A CHARACTERIZATION OF INTERIM EFFICIENCY WITH PUBLIC GOODS

By John O. Ledyard and Thomas R. Palfrey[1]

1. INTRODUCTION

In this paper, we consider the following classical public goods problem. A group of individuals must decide on a level of a public good that is produced according to constant returns to scale up to some capacity constraint. In addition to deciding the level of public good, the group must decide how to tax the individuals in the group in order to cover the cost. The distribution of the burden of taxation is important because different individuals have different marginal rates of substitution between the private good (taxes) and the public good, and may have different incomes as well. These individual marginal rates of substitution are private information; that is, each individual knows his or her own marginal rate of substitution, but not those of the other members of the group. Adopting a Bayesian mechanism design framework, we assume that the distribution of marginal rates of substitution is common knowledge.

We are interested in characterizing efficiency in this environment and are also interested in characterizing those mechanisms that one might expect to actually arise in practice. This suggests two approaches, one from normative considerations and one from positive considerations. On the normative side, we ask: What should an active planner (a mechanism designer) do? A well-known special case of this problem has been solved for one particular social welfare function (e.g., d'Aspremont and Gérard-Varet (1979)) that is insensitive to the distribution of cost shares. What distinguishes our work here is that we consider a planner who is maximizing a welfare function that is sensitive to the allocation of cost shares over the different valuation-types. Simply put, the planner may care who pays. This is represented formally by type-contingent welfare weights.

Why might the consideration of such distributional goals be relevant? What rationale can be given for nonconstant welfare weights? Perhaps the simplest example to answer these questions corresponds to public decisions with zero production costs. Such cases are well-approximated in the real world by social legislation such as blue laws, smoking and drinking prohibitions, clothing requirements at beaches, and so forth. Suppose one is considering the implementation of one such social regulation. Many would argue that if implementation takes place, then the losers (i.e. those with a negative valuation to the

proposed regulation) should be compensated.[2] But one runs into the (incentive compatibility) problem that if you naively say you are going to compensate all losers, then everyone will claim to be a loser, possibly leading to production never occurring. A planner might want, therefore, to give some weight to the losers but not to the exclusion of all others. Obviously, in order to compensate the losers in such decisions, incentive taxes need to be carefully constructed that will achieve such type-contingent redistribution, at least to the extent limited by incentive compatibility constraints. As we will show below, there is a direct and intuitive link between the desired degree of such compensation and the corresponding distortions away from the Lindahl-Samuelson optimum. In this particular example, significant compensation of losers would necessitate a corresponding degree of underproduction relative to the classic solution. Other weighting schemes would correspond to other type-distributional goals, and could lead to either under- or overproduction.

A second reason to consider nonconstant welfare weights arises if one concedes that this partial equilibrium model is embedded in a richer general equilibrium structure, where income or wealth distribution is a goal of the planner. If preferences for the public good are correlated with income or·wealth in a systematic way, then the public good mechanism can be used as an instrument for redistribution, and unequal welfare weights would be a reflection of the planner's redistributive goals.

A third rationale for unequal weights is more direct. For reasons that may have to do only remotely with issues of compensating losers or wealth redistribution, certain kinds of type-dependent cost-sharing may be deemed desirable on their own merits. A classic example of this is the class of proportional cost-sharing rules, whereby individuals valuing the public good more *should* bear a proportionally larger share of the costs (e.g., Jackson and Moulin (1992)). Such normative goals would correspond to a system of welfare weights that decrease in valuations in a particular way.

For the positive approach to the mechanism design problem, we ask: What would we expect to see in practice? Here we are looking for a concept of efficiency or stability because one would expect inefficient or unstable mechanisms to be replaced by others. Under complete information, these concepts correspond to Pareto optimality and the core, respectively. Under asymmetric information the problem is a bit more subtle, and there remains no true consensus on the appropriate equivalent concepts.[3] We therefore take a minimalist approach, and look at a natural extension of Pareto optimality to asymmetric information. In the analysis below we assume that all decisions, including whether to change the mechanism, are made at the interim stage—that is, when each agent knows his or her type, but not anyone else's type. If there is no communication, then the set of interim incentive efficient mechanisms consists of those incentive compatible mechanisms for which it cannot be common knowledge that there is another mechanism which generates a unanimous improvement. We would expect therefore that surviving institutions would be, minimally, interim incentive efficient.

Luckily we do not have to choose between normative and positive approaches to this problem. As pointed out in Holmström and Myerson (1983), a mechanism is interim efficient if and only if there exist type-dependent social welfare weights for which that mechanism solves the planner's optimization problem subject to feasibility and incentive compatibility constraints. Thus, by varying the welfare weights in our planner's problem,

---

[2]Sometimes, this requirement is implicitly imposed as a voluntary participation constraint.
[3]See Holmstrom and Myerson (1983), Cramton and Palfrey (1995), and Crawford (1985) for good discussions of the difficulties of extending these concepts to asymmetric information.

we map out the entire set of mechanisms that are interim incentive efficient. Thus, a complete solution to this problem, posed either from a normative or positive standpoint, is equivalent to fully characterizing the set of interim efficient mechanisms for the production of public goods in this framework.

