasing in θ^2, we have

$$\frac{df}{d\theta^2}(\tilde{\theta}^2) < 0$$

If $(\tilde{\theta}^1, \tilde{\theta}^2, \tilde{\theta}^1) \notin A$, then

$$0 = \sum h^i = -2e(\tilde{\theta}^1) - 2f(\tilde{\theta}^2) + 2(a^1)^{-3}(\tilde{\theta}^1) + 2a^1(\tilde{\theta}^2) + 2b^3(\tilde{\theta}^1, \tilde{\theta}^2) + h^2(\tilde{\theta}^1, \tilde{\theta}^1) \tag{3.32}$$

Differentiating (3.32) with respect to θ^2, we have

$$0 = -2\frac{df}{d\theta^2}(\tilde{\theta}^2) + 2\frac{da^1}{d\theta^2}(\tilde{\theta}^2) + 2\frac{\partial b^3}{\partial\theta^2}(\tilde{\theta}^1, \tilde{\theta}^2) \tag{3.33}$$

Because $\partial b^3/\partial\theta^2 \leqslant 0$, we conclude that, regardless of whether or not $(\tilde{\theta}^1, \tilde{\theta}^2, \tilde{\theta}^1) \in A$,

$$\frac{\partial f}{\partial\theta^2}(\tilde{\theta}^2) < 0 \tag{3.34}$$

Now choose $\hat{\theta} \in N$ such that $\hat{\theta}^2 = \tilde{\theta}^2$, $\hat{\theta}^1 < a^2(\hat{\theta}^2)$, and such that e is differentiable at $\hat{\theta}^1$ and h^2 at $(\hat{\theta}^1, \hat{\theta}^2, \hat{\theta}^1)$. From symmetry, $h^1(\hat{\theta}^2, \hat{\theta}^1) = h^3(\hat{\theta}^1, \hat{\theta}^2) = -e(\hat{\theta}^1) - f(\hat{\theta}^2)$. If $(\hat{\theta}^1, \hat{\theta}^2, \hat{\theta}^1) \notin A$, then

$$(3.35) \quad 0 = \sum h^i = h^2(\hat{\theta}^1, \hat{\theta}^1) - 2e(\hat{\theta}^1) - 2f(\hat{\theta}^2)$$

Differentiating (3.35) with respect to θ^2, we obtain

$$0 = \frac{df}{d\theta^2}(\hat{\theta}^2)$$

which contradicts (3.34), as $\hat{\theta}^2 = \tilde{\theta}^2$. Therefore, suppose $(\hat{\theta}^1, \hat{\theta}^2, \hat{\theta}^1) \in A$. But then,

$$\begin{aligned} & 0 = \sum t^i = -\sum a^i + \sum h^i \\ (3.36) \quad & \quad = -a^1(\hat{\theta}^2, \hat{\theta}^1) - a^2(\hat{\theta}^1, \hat{\theta}^1) - a^3(\hat{\theta}^1, \hat{\theta}^2) \\ & \quad\quad - 2e(\hat{\theta}^1) - 2f(\hat{\theta}^2) + h^2(\hat{\theta}^1, \hat{\theta}^1) \end{aligned}$$

Differentiating (3.36) with respect to θ^2, we obtain

$$(3.37) \quad 0 = -\frac{\partial a^1}{\partial \theta^2}(\hat{\theta}^2, \hat{\theta}^1) - \frac{\partial a^3}{\partial \theta^2}(\hat{\theta}^1, \hat{\theta}^2) - 2\frac{df}{d\theta^2}(\hat{\theta}^2)$$

But $\partial a^1/\partial\theta^2$ and $\partial a^3/\partial\theta^2$ are nonpositive, and so, from (3.37), we infer that $df/d\theta^2(\hat{\theta}^2)$ is nonnegative, contradicting (3.34). We conclude, therefore, that Case I is impossible.

Case II: a^1 is strictly decreasing in both θ^2 and θ^3 in a neighborhood N of $\bar{\theta}$.

