ON THE DIFFICULTY OF ATTAINING DISTRIBUTIONAL GOALS WITH IMPERFECT INFORMATION ABOUT CONSUMERS*

Jean-Jacques Laffont

Ecole Polytechnique, Paris and Université des Sciences Sociales de Toulouse, France

Eric Maskin

Massachusetts Institute of Technology, Cambridge, Mass., USA

Abstract

Recently, mechanisms which overcome the free rider problem and achieve Pareto optimality under imperfect information have been constructed. In this paper we provide various impossibility theorems which show the difficulty of achieving distributional goals when consumers' tastes are unknown. The results are developed for a particular game theoretic solution concept, that of dominant strategy; they could be extended if, instead, Bayesian equilibrium were the solution concept. As a way out we propose a second-best approach to welfare optimization.

I. Introduction

Recently a solution to the free rider problem proposed by Vickrey, Clarke and Groves has received a great deal of attention.[1] Mechanisms have been constructed to elicit private information; in particular, individuals' preferences for public goods.[2] These mechanisms use the information to optimize the welfare criterion corresponding to the sum of individuals' willingnesses to pay. In a model of public goods and a single private good (say, money) where utility functions are additively separable between the public and private goods and where all agents have the same constant marginal utility of money, this social objective function yields the same allocation of public goods as the Pareto optimum in which all individuals' utilities are given equal weight. Because our distribution goals may not be strictly utilitarian, however, we may be interested in optimizing social welfare functions (SWF) which are not simply sums of

* This work has been supported by the Sonderforschungsbereich 21 and the National Science Foundation.

[1] See, for example, Green & Laffont (1978) and Laffont (1979).

[2] Economies with a single private good are considered. True tastes are elicited as dominant strategies in the revelation game set up by the mechanisms.

utilities. In particular, we may wish to consider SWF's which take account of an agent's ability to pay as well as his willingness to pay for a public project. This paper investigates the extent to which this and other distributional goals are attainable. The results are, for most part, negative.

In Section II, we set up a framework in which to study these issues. Section III provides a general impossibility theorem which shows that under imperfect information no strictly concave Bergsonian SWF can be optimized. This result leaves us with essentially just the linear SWF's. Section IV extends this result to SWF's which are more general than the Bergsonian variety. It also elucidates the difficulties of optimizing SWF's—even linear SWF's—which incorporate information about abilities to pay for public projects. Section V illustrates comparable difficulties for the use of information which is not taste related.

The results of this paper are developed for a particular game theoretic solution concept—that of dominant strategies. We note in the conclusion that the results would all go through if, instead, Bayesian equilibrium were the solution concept; see Harsanyi (1968). As a way out of the pessimism of this paper, we propose a second-best approach to welfare optimization which we hope to pursue in future work.

II. The Model

We consider an economy with n $(n \geqslant 2)$ consumers, indexed by $i=1, \ldots, n$, and two commodities, one public and one private. The utility function of consumer i, $u_i(K, x_i)$, is additively separable between the public good K and the private good x_i, $i=1, \ldots, n$, and furthermore each agent is assumed to have the same constant marginal utility of private good. Hence,

$$u_i(K, x_i) = v_i(K) + x_i.^1$$

The family of utility functions is further restricted by:

Assumption 1. For $i=1, \ldots, n$ let Θ_i be an open interval of $\mathbf{R}$ and let v_i: $\overset{0}{\mathbf{R}}_+ \times \Theta_i \to \mathbf{R}$ be a continuously differentiable function such that for any $\theta \in \Theta = \prod_{i=1}^n \Theta_i$, for any $\lambda = (\lambda_1, \ldots, \lambda_n)$ such that $\sum_{i=1}^n \lambda_i = 1$ and $\lambda_i \geqslant 0$, $i=1, \ldots, n$, there exists $K^*(\theta) \in \overset{0}{\mathbf{R}}_+$ for which

(i) $\sum_{i=1}^n \lambda_i v_i(K^*(\theta), \theta_i) = \max_{K \in R_+} \sum_{i=1}^n \lambda_i v_i(K, \theta_i)$

(ii) $K^*(\theta)$ is continuously differentiable.