A complete characterization of interim efficiency has been done for the special case where the types are identically distributed and can only take on two values (Ledyard and Palfrey (1994)). There it was shown that optimal production always takes a special form in which the public good is provided if and only if the number of high valuation types exceeds a threshold number that depends on the welfare weights and the distribution of types. The greater the welfare weight on high valuation types, the lower the optimal threshold. With more than two types (as in this paper) the optimal mechanism generally depends on the exact profile of types in a more complicated way. In this paper, we fully characterize interim efficient mechanisms and obtain some comparative statics about how the optimal mechanism changes with the underlying distribution of types and with the welfare weights of the welfare function.

Section 2 presents the basic model. Section 3 presents the characterization and its proof. Section 4 explains the intuition behind the characterization. In Section 5, we summarize these findings and offer some concluding remarks about some possible future directions of research.

2. THE MODEL

There are N people who must decide on the quantity, q, of a public good that is produced according to constant returns to scale[4] and has a maximum level $Y = 1$. The cost of producing $q \in [0,1]$ is equal to Kq. In addition, they must decide how to distribute the production costs. Because of the linear production technology, the optimal level of the public good will always be either 0 or 1, so this is equivalent to a problem of deciding on whether or not to produce a discrete public good. We let a^i denote individuals i's share of the cost, in units of the consumption of the private good, and assume it can take any real value. Therefore the set of feasible levels of production and cost shares are given by

$$(a^1, \ldots, a^N, q) \in \Re^N \times [0,1]$$

such that

$$Kq \le \sum_{i=1}^{N} a^i.$$

Individual *preferences* are assumed to be risk-neutral and quasilinear in the level of public good production and the taxes (cost shares), so the utility to type v^i of agent i for an allocation (q, a) is given by

$$V^i = v^i q - a^i.$$

Thus, v^i represents the marginal rate of substitution between the public and private good. We refer to v^i as player i's "value." We assume that each individual knows his own value, v^i, and does not know the values of the other individuals. We assume that the

---

[4] We explain in the next section how the model and the results are easily extended to arbitrary production technologies.

individual values (v^i) are independently distributed, with the (common knowledge) cdf of i's value denoted $F_i(\cdot)$ and the support of F_i is $V^i = [\underline{v}^i, \bar{v}^i]$, where $\underline{v}^i < K/N < \bar{v}^i$. We assume F_i has a continuous positive density on V^i. Note that $\underline{v}^i < 0$ is allowed.

Clearly under these assumptions, our choice of normalization of the utility function is arbitrary up to an affine transformation. In particular, it is equivalent (in terms of individual decision theory) to the models of asymmetric information about contribution costs (a^i), where utilities are normalized[5] so that the marginal utility of the public good (v) equals 1, so that $u^i = q - (1/v^i)a^i$. However, the class of ex-ante incentive efficient mechanisms (in the sense of Holmström and Myerson (1983)) will be different under the two normalizations.[6] So, below, we will focus on the set of interim-incentive efficient mechanisms. That set is independent of whatever (type dependent) normalization one chooses.

A *mechanism* consists of a message space for each agent and an outcome function mapping message profiles into probability distributions over the set of feasible allocations. By the revelation principle, the properties (in terms of allocations) of any optimal mechanism can be duplicated by an incentive compatible, direct mechanism in which the message space for agent i is simply the set of possible types (values) in the support of F_i. A strategy for i is a mapping $\sigma^i: V^i \rightarrow V^i$, that is, a decision rule that specifies a reported type for each possible type. We refer to the identity mapping as the truthful strategy. By the linearity of the individual utility functions, there is also no loss of generality in restricting attention to deterministic mechanisms. Thus, we denote a feasible direct mechanism simply as a function

$$\eta: V^N \rightarrow \left\{ (a^1, \ldots, a^N, q) \in R^N \times [0,1] \Big| \sum_{i=1}^{N} a^i \geq Kq \right\}.$$

We denote the public good allocation component of η at type profile v by $q(v)$, and the private good tax for i by $a^i(v)$.

Besides feasibility, the main restriction on η is that it be *incentive compatible*, which means that it is a Bayesian equilibrium of η for all agents to adopt a strategy of truthfully reporting their type. Given a strategy profile $\sigma^i: V^i \rightarrow V^i$ and a mechanism, η, let the interim utility of type v^i of agent i, assuming all others truthfully report their type, be denoted by:

$$\hat{u}^i(\eta, v^i, \sigma^i) = \int_{V^{-i}} [v^i q(\sigma^i(v^i), v^{-i}) - a^i(\sigma^i(v^i), v^{-i})] dF(v|v^i).$$

Let $u^i(\eta, v^i) \equiv \hat{u}^i(\eta, v^i, I)$ where I denotes the truthful strategy $I(v) = v$. Then η is incentive compatible if and only if $u^i(\eta, v^i) \geq \hat{u}^i(\eta, v^i, \sigma^i)$ for all v^i, σ^i.