Consider $\theta \in N$ on the boundary of A. Then $a^1(\theta^2, \theta^3) \leqslant \theta^1$. If the inequality holds strictly, then for $(\tilde{\theta}^2, \tilde{\theta}^3)$ slightly less than (θ^2, θ^3),

$$(3.38) \quad (a^1(\tilde{\theta}^2, \tilde{\theta}^3), \tilde{\theta}^2, \tilde{\theta}^3) < (\theta^1, \theta^2, \theta^3)$$

But $(a^1(\tilde{\theta}^2, \tilde{\theta}^3), \tilde{\theta}^2, \tilde{\theta}^3) \in A$. Thus, (3.38) contradicts the assumption that θ is on the boundary. Hence

$$(3.39) \quad a^1(\theta^2, \theta^3) = \theta^1$$

If, for $\tilde{\theta}^2 < \theta^2$, $(\theta^1, \tilde{\theta}^2, \theta^3) \in A$, then as a^1 is strictly decreasing θ^2, $a^1(\tilde{\theta}^2, \theta^3) > \theta^1$. But $a^1(\tilde{\theta}^2, \theta^3) \leqslant \theta^1$, as $(\theta^1, \tilde{\theta}^2, \theta^3) \in A$. Therefore,

$$(3.40) \quad a^2(\theta^1, \theta^3) = \theta^2$$

Similarly,

$$(3.41) \quad a^3(\theta^1, \theta^2) = \theta^3$$

Hence, from (3.28) and (3.39)–(3.41),

(3.42) $\theta^1 + \theta^2 + \theta^3 = 0$

Because (3.42) holds for all boundary points in N, we conclude that the incentive scheme is locally successful. From the same argument that establishes the corollary to Theorem 3.3, we thus generate a contradiction. Case II is therefore impossible, and we have verified that in the neighborhood N of $\bar{\theta}$, $\partial a^1/\partial\theta^2 = \partial a^1/\partial\theta^3 = 0$. We can conclude that $(\partial a^i/\partial\theta^j)(\theta^{-i}) = 0$ for any $i \neq j$ and θ^{-i} for which $\partial a^i/\partial\theta^j$ is defined.

Because $a^1(\bar{\theta}^2, \bar{\theta}^3) < 0$, (3.28) implies that either $a^2(\bar{\theta}^1, \bar{\theta}^3) > 0$ or $a^3(\bar{\theta}^1, \bar{\theta}^2) > 0$. Without loss of generality, assume the former. Because $\bar{\theta}$ is on the boundary of A, either $a^2(\bar{\theta}^1, \bar{\theta}^3) = \bar{\theta}^2$ or $a^3(\bar{\theta}^1, \bar{\theta}^2) = \bar{\theta}^3$. Assume the former (if $a^3(\bar{\theta}^1, \bar{\theta}^2) = \bar{\theta}^3$, the argument is entirely analogous). Consider points of the form $(\bar{\theta}^1, \bar{\theta}^2, \theta^3)$ for $\theta^3 \geqslant \bar{\theta}^3$. Let

$$\bar{\bar{\theta}}^3 = \min\{\theta^3 \geqslant \bar{\theta}^3 \mid a^2(\bar{\theta}^1, \theta^3) \neq a^2(\bar{\theta}^1, \bar{\theta}^3)\}.$$

Because $\bar{\theta}^2 = a^2(\bar{\theta}^1, \theta^3)$ for all $\theta^3 < \bar{\bar{\theta}}^3$, $(\bar{\theta}^1, \bar{\theta}^2, \theta^3)$ is on the boundary of A for such θ^3. Therefore, $(\bar{\theta}^1, \bar{\theta}^2, \bar{\bar{\theta}}^3)$ is on the boundary of A. Thus, from (3.28),

(3.43) $a^1(\bar{\theta}^2, \bar{\bar{\theta}}^3) + a^2(\bar{\theta}^1, \bar{\bar{\theta}}^3) + a^3(\bar{\theta}^1, \bar{\theta}^2) = 0$

But from the definition of $\bar{\bar{\theta}}^3$, $a^1(\bar{\theta}^2, \bar{\bar{\theta}}^3) \leqslant a^1(\bar{\theta}^2, \bar{\theta}^3)$ and $a^2(\bar{\theta}^1, \bar{\bar{\theta}}^3) < a^2(\bar{\theta}^1, \bar{\theta}^3)$, a contradiction of (3.43). Therefore, $\bar{\bar{\theta}}^3$ does not exist, and, for all $\theta^3 \geqslant \bar{\theta}^3$, $a^2(\bar{\theta}^1, \theta^3) = \bar{\theta}^2$. Choose $\theta^3 > 0$. By similar argument, $a^2(\theta^1, \theta^3) = \bar{\theta}^2$ for all $\theta^1 \geqslant \bar{\theta}^1$. Choose $\theta^1 > 0$. Then $(\theta^1, \bar{\theta}^2, \theta^3)$ lies on the boundary of A because $\bar{\theta}^2 = a^2(\theta^1, \theta^3)$. But $(\theta^1, \bar{\theta}^2, \theta^3) > 0$, so by weak efficiency it lies in the interior of A. We conclude that $a^1(\bar{\theta}^2, \bar{\theta}^3)$ cannot differ from zero after all. Q.E.D.