An agent is characterized by his valuation function, v_i, his taste characteristic, θ_i, and a vector, $\eta_i \in H_i$, of welfare relevant characteristics other than

[1] $v_i(K)$ is the *net* willingness to pay for the public good; i.e., the willingness to pay less the imputed cost share.

his tastes for the public good; η_i might represent, for example, endowment or productivity.

The functional forms $v_i(\cdot,\cdot)$ are assumed to be known publicly, but the true value $\hat{\theta}_i$ of the parameter θ_i is known only to agent i, *a priori*. Similarly, $\hat{\eta}_i$, the true value of η_i, is, at the beginning, strictly private information. A mechanism is a procedure where agents announce messages, on the basis of which a public good level is chosen. The purpose of a mechanism is to determine an "optimal" level of the public good. An optimal level is one which maximizes a given social welfare function. This paper studies the class of social welfare functions which can be optimized by mechanisms in which agents announce *characteristics* as messages and where revelation of true characteristics is a dominant strategy. A mechanism where agents announce characteristics (not necessarily their true characteristics) as strategies is a revelation mechanism.

A mechanism is formally defined as a mapping, $f=(d,t)$, from the strategy spaces $\prod_{i=1}^n H_i \times \prod_{i=1}^n \Theta_i$ into $\overset{0}{\mathbf{R}}_+ \times \mathbf{R}^n$, composed of a decision function, $d(\cdot)$, with range $\overset{0}{\mathbf{R}}_+$ and a n-tuple of transfer functions, $t(\cdot)=[t_1(\cdot), ..., t_n(\cdot)]$, each with range $\mathbf{R}$. $d(\cdot)$ associates to any $2n$-tuple of announced parameters a quantity of public good, while $t_i(\cdot)$ associates a transfer of private good to agent i, $i=1, ..., n$.

A revelation mechanism is said to be C^1 when the function $f(\cdot)$ is continuously differentiable.

A revelation mechanism, $f(\cdot)=[d(\cdot), t(\cdot)]$ is said to be *strongly individually incentive compatible* (s.i.i.c.) if the truth is dominant strategy for each consumer, that is, if, for any i, any $(\eta,\theta)\in\prod_{j=1}^n H_j \times \prod_{j=1}^n \Theta_j$, and any $(\hat{\eta}_i,\hat{\theta}_i)\in H_i\times\Theta_i$

$$\begin{aligned} &v_i(d(\hat{\eta}_i,\eta_{-i},\hat{\theta}_i,\theta_{-i}),\hat{\theta}_i)+t_i(\hat{\theta}_i,\theta_{-i},\hat{\eta}_i,\eta_{-i}) \\ &\qquad \geqslant v_i(d(\eta_i,\eta_{-i},\theta_i,\theta_{-i}),\hat{\theta}_i)+t_i(\theta_i,\theta_{-i},\eta_i,\eta_{-i}), \end{aligned}$$

where

$$\theta_{-i}=(\theta_1, ..., \theta_{i-1}, \theta_{i+1}, ..., \theta_n)$$

and

$$\eta_{-i}=(\eta_1, ..., \eta_{i-1}, \eta_{i+1}, ..., \eta_n).$$

A social welfare function is a real valued function F of $[v_1(K,\hat{\theta}_1), ..., v_n(K,\hat{\theta}_n), t_1, ..., t_n, \hat{\theta}_1, ..., \hat{\theta}_n, \hat{\eta}_1, ..., \hat{\eta}_n]$.

We say that a social welfare function F is implementable if there exists a s.i.i.c. mechanism whose outcome maximizes F when everyone tells the truth.

From Green & Laffont (1978), we know that the social welfare functions $\sum_{i=1}^n \lambda_i v_i(K,\hat{\theta}_i)$ where the weights λ_i $(\sum_{i=1}^n \lambda_i=1, \lambda_i>0, i=1, ..., n)$ are constants, are implementable by the following obvious generalization of the Groves mechanisms:

$$d(\theta) \text{ maximizes } \sum_{i=1}^{n} \lambda_i v_i(K, \theta_i) \quad \text{in } K$$

$$t_i(\theta) = \frac{1}{\lambda_i} \sum_{j \neq i} \lambda_j v_j(d(\theta), \theta_j)$$

III. Bergsonian Social Welfare Functions

By a Bergsonian SWF, we mean a real-valued function of $v_1, ..., v_n, t_1, ..., t_n$. This is actually a more general formulation than the usual definition, which makes F a function only of $v_1 + t_1, ..., v_n + t_n$. We shall demonstrate that strictly concave F's are not implementable.

We adopt the following assumptions.

Assumption 2. F is twice continuously differentiable and its matrix of second order partial derivatives with respect to the v_j's is negative definite. Furthermore $\partial F / \partial v_i > 0$ for all i.

To prove an impossibility theorem, we can work with a small set of valuation functions, since any superset will then lead to impossibility *a fortiori*.

Consider the class V^Q of valuation function n-tuples consisting of functions which are quadratic with constant term:

$$V^Q = \left\{ \theta_1 K - \frac{K^2}{2} + \alpha_1, \ldots, \theta_n K - \frac{K^2}{2} + \alpha_n \,\middle|\, \theta \in \Theta \right.$$

$$\left. = \prod_i \Theta_i, \alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbf{R}^n, K \in \overset{0}{\mathbf{R}}_+ \right\}$$

$$= \{v_1(K, \theta_1, \alpha_1), \ldots, v_n(K, \theta_n, \alpha_n) \,|\, \theta \in \Theta, \alpha \in \mathbf{R}^n, K \in \mathbf{R}\},$$

where Θ_i is a bounded open interval of the positive real line.

Assumption 3. $\forall \theta \in \Theta$, $\forall \alpha \in \mathbf{R}^n$, there exists a unique continuously differentiable $K^*(\theta, \alpha)$ which maximizes $F(v_1(K, \theta_1, \alpha_1), ..., v_n(K, \theta_n, \alpha_n), t_1(\theta, \alpha), ..., t_n(\theta, \alpha))$.

Theorem 1. *Under Assumptions 2 and 3, there exists no implementable Bergsonian SWF.*

Proof. We first parameterize v_i by rewriting it as

$$\bar{v}_i = \bar{\theta}_i K - \frac{K^2}{2} + \beta_i \bar{\theta}_i + \gamma_i, \quad \text{where } \bar{\theta}_i, \beta_i, \gamma_i \in \mathbf{R}$$

The class of valuation functions $\{\bar{v}_i\}$ is the same as the class $\{v_i\}$ so that nothing substantive has changed. Take

$\theta_i = (\bar{\theta}_i, \beta_i, \gamma_i)$.

Suppose that F is a Bergsonian SWF satisfying Assumptions 2 and 3 and which is implemented by the mechanism $(K^*(\theta), t_1, ..., t_n)$. Agent i's maximization problem is

$$\max_{\theta_i} \bar{v}_i(K^*(\theta), \hat{\theta}_i) + t_i(\theta).$$

The first-order condition in $\bar{\theta}_i$ is

$$\frac{\partial t_i}{\partial \bar{\theta}_i}(\theta) = -\frac{\partial \bar{v}_i}{\partial K}(K^*(\theta), \hat{\theta}_i)\frac{\partial K^*}{\partial \bar{\theta}_i}.$$

Evaluated at the truth, $\bar{\theta}_i = \hat{\theta}_i$, the second-order condition in $\bar{\theta}_i$ is:

$$\frac{-\partial^2 \bar{v}_i}{\partial K \partial \bar{\theta}_i}(K^*(\hat{\theta}_i, \theta_{-i}), \hat{\theta}_i)\frac{\partial K^*}{\partial \bar{\theta}_i}(\hat{\theta}_i, \theta_{-i}) < 0. \tag{1}$$

Since K^* maximizes F, $\sum_j \frac{\partial F}{\partial v_j}\frac{\partial \bar{v}_j}{\partial K} = 0$ when $K = K^*$.