Theorem 3.14: There exists no weakly efficient, regular CDSIC incentive scheme.

Proof: Suppose that (x, t) is a weakly efficient regular CDSIC scheme. It is convenient to suppose that (x, t) is simple. For $\theta \notin A$, the payoff to the coalition of agents 1 and 2 is

(3.62) $$\theta^1 x^-(\theta) - \int_0^{\theta^1} s \frac{\partial x^-}{\partial\theta^1}(s, \theta^{-1})\, ds + \theta^2 x^-(\theta) - \int_0^{\theta^2} s \frac{\partial x^-}{\partial\theta^2}(s, \theta^{-2})\, ds + h^1(\theta^{-1}) + h^2(\theta^{-2})$$

Given that the coalition chooses θ^1 optimally, (3.62) implies

$$\theta^2 \frac{\partial x^-}{\partial \theta^1}(\theta) - \int_0^{\theta^2} s \frac{\partial^2 x^-}{\partial \theta^1 \partial \theta^2}(s, \theta^{-2})\, ds + \frac{\partial h^2}{\partial \theta^1}(\theta^{-2}) = 0 \tag{3.63}$$

Because (3.63) holds for all θ^2 locally, we obtain

$$\frac{\partial x^-}{\partial \theta^1} = 0 \tag{3.64}$$

Thus

$$\frac{\partial x^-}{\partial \theta^j} = 0 \quad \text{for all } j \tag{3.65}$$

For $\theta \in A$, the coalition of agents 1 and 2 has payoff

$$\begin{aligned}(\theta^1 + \theta^2) x^+(\theta) &- \int_0^{\theta^1} s \frac{\partial x^+}{\partial \theta^1}(s, \theta^{-1})\, ds - \int_0^{\theta^2} s \frac{\partial x^+}{\partial \theta^2}(s, \theta^{-2})\, ds \\ &+ \int_0^{a^1(\theta^{-1})} (x^-(s, \theta^{-1}) - x^+(s, \theta^{-1}))\, ds \\ &+ \int_0^{a^2(\theta^{-2})} (x^-(s, \theta^{-2}) - x^+(s, \theta^{-2}))\, ds \\ &+ h^1(\theta^{-1}) + h^2(\theta^{-2})\end{aligned} \tag{3.66}$$

If a^2 is differentiable as a function of θ^1, optimal choice of θ^1 in (3.66) implies

$$\begin{aligned}\theta^2 \frac{\partial x^+}{\partial \theta^1} &= \int_0^{\theta^2} s \frac{\partial^2 x^+}{\partial \theta^1 \partial \theta^2}(s, \theta^{-2})\, ds + (x^-(a^2(\theta^{-2}), \theta^{-2}) \\ &\quad - x^+(a^2(\theta^{-2}), \theta^{-2})) \frac{\partial a^2}{\partial \theta^1}(\theta^{-2}) + \frac{\partial h^2}{\partial \theta^1}(\theta^{-2}) \\ &= 0\end{aligned} \tag{3.67}$$

Because (3.67) holds locally for all θ^2,

$$\frac{\partial x^+}{\partial \theta^1} = 0$$

Therefore,

(3.68) $$\frac{\partial x^+}{\partial \theta^i} = 0 \quad \text{for all } i$$

From weak efficiency, (3.65) and (3.68) imply that $x^- \equiv 0$ and $x^+ \equiv 1$. That is, the scheme is simple. Choose θ on the boundary of A. Then $\theta^i = a^i(\theta^{-i})$ for some i. Suppose, without loss of generality, that $\theta^1 = a^1(\theta^{-1})$. Then, for $\hat{\theta}^1 < a^1(\theta^{-1})$, the payoff to a coalition of 1 and 2 for parameters $(\hat{\theta}^1, \theta^{-1})$ is