Differentiating this last condition with respect to $\bar{\theta}_i$ and solving for $\partial K^*/\partial \bar{\theta}_i$, we obtain

$$\frac{\partial K^*}{\partial \bar{\theta}_i} = -\frac{\frac{\partial F}{\partial v_i}\frac{\partial^2 \bar{v}_i}{\partial K \partial \bar{\theta}_i} + \sum_{j=1}^{n}\frac{\partial \bar{v}_j}{\partial K}\frac{\partial^2 F}{\partial v_j \partial v_i}\frac{\partial \bar{v}_i}{\partial \bar{\theta}_i} + \sum_j \sum_s \frac{\partial \bar{v}_j}{\partial K}\frac{\partial^2 F}{\partial v_j \partial t_s}\frac{\partial t_s}{\partial \bar{\theta}_i}}{\sum_{j=1}^{n}\frac{\partial F}{\partial v_j}\frac{\partial^2 \bar{v}_j}{\partial K^2} + \sum_r \sum_s \frac{\partial \bar{v}_r}{\partial K}\frac{\partial^2 F}{\partial v_r \partial v_s}\frac{\partial v_s}{\partial K}} \tag{2}$$

Let $D_i(\theta)$ be the denominator of the right-hand side of (2). Using the quadratic specification of the valuation functions, (1) can be rewritten as:

$$\hat{\beta}_i G_i(\hat{\theta}_i, \theta_{-i}) + H_i(\hat{\theta}_i, \theta_{-i}) < 0,$$

where

$$G_i(\hat{\theta}_i, \theta_{-i}) = \frac{\sum_{j \neq i}(\hat{\theta}_j - K^*)\frac{\partial^2 F}{\partial v_j \partial v_i} + (\hat{\theta}_i - K^*)\frac{\partial^2 F}{\partial v_i^2}}{D_i(\hat{\theta}_i, \theta_{-i})}$$

$$H_i(\hat{\theta}_i, \theta_{-i}) = \frac{\frac{\partial F}{\partial v_i} + \sum_{j \neq i}(\theta_j - K^*)\left[\frac{\partial^2 F}{\partial v_j \partial v_i}K^* + \sum_j \sum_s \frac{\partial^2 F}{\partial v_j \partial t_s}\frac{\partial t_s}{\partial \bar{\theta}_i}\right]}{D_i(\hat{\theta}_i, \theta_{-i})} + \frac{(\hat{\theta}_i - K^*)\left[\frac{\partial^2 F}{\partial v_i^2}K^* + \sum_s \frac{\partial^2 F}{\partial v_i \partial t_s}\frac{\partial t_s}{\partial \bar{\theta}_i}\right]}{D_i(\hat{\theta}_i, \theta_{-i})}$$

Now choose $\theta \in \Theta$ so that, for some $t, u \in \{1, ..., n\}$, $\bar{\theta}_t \neq \bar{\theta}_u$. Then because the matrix of second partials $(\partial^2 F/\partial v_s \partial v_r)$ is non-singular, the numerator of $G_i(\theta)$

does not vanish for some i. For such i, choose $\tilde{\beta}_i \in \mathbf{R}$ such that $\tilde{\beta}_i G_i(\theta_i, \theta_{-i}) + H_i(\theta_i, \theta_{-i}) > 0$.

Select $\tilde{\gamma}_i \in \mathbf{R}$ so that $\tilde{\gamma}_i + \tilde{\beta}_i \theta_i = \gamma_i + \beta_i \theta_i$. This implies that $\tilde{\beta}_i G_i(\bar{\theta}_i, \tilde{\beta}_i, \tilde{\gamma}_i, \theta_{-i}) + H_i(\bar{\theta}_i, \tilde{\beta}_i, \tilde{\gamma}_i, \theta_{-i}) > 0$. So second-order conditions are violated at $(\bar{\theta}_i, \tilde{\beta}_i, \tilde{\gamma}_i, \theta_{-i})$, and, therefore, incentive compatibility cannot hold. Q.E.D.