(3.69) $$h^1(\theta^{-1}) + h^2(\hat{\theta}^1, \theta^{-1-2})$$

Because θ^2 is chosen optimally,

$$\frac{\partial h^1}{\partial \theta^2}(\theta^{-1}) = 0$$

For $\hat{\theta}^1 \geqslant a^1(\theta^2, \theta^3)$, the payoff to the coalition is

$$\hat{\theta}^1 + \theta^2 - a^1(\theta^{-1}) - a^2(\hat{\theta}^1, \theta^{-1-2}) + h^1(\theta^{-1}) + h^2(\hat{\theta}^1, \theta^{-1-2})$$

Optimal choice of θ^2 implies

(3.70) $$-\frac{\partial a^1}{\partial \theta^2} + \frac{\partial h^1}{\partial \theta^2} = 0$$

Thus, from (3.60) and (3.61),

$$\frac{\partial a^1}{\partial \theta^2}(\theta^{-1}) = 0$$

Similarly,

$$\frac{\partial a^1}{\partial \theta^j}(\theta^{-1}) = 0 \quad \text{for all} \quad j \neq 1$$

That is, a^1 is locally constant. Now consider $\hat{\theta}^1$ slightly less than θ^1. Because it is optimal for the coalition of 1 and 2 to play truthful strategies, we have, when parameters are θ,

$$\theta^1 + \theta^2 - a^1(\theta^{-1}) - a^2(\theta^{-2}) + h^1(\theta^{-1}) + h^2(\theta^{-2}) \geqslant h^1(\theta^{-1}) + h^2(\hat{\theta}^1, \theta^{-1-2})$$

Therefore,

(3.71) $$\theta^2 - a^2(\theta^{-2}) + h^2(\theta^{-2}) \geqslant h^2(\hat{\theta}^1, \theta^{-1-2})$$

Similarly, if the true parameters are $(\hat{\theta}^1, \theta^{-1})$, $h^1(\theta^{-1}) + h^2(\hat{\theta}^1, \theta^{-1-2}) \geqslant \hat{\theta}^1 + \theta^2 - \theta^1 - a^2(\theta^{-2}) + h^1(\theta^{-1}) + h^2(\theta^{-2})$, so that

$$(3.72) \quad h^2(\hat{\theta}^1, \theta^{-1-2}) \geqslant \theta^2 + (\hat{\theta}^1 - \theta^1) - a^2(\theta^{-2}) + h^2(\theta^{-2})$$

Inequalities (3.71) and (3.72) together imply that

$$\theta^2 - a^2(\theta^{-2}) \geqslant h^2(\hat{\theta}^1, \theta^{-1-2}) - h^2(\theta^{-2}) \geqslant \theta^2 + (\hat{\theta}^1 - \theta^1) - a^2(\theta^{-2})$$

which is impossible, because the middle expression does not depend on θ^2. Thus the scheme cannot be CDSIC after all. Q.E.D.

NOTES

1 There are many other synonyms as well.
2 The term "members of society" may be misleading. Incentives theory applies to many purely "private" situations as well (e.g., the employer-employee relationship).
3 Like "planner" and "agent," "incentive scheme" goes under a variety of different names, depending on the area of application. For example, the term "contract" is often used in work on insurance, whereas "mechanism" applies in the allocation literature, and "voting scheme" or "game form" applies in the social choice context.
4 An incentive scheme is, in effect, a promise by the planner to react in a specified way to what agents do or reveal. The literature does not generally consider how the promise is enforced.
5 For an alternative model that is, in fact, somewhat more general, see the work of Myerson (1980).
6 It may seem peculiar that θ should enter the planner's payoff function, because it was assumed to be unobservable. We have included θ to allow, for example, the planner's objective function to be a social welfare function whose arguments are individual agents' utilities. Of course, if θ is unobservable, the planner can only maximize the expectation of his payoff with respect to θ.
7 In principle, each individual could attach a different subjective distribution to θ, and these distributions could jointly be common knowledge.
8 However, some recent progress in laying the theoretical foundations of "experimental" Nash equilibrium has been made by Levine (1981).
9 An example of this possibility is provided in Dasgupta and associates (1979, p. 195).
10 A domain of preferences Θ^i is "rich" if and only if for all $\theta_1, \theta_2 \in \Theta^i$ and all $d_1, d_2 \in D$ for which d_1 is preferred (strictly preferred) to d_2 under θ_1 implies that d_1 is preferred (strictly preferred) to d_2 under θ_1, there exists θ_3 such that for all c and $j = 1, 2$, d_j is preferred to c under θ_3, if d_j is preferred to c under θ_j.
11 The nonveto property requires that if all agents, except possibly one, prefer d to all other decisions, then $d \in f(\theta)$.
12 An allocation rule is individually rational if it assigns allocations that all agents prefer to their initial endowments.

13 Actually, the strong equilibria coincide with the Nash equilibria in Schmeidler's scheme, as in the work on double implementation.

14 Hurwicz (1972) and J. Roberts (1979) have shown that truthful behavior cannot constitute a global Nash equilibrium. However, Champsaur and Laroque (1980) have demonstrated that if the procedure is truncated at time T, then the global Nash equilibrium allocations converge to Lindahl equilibria as T tends to infinity. Truchon (1980) has shown that by introducing thresholds in the adjustment of public goods, a large class of *global* Nash equilibria leads to individually rational Pareto-optimal outcomes.

15 Green and Stokey (1980*b*) considered a similar model; but where the planner cannot commit himself to a scheme in advance. Their paper, therefore, does not fit within our framework.

16 We assume here either that the project is costless or that θ^i is the consumer's willingness to pay *net* of his share of the cost (the rule for dividing the costs is taken to be exogenous).

17 We have not allowed for random transfers because the space of transfers is convex and consumers are risk-neutral with respect to the private good. Thus, according to the reasoning of Section 1.4, there is no need to randomize transfers.

REFERENCES

Arrow, K. J. 1951. *Social Choice and Individual Values.* Cowles Foundation Monograph 12.

Arrow, K. J. 1979. "The Property Rights Doctrine and Demand Revelation under Incomplete Information" in *Economics and Human Welfare,* edited by M. Boskin, New York: Academic Press.

Barberá, S. 1977. "Manipulation of Social Decision Functions." *Journal of Economic Theory* 15:266–78.

Barberá, S. 1977*b*. "Manipulation of Voting Schemes that Do Not Leave too Much to Chance." *Econometrica,* 15:1573–88.

Bennett, E., and Conn, D. 1977. "The Group Incentive Properties of Mechanisms for the Provision of Public Goods." *Public Choice* 24-2:95–102.

Champsaur, P. 1976. "Neutrality of Planning Procedures in an Economy with Public Goods." *Review of Economic Studies* 43:293–300.

Champsaur, P., Drèze, J., and Henry, C. 1977. "Stability Theorems with Economic Applications." *Econometrica* 45:273–94.

Champsaur, P., and Laroque, G. 1980. "Strategic Behavior in Decentralized Planning Procedures." Unpublished mimeograph.

Chichilnisky, G., and Heal, G. 1980*a*. "Public Decision Rules Implementable by Non-cooperative Games: A Characterization." Unpublished mimeograph.

Chichilnisky, G., and Heal, G. 1980*b*. "Games to Prevent Free Riding: Existence and Construction." Unpublished mimeograph.

Clarke, E. 1971. "Multipart Pricing of Public Goods." *Public Choice* 8:19–33.

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

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

Diamond, P., and Mirrlees, J. 1971. "Optimal Taxation and Public Production, I." *American Economic Review* 61:261–78.