IV. More General Social Welfare Functions

Theorem 1 demonstrates that if the family of possible valuation functions is large enough to include the quadratic class, one cannot implement Bergson SWF's which are strictly concave. This goes a long way toward ruling out non-linear SWF's. We next investigate the implementability of more general SWF's. That is, we shall consider SWF's which depend directly on the taste parameters θ. The results, however, are not much more positive than those for Bergson SWF's.

Consider again the class V^Q of quadratic valuation functions with constant terms. Let F be a real valued function of $v_1, ..., v_n, t_1, ..., t_n, \theta_1, ..., \theta_n, \alpha_1, ..., \alpha_n, K$, where $(v_1, ..., v_n) \in V^Q$.

Assumption 2'. F is twice continuously differentiable, and the matrices $(\partial^2 F/\partial v_i \partial v_j)$ and $(\partial^2 F/\partial \alpha_i \partial v_j)$ are negative definite. Furthermore $\partial F/\partial v_i > 0$ for all i.

Assumption 3'. $\forall \theta \in \Theta$, $\forall \alpha \in \mathbf{R}^n$, there exists a unique, continuously differentiable $K^*(\theta, \alpha)$ which maximizes $F(v_1(K, \theta_1, \alpha_1), ..., v_n(K, \theta_n, \alpha_n), t_1, ..., t_n, \theta_1, ..., \theta_n, \alpha_1, ..., \alpha_n, K)$.

Theorem 2. *No SWF satisfying Assumptions 2' and 3' is implementable.*

Proof. Almost identical to the proof of Theorem 1.

Theorem 2 is a strongly negative result but still does not bear on a large class of potentially implementable SWF's; viz., those that are linear in the valuation functions. For example, suppose the class of valuation functions is quadratic: $\theta_i K - (K^2/2)$, $\theta_i > 0$. Then the second order conditions for implementability becomes

$$\frac{\partial K^*}{\partial \theta_i}(\theta) > 0 \quad \forall \theta \, \forall i. \tag{4}$$

If (4) holds for K^*, then any SWF for which K^* is maximizing is implementable. In particular the following two-person SWF is implementable for v_1 and v_2 quadratic:

$$F \equiv \theta_2 v_1(K, \theta_1) + \theta_1 v_2(K, \theta_2).$$

The maximizing K for this SWF is

$$K^*(\theta) = \frac{2\theta_1 \theta_2}{\theta_1 + \theta_2},$$

the fact that $(\partial K^*/\partial \theta_i) = (2\theta_i^2/(\theta_1 + \theta_2))^2 > 0$ implies implementability.

The transfer functions of the implementing mechanism are of the form

$$t_1(\theta) = \int \frac{2\theta_1 \theta_2^2(\theta_2 - \theta_1)}{(\theta_1 + \theta_2)^3} d\theta_1 + h_1(\theta_2)$$

$$t_2(\theta) = \int \frac{2\theta_2 \theta_1^2(\theta_1 - \theta_2)}{(\theta_1 + \theta_2)^3} d\theta_2 + h_2(\theta_1) \tag{5}$$

It may easily be verified that this mechanism is not of the Clarke–Groves variety. That is, the transfers are not chosen so as to make the individual and social maximands coincide.

This example should not, however, make us too sanguine about the possibility of implementability in the linear case. From Robert's paper in Laffont (1979), we know that if the class of valuation functions is entirely unrestricted, then only the SWF's of the form $\sum_i \lambda_i v_i$, where the λ_i's are constants, can be implemented.

This suggests a general result: the larger the class of valuation functions is, the less freedom one has to have the weights depend on the taste parameters. The conjectured result is illustrated by the following example.