Drèze, J., and de la Vallée Poussin, D. 1971. "A Tâtonnement Process for Public Goods." *Review of Economic Studies* 38:133–50.
Dubins, L. 1974. "Group Decision Devices." Unpublished mimeograph.
Dutta, B., and Pattanaik, P. K. 1978. "On Nicely Consistent Voting Systems." *Econometrica* 46:163–70.
Farquharson, R. 1969. *Theory of Voting*, New Haven: Yale University Press.
Ferejohn, J., Grether, D., and McKelvey, R. 1980. "Implementation of Democratic Social Choice Functions." Unpublished mimeograph.
Fugigaki, Y., and Sato, K. 1981. "Incentives in the Generalized MDP Procedure for the Provision of Public Goods." *Review of Economic Studies* 48:473–86.
Gärdenfors, P. 1976. "Manipulation of Social Choice Functions." *Journal of Economic Theory* 13:217–28.
Gibbard, A. 1973. "Manipulation of Voting Schemes: A General Result." *Econometrica* 41:587–602.
Gibbard, A. 1977. "Manipulation of Schemes that Mix Voting with Chance." *Econometrica* 45:665–82.
Goldman, S., Leland, H., and Sibley, D. 1977. "Optimal Nonuniform Prices." Unpublished mimeograph.
Green, J. 1979. "Statistical Decision Theory Requiring Incentives for Information Transfer." Unpublished mimeograph.
Green, J. 1980. "Wage-Employment Contracts." Unpublished mimeograph.
Green, J., and Laffont, J. J. 1976. "Révélation des Préférences pour les Biens Publics." *Cahiers du Séminaire d'Econométrie.*
Green, J., and Laffont, J. J. 1977. "Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods." *Econometrica* 45:427–38.
Green, J., and Laffont, J. J. 1979*a*. *Incentives in Public Decision-Making.* Amsterdam: North-Holland.
Green, J., and Laffont, J. J. 1979*b*. "On Coalition Incentive Compatibility." *Review of Economic Studies* 46:243–254.
Green, J., and Stokey, N. 1980*a*. "The Value of Social Information in the Delegation Problem." Unpublished mimeograph.
Green, J., and Stokey, N. 1980*b*. "A Two-Person Game of Information Transmission." Unpublished mimeograph.
Grossman, S., and Hart, O. 1980. "An Analysis of the Principal-Agent Problem." Unpublished mimeograph.
Grossman, S., and Hart, O. 1981. "Implicit Contracts, Moral Hazard, and Unemployment." *American Economic Review*, 71:301–7.
Groves, T. 1973. "Incentives in Teams." *Econometrica* 41:617–31.
Groves, T., and Ledyard, J. 1977. "Optimal Allocation of Public Goods: A Solution to the Free Rider Problem." *Econometrica* 45:783–810.
Groves, T., and Loeb, M. 1975. "Incentives and Public Inputs." *Journal of Public Economics* 4:311–26.
Guenerie, R., and Laffont, J. J. 1979. "Regulating a Monopolist." Unpublished mimeograph.
Hammond, P. 1979. "Straightforward Individual Incentive Compatibility in Large Economies." *Review of Economic Studies* 46:63–82.
Harris, M., and Raviv, A. 1978. "Allocation Mechanisms and the Design of Auctions." Unpublished mimeograph.

Harris, M., and Raviv, A. 1979. "Optimal Incentive Contracts with Imperfect Information." *Journal of Economic Theory* 20:231-59.

Harris, M., and Raviv, A. 1981. "A Theory of Monopoly Pricing Schemes with Demand Uncertainty." *American Economic Review* 71:347-69.

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

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

Henry, C. 1977. "The Free Rider Problem in the MDP Procedure," unpublished mimeograph.

Holmstrom, B. 1979*a*. "Groves Schemes on Restricted Domains." *Econometrica,* 47:1137-1144.

Holmstrom, B. 1979*b*. "Moral Hazard and Observability." *Bell Journal of Economics* 10:74-91.

Holt, C. 1980. "Competitive Bidding for Contracts under Alternative Auction Procedures." *Journal of Political Economy* 88:433-45.

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

Hurwicz, L. 1975. "On the Existence of Allocation Systems whose Manipulative Nash Equilibria are Pareto Optimal." Unpublished mimeograph.

Hurwicz, L. 1979*a*. "Outcome Functions Yielding Walrasian and Lindahl Allocations at Nash Equilibrium Points." *Review of Economic Studies* 46:217-25.

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

Hurwicz, L., Postlewaite, A., and Maskin, E. 1979. "Feasible Implementation of Social Choice Correspondences by Nash Equilibria." Unpublished mimeograph.

Hurwicz, L., and Schmeidler, D. 1978. "Outcome Functions which Guarantee the Existence and Pareto Optimality of Nash Equilibria." *Econometrica* 46:1447-74.

Hurwicz, L., and Shapiro, L. 1978. "Incentive Structures Maximizing Residual Gain under Incomplete Information." *Bell Journal of Economics* 9:180-91.

Kalai, E., and Müller, E. 1977. "Characterization of Domains Admitting Nondictatorial Social Welfare Functions and Non Manipulatible Voting Procedures." *Journal of Economic Theory* 16:456-69.

Kalai, E., Postlewaite, A., and Roberts, J. 1977. "A Group Incentive Compatible Mechanism Yielding Core Allocations." Unpublished mimeograph.

Kelly, J. 1977. "Strategy-Proofness and Social Choice Functions without Single-valuedness." *Econometrica* 45:439-46.