Consider the SWF $\sum_{i=1}^n \lambda_i(\theta_i) v_i(K, \theta_i)$ in the quadratic case. K^* is defined by

$$\sum_{i=1}^{n} \lambda_i(\theta_i)\,(\theta_i - K^*) = 0$$

or

$$K^*(\theta) = \frac{\sum_{j=1}^{n} \lambda_j(\theta_j)\,\theta_j}{\sum_{j=1}^{n} \lambda_j(\theta_j)}$$

$$\frac{\partial K^*}{\partial \theta_i} = \frac{(\sum \lambda_j)\,[\lambda_i(\theta_i) + \theta_i \lambda_i'(\theta_i)] - (\sum \lambda_j \theta_j)\,\lambda_i'(\theta_i)}{(\sum \lambda_j)^2} \tag{6}$$

$$\text{sign}\,\frac{\partial K^*}{\partial \theta_i} = \text{sign}\,\sum_j \lambda_i \lambda_j \left[1 + \frac{\lambda_i'}{\lambda_i}(\theta_i - \theta_j)\right]$$

Suppose that $\lambda_i'/\lambda_i \neq 0$. When the range of θ_j is small enough, condition (4) is satisfied. However, when this range increases, sign $\partial K^*/\partial \theta_i$ becomes negative for some values of θ, showing that as the class expands the SWF of the type chosen here becomes non-implementable.

As we noted in the introduction, one question of particular interest is whether it is possible to implement SWF's that take account of an individual's *ability* (rather than desire) to pay for a public good. Taking account of abilities is a simple matter if these are known; one can simply arrange lump-sum transfers accordingly. Eliciting this information may be difficult, however, as the following argument suggests. Suppose that agent i's preferences can be represented by the utility function $v_i(K, \theta_i) + \lambda_i(\eta_i)t_i$, where $\lambda_i(\cdot)$ is a known function of the publicly unknown characteristic η_i. Taking the private good as *numéraire*, λ_i is the marginal utility of money. One might interpret η_i as marketable endowments (whether in human or physical form). λ_i is presumably lower the more richly endowed is i and is, therefore, a useful index of ability to pay. From Theorem 2, we know there is little point in considering SWF's which are strictly concave functions of the v_j's; so consider SWF's of the form

$$\sum_{j=1}^{n} \phi_j(\eta_j)\, v_j(K, \theta_j)$$

where $\phi_j > 0$ for all j.

Assumption 4. λ_i and ϕ_i are strictly positive, continuously differentiable functions for all i and there is a unique twice continuously differentiable function $K^*(\theta, \eta)$ such that $K = K^*(\theta, \eta)$ maximizes $\sum_{j=1}^{n} \phi_j(\eta_j) v_j(K, \theta_j)$.

Assumption 5.

$$\frac{\partial^2 v_i}{\partial K\, \partial \theta_i}(K^*(\theta, \eta), \theta_i) \neq 0 \quad \forall \theta, \eta.$$

Theorem 3. *Under Assumptions* 1, 4, *and* 5, *an SWF of the form* $\sum_{i=1}^{n} \phi_i(\eta_i) v_i(K, \theta_i)$ *is implementable if and only if*

$$\phi(\eta_i) = \frac{A_i}{\lambda_i(\eta_i)},$$

where A_i is a positive constant.

Proof. Maximizing $\sum_j \phi_j(\eta_j) v_j(K, \theta_j)$ leads to the first-order condition.

$$\sum_j \phi_j(\eta_j) \frac{\partial v_j}{\partial K}(K^*(\theta, \eta), \theta_j) = 0 \tag{7}$$

Differentiating (7), we obtain

$$\frac{\partial K^*}{\partial \eta_i} = \frac{-\dfrac{\partial \phi_i}{\partial \eta_i}\dfrac{\partial v_i}{\partial K}}{\sum \phi_j \dfrac{\partial^2 v_j}{\partial K^2}} \quad \text{and} \quad \frac{\partial K^*}{\partial \theta_i} = \frac{-\phi_i \dfrac{\partial^2 v_i}{\partial K\, \partial \theta_i}}{\sum_j \phi_j \dfrac{\partial^2 v_j}{\partial K^2}} \tag{8}$$

Incentive compatibility implies the identities

$$\frac{\partial t_i}{\partial \theta_i} \equiv -\frac{1}{\lambda_i}\frac{\partial v_i}{\partial K}\frac{\partial K^*}{\partial \theta_i}$$

$$\frac{\partial t_i}{\partial \eta_i} \equiv -\frac{1}{\lambda_i}\frac{\partial v_i}{\partial K}\frac{\partial K^*}{\partial \eta_i},$$

since the first-order conditions must be satisfied for any θ_{-i} and any true parameters.

Using the fact that $\partial^2 t_i/(\partial\theta_i\partial\eta_i)$ must equal $\partial^2 t_i/(\partial\eta_i\partial\theta_i)$ (Young's theorem), we obtain

$$\frac{\dfrac{\partial \lambda_i}{\partial \eta_i}}{\lambda_i}\frac{\partial v_i}{\partial K}\frac{\partial K^*}{\partial \theta_i} = \frac{-\partial^2 v_i}{\partial K\partial\theta_i}\frac{\partial K^*}{\partial \theta_i}. \tag{9}$$

Using (8) and invoking Assumption 5, we can reduce (9) to

$$\frac{\dfrac{\partial \lambda_i}{\partial \eta_i}}{\lambda_i} = \frac{\dfrac{-\partial \phi_i}{\partial \eta_i}}{\phi_i}$$

or

$$\phi_i(\eta_i) = \frac{A_i}{\lambda_i(\eta_i)}.$$

It remains only to check the second-order conditions. Since $v_i(K, \theta_i)$ is strictly concave in K, the matrix of second-order derivatives of agent i's objective function is negative semi-definite, implying pseudo-concavity. Thus the first-order conditions suffice. Q.E.D.

Theorem 3 is a decidedly negative result for those who are equity minded. It states, in effect, that the only implementable SWF's which take into account ability to pay are those which weight richer agents more heavily than poorer.

V. On the Use of Non-Taste-related Information

Theorem 3 is discouraging as to the possibility of using information about endowments. As we formulated preferences, however, endowments entered the agent's utility function through his marginal utility of income. A natural inquiry is to ask whether characteristics which do not affect utility at all (but still may be relevant to social welfare) may be incorporated by an implementable SWF. Unfortunately, the answer is, once again, no, as we shall now demonstrate. In view of Theorem 1, we can immediately exclude SWF's which are strictly concave functions of the v_i's. Therefore, consider a linear SWF of the form:

$$F = \sum_i \lambda_i(\eta)\, v_i(K, \theta_i) \tag{10}$$

where $\eta = (\eta_1, \ldots, \eta_n)$, and η_i is a non-taste-related characteristic of individual i.

Assumption 6. Let V be a class of utility functions such that for all C^1-decision functions, $d(\eta, \theta)$ for which

$$\sum_{i=1}^{n} \lambda_i(\eta)\, v_i(d(\eta, \theta), \theta_i) = \max_{K>0} \sum_{i=1}^{n} \lambda_i(\eta)\, v_i(K, \theta),$$

$$\frac{\partial d}{\partial \eta_i}(\eta, \theta) \neq 0 \quad \text{and} \quad \frac{\partial^2 v_i}{\partial K\, \partial \theta_i}(d(\eta, \theta), \theta_i) \neq 0$$

for each i and almost any (η, θ).

The condition $\partial d/\partial \eta_i \neq 0$ guarantees that the optimal public decision is genuinely dependent on the parameters η_i, while $\partial^2 v_i/(\partial K \partial \theta_i) \neq 0$ ensures that the decision depends on the θ_i's. Assumption 6, therefore, serves only to rule out uninteresting cases.