Laffont, J. J., and Maskin, E. 1979*a*. "A Differential Approach to Expected Utility Maximizing Mechanisms." In: *Aggregation and Revelation of Preferences,* edited by J. J. Laffont, pp. 289-308. Amsterdam: North-Holland.

Laffont, J. J., and Maskin, E. 1979*b*. "On the Difficulty of Attaining Distributional Goals with Imperfect Information." *Scandinavian Journal of Economics* 81:227-37.

Laffont, J. J., and Maskin, E. 1980*a*. "A Differential Approach to Dominant Strategy Mechanisms." *Econometrica* 48:1507–20.

Laffont, J. J., and Maskin, E. 1980*b*. "Nash and Dominant Strategy Impenetration in Economic Environments." Unpublished mimeograph.

Laffont, J. J., and Maskin, E. 1980*c*. "A Characterization of SLIIC Planning Procedures with Public Goods." Unpublished mimeograph.

Ledyard, J. 1977. "Incentive Compatibility and Incomplete Information." Unpublished mimeograph.

Levine, D. 1981. *The Enforcement of Collusion in Oligopoly,* unpublished MIT Ph.D. dissertation.

Malinvaud, E. 1971. "A Planning Approach to the Public Good Production." *Swedish Journal of Economics* 73:96–112.

Malinvaud, E. 1972. "Prices for Individual Consumption, Quantity Indicators for Collective Consumption." *Review of Economic Studies* 39:385–405.

Marschak, J., and Radner, R. 1972. *Economic Theory of Teams.* Cowles Foundation Monograph 22.

Mas-Colell, A. 1978. "An Axiomatic Approach to the Efficiency of Noncooperative Equilibrium in Economics with a Continuum of Traders." Unpublished mimeograph.

Maskin, E. 1976. "Social Welfare Functions on Restricted Domains." Unpublished mimeograph.

Maskin, E. 1977. "Nash Equilibrium and Welfare Optimality." Unpublished mimeograph.

Maskin, E. 1979*a*. "Implementation and Strong Nash Equilibrium." In: *Aggregation and Revelation of Preferences,* edited by J. J. Laffont, pp. 433–49, Amsterdam: North-Holland.

Maskin, E. 1979*b*. "Incentive Schemes Immune to Group Manipulation." Unpublished mimeograph.

Maskin, E. 1980*a*. "On First Best Taxation." In *Income Distribution: the Limits to Redistribution,* edited by D. Collard, R. Lecomber, and M. Slater, pp. 9–22. Bristol: Scientechnica.

Maskin, E. 1980*b*. "Randomization in Incentive Problems." Unpublished mimeograph.

Maskin, E. and Riley, J. 1980*a*. "Monopoly Selling Strategies when Information is Incomplete, Price Discrimination, and Bundling." Unpublished mimeograph.

Maskin, E., and Riley, J. 1980*b*. "Auctioning an Indivisible Object." Unpublished mimeograph.

Mirrlees, J. 1971. "An Exploration in the Theory of Optimum Income Taxes." *Review of Economic Studies,* 38:175–208.

Mirrlees, J. 1975. "The Theory of Moral Hazard and Unobservable Behaviour —Part I" Unpublished mimeograph.

Moulin, H. 1979. "Prudence versus Sophistication in Voting Strategy." Unpublished mimeograph.

Moulin, H. 1980*a*. "Implementing Efficient, Anonymous, and Neutral Social Choice Functions." *Journal of Mathematical Economics* 7:249–69.

Moulin, H. 1980*b*. "On Strategy Proofness and Single Peakedness." *Public Choice* 35:437–55.

Moulin, H., and Peleg, B. 1980. "Stability and Implementation of Effectivity Functions." Unpublished mimeograph.