Theorem 4. *If a class of valuation functions satisfies Assumptions* 1 *and* 6, *then if a SWF F on V satisfies* (10), *F is not implementable.*

Proof. Consider the maximization program of an agent i faced with a C^1-mechanism, $f(\cdot) = (d(\cdot), t(\cdot))$.

$$\max_{\theta_i, \eta_i} v_i(d(\eta, \theta), \hat{\theta}_i) + t_i(\eta, \theta)$$

The first-order conditions of this program are:

$$\frac{\partial v_i}{\partial K}(d(\eta_i, \eta_{-i}, \theta_i, \theta_{-i}), \hat{\theta}_i) \cdot \frac{\partial d}{\partial \eta_i}(\eta, \theta) + \frac{\partial t_i}{\partial \eta_i}(\eta, \theta) = 0$$

$$\frac{\partial v_i}{\partial K}(d(\eta, \theta), \hat{\theta}_i) \frac{\partial d}{\partial \theta_i}(\eta, \theta) + \frac{\partial t_i}{\partial \theta_i}(\eta, \theta) = 0$$

In order for the true values $(\hat{\eta}_i, \hat{\theta}_i)$ to be dominant strategies, the above conditions must hold as identities when evaluated at the truthful point. Thus,

$$\frac{\partial t_i}{\partial \eta_i}(\eta, \theta) \equiv \frac{-\partial v_i}{\partial K}(d(\eta, \theta), \theta_i) \cdot \frac{\partial d(\theta)}{\partial \eta_i}$$

$$\frac{\partial t_i}{\partial \theta_i}(\eta, \theta) \equiv \frac{-\partial v_i}{\partial K}(d(\eta, \theta), \theta_i) \cdot \frac{\partial d(\theta)}{\partial \theta_i}$$

The equality of the cross derivatives of $t_i(\eta, \theta)$ implies

$$\frac{\partial^2 v_i}{\partial K\, \partial \theta_i} \cdot \frac{\partial d(\theta)}{\partial \eta_i} \equiv 0$$

which is impossible from Assumption 6. Q.E.D.

VI. Conclusions

This paper demonstrates that, even for highly restricted domains of preferences, the possibilities of implementing SWF's other than a weighted sum (with constant weights) of marginal rates of substitution between public and private goods are highly limited. In particular, SWF's which incorporate elicited information about ability to pay rather than willingness to pay for a public good appear impossible, in general, to implement.

Throughout, the solution concept we have imposed for implementing mechanisms is that of dominant strategies. One avenue for attaining more optimistic results would seem to be to weaken the solution concept. As we argued elsewhere,[1] Bayesian equilibrium[2] does not generally permit a wider range of SWF's to be implemented than does the dominant strategy equilibrium. Indeed, the four theorems of this paper all go through when Bayesian equilibrium become the solution concept.

Alternatively, one might adopt Nash equilibrium as the solution concept. More work needs to be done to determine the possibilities in this case. Maskin (1977) has shown that any social welfare function satisfying the properties of monotonicity and no veto power can be implemented by a Nash mechanism. On the other hand, Roberts[3] has shown that when valuation functions are unrestricted and the social welfare function has a unique optimum, it can be Nash-implemented only if it is dominant–strategy–implementable.

An alternative direction, which we hope to explore in future work, is second-best optimization. The results of this paper show that only a limited class of social welfare functions may be implemented. If a SWF of interest falls outside this class, one can "partially" implement it by optimizing instead the implementable SWF which best approximates it (in an expected welfare sense). Obviously, the worse the approximation, the more partial the implementation.

References

Green, J. & Laffont, J. J.: *Incentives in public decision making.* North-Holland, Amsterdam, 1978.

Harsanyi, J.: Games with incomplete information played by Bayesian players, I–III. *Management Science,* 1968.

Laffont, J. J. (ed.): *Aggregation and revelation of preferences.* North-Holland, Amsterdam, 1979.

Maskin, E.: Nash equilibria and welfare optimality. Mimeo, 1977.

Sen, A.: On weights and measures: Informational constraints in social welfare analysis. *Econometrica 47,* 1539–1572, 1977.

---

[1] See Laffont & Maskin, "A differential approach to expected utility maximizing mechanisms" in Laffont (1979).
[2] Sen (1977) discusses the use of non-utility components of social welfare functions.
[3] In Laffont (1979).