Myerson, R. 1979. "Incentive Compatibility and the Bargaining Problem." *Econometrica* 47:61–73.
Myerson, R. 1981*a*. "Optimal Coordination Mechanisms in Principal Agent Problems." Unpublished mimeograph.
Myerson, R. 1981*b*. "Optimal Auction Design." *Mathematics of Operations Research* 6:58–73.
Pattanaik, P. K. 1975. "Strategic Voting without Collusion under Binary and Democratic Group Decision Rules." *Review of Economic Studies* 42:93–104.
Pattanaik, P. K. 1976. "Counter-threats and Strategic Manipulation under Voting Schemes." *Review of Economic Studies* 43:11–18.
Pattanaik, P. K., and Sengupta, M. 1977. "Restricted Preferences and the Strategy-Proofness of a Class of Group Decision Rules," unpublished mimeograph.
Peleg, B. 1978*a*. "Consistent Voting Systems." *Econometrica* 46:153–61.
Peleg, B. 1978*b*. "Representation of Simple Games by Social Choice Functions." *International Journal of Game Theory* 7:81–94.
Postlewaite, A. 1979. "Manipulation via Endowments." *Review of Economic Studies* 44:255–62.
Radner, R. 1980. "Monitoring Cooperative Agreements in a Repeated Principal-Agent Relationship." Unpublished mimeograph.
Riley, J., and Samuelson, W. 1981. "Optimal Auctions." Unpublished mimeograph.
Roberts, J. 1979. "Incentives in Planning Procedures for the Provision of Public Goods." *Review of Economic Studies* 55:283–92.
Roberts, J., and Postlewaite, A. 1976. "The Incentives for Price-taking Behavior in Large Exchange Economies." *Econometrica* 44:115–28.
Roberts, K. 1979*a*. "The Characterization of Implementable Rules." In: *Aggregation and Revelation of Preferences*, edited by J. J. Laffont, pp. 321–48. Amsterdam: North-Holland.
Roberts, K. 1979*b*. "Welfare Considerations of Nonlinear Pricing," *Economic Journal* 89:66–83.
Rosenthal, R. 1979. "Arbitration of Two-part Disputes under Uncertainty." *Review of Economic Studies* 45:595–604.
Ross, S. 1973. "The Economic Theory of Agency: The Principal's Problem." *American Economic Review* 63:134–9.
Rubinstein, A., and Yaari, M. 1980. "Repeated Insurance Contracts and Moral Hazard," unpublished mimeograph.
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. 1976. "Straightforward Allocation Mechanisms." Unpublished mimeograph.
Satterthwaite, M., and Sonnenschein, H. 1981. "Strategy Proof Allocation Mechanisms." *Review of Economic Studies* 48:587–98.
Schmeidler, D. 1980. "Walrasian Analysis via Strategic Outcome Functions," *Econometrica* 48:1585–94.
Schoumaker, F. 1977. "Best Reply Strategies in the Champsaur, Drèze and Henry Procedure." Unpublished mimeograph.
Selten, R. 1975. "Reexamination of the Perfectness Concept for Equilibrium

Points in Extensive Games." *International Journal of Game Theory* 4:25-55.

Sengupta, M. 1979. "Implementable Social Choice Rules: Characterization and Correspondence Theorems under Strong Nash Equilibrium." Unpublished mimeograph.

Shavell, S. 1979*a*. "On Moral Hazard and Insurance." *Quarterly Journal of Economics* 93:541-62.

Shavell, S. 1979*b*. "Risk Sharing and Incentives in the Principal and Agent Relationship." *Bell Journal of Economics* 10:55-73.

Smets, H. 1972. "Le Principe de la Compensation Réciproque: Un Instrument Economique pour la Solution de Certains Problèmes de Pollution Transfrontière." Paris: O.C.D.E., Direction de l'Environment.

Spence, M. 1977. "Nonlinear Prices and Welfare." *Journal of Public Economics.*

Stiglitz, J. 1974. "Incentives and Risk Sharing in Sharecropping." *Review of Economic Studies* 61:219-56.

Thomson, W. 1979. "Maximin Strategies and the Elicitation of Preferences." In: *Aggregation and Revelation of Preferences,* edited by J. J. Laffont, pp. 245-68. Amsterdam: North-Holland.

Thomson, W. 1980. "On the Manipulability of Resource Allocation Mechanisms." Unpublished mimeograph.

Tideman, N. 1972. "The Efficient Provision of Public Goods." In: *Public Prices for Public Products,* edited by S. Mushkin, Washington: The Urban Institute.

Truchon, M. 1980. "Implementation of Pareto Optima with the MDP Procedure when Consumers are Non-Myopic." Unpublished mimeograph.

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

Walker, M. 1977. "An Informationally Efficient Auctioneerless Mechanism for Attaining Lindahl Allocations." Unpublished mimeograph.

Walker, M. 1980. "On the Impossibility of a Dominant Strategy Mechanism to Optimally Decide Public Questions." *Econometrica* 48:1521-40.

Weitzman, M. 1976. "The New Soviet Incentive Model." *Bell Journal of Economics* 7:251-7.

Wilson, R. 1978. "Competitive Exchange." *Econometrica* 46:577-85.