The set of interim incentive efficient allocation rules[7] can be represented as the solutions to a set of maximization problems. Let $\lambda > 0$ be a system of welfare weights, a profile of measurable functions mapping types into the positive real line such that $\int_{\underline{v}^i}^{\bar{v}^i} \lambda_i(v^i) dF_i(v^i) = 1 \ \forall i$, where $\lambda_i(v^i)$ represents the welfare weight assigned to type v^i of

---

[5]This normalization can be made as long as $\underline{v}^i > 0$.

[6]The fact that ex ante efficiency is sensitive to utility normalizations is discussed in Ledyard and Palfrey (1994, p. 333).

[7]For the remainder, we simply refer to such allocations as "interim efficient."

agent i. Then η is interim efficient if and only if there is a λ such that η maximizes $\Sigma_i \int_{\underline{v}^i}^{\bar{v}^i} \lambda_i(v^i) u^i(\eta, v^i) dF_i(v^i)$ over the set of feasible and incentive compatible mechanisms.[8] We now proceed to characterize this set.

3. THE CHARACTERIZATION

As indicated above, we represent interim efficient rules as a solution to a constrained maximization problem. First we need to identify incentive compatible mechanisms in a useful way.

For smooth mechanisms, when preferences are linear, the characterization of incentive compatibility in terms of derivatives is well-known. There are basically two features of such mechanisms. First, an envelope condition is satisfied, namely that the total derivative of the interim utility for i with respect to type when players adopt truthful strategies is equal to the partial derivative with respect to type (i.e., fixing the reports of all agents). Second, the interim utility to i under truthful reporting is convex in i's type. This is stated formally below, without proof.

LEMMA (Rochet (1987)): *If $\hat{u}^i$ is linear in v^i and η is twice continuously differentiable, then η is incentive compatible if and only if*

(i) $\nabla_{v^i} u^i(\eta, v^i) = \nabla_{v^i} \hat{u}^i(\eta, v^i, I),$

(ii) $u^i(\eta, v^i)$ *is convex in* v^i.

For our problem $\nabla_{v^i} u^i(\eta, v^i) = Q_i(v^i) \equiv \int_{V^{-i}} q(v) dF(v|v^i)$. So u^i is convex in v^i if and only if $Q_i'(v^i) \geq 0 \ \forall v^i$. Using these facts we can see that a mechanism (q, a) is interim efficient if and only if there is a λ such that (q, a) solves $\max \int_V \Sigma_i \lambda_i(v^i)(v^i q(v) - a^i(v)) dF(v)$ subject to $0 \leq q(v) \leq 1 \ \forall v$, $Q_i'(v^i) \geq 0 \ \forall i, v^i$, $\nabla_{v^i} u^i(v^i) = \nabla_{v^i} \hat{u}^i(v^i, I) \ \forall i, v^i$, and $\Sigma_i a^i(v) = Kq(v) \ \forall v$.

Using the approach of Mirrlees (1971) and Wilson (1993) we construct the Lagrangian equivalent problem

$$\max_{\eta} \min_{\psi, \delta} \Sigma_i \int_{\underline{v}^i}^{\bar{v}^i} \lambda_i(v^i) u^i(\eta, v^i) dF_i(v^i)$$

$$+ \Sigma_i \int_{\underline{v}^i}^{\bar{v}^i} \psi_i(v^i) [u^i_{v^i}(\eta, v^i) - \hat{u}^i_{v^i}(\eta, v^i, I)] dv^i$$

$$+ \int_V \delta(v) \left[\Sigma_i a^i(v) - Kq(v) \right] dv \qquad \text{subject to:}$$

$$0 \leq q(v) \leq 1 \qquad \forall v \in V,$$
$$Q_i'(v^i) \geq 0 \qquad \forall i, v^i \in V^i,$$

where ψ_i and δ are multipliers for (first order) incentive compatibility and feasibility, respectively. Applying Green's Theorem and substituting the identity $u^i(\eta, v^i) =$

[8] See Holmstrom and Myerson (1983).

$\hat{u}^i(\eta, v^i, I)$ converts the maximization problem to:

$$\max_{\eta} \min_{\psi, \delta} \sum_i \int_{\underline{v}^i}^{\bar{v}^i} \{\hat{u}^i(\eta, v^i, I)[\lambda_i(v^i)f_i(v^i) - \psi_i'(v^i)]$$

$$- \psi_i(v^i)\hat{u}^i_{v^i}(\eta, v^i, I)\} dv^i + \int_V \delta(v)\left(\sum_i a^i(v) - Kq(v)\right) dv$$

$$+ \sum_i \int_{\partial V^i} \hat{u}^i(\eta, v^i, I)\psi_i(v^i)\xi_i(v^i)\, dv^i \qquad \text{subject to:}$$

$$0 \le q(v) \le 1 \qquad \forall v \in V,$$
$$Q_i'(v^i) \ge 0 \qquad \forall i, v^i \in V^i,$$

where ∂V^i denotes the boundary of V^i and ξ_i points outward at v^i.

We are now in a position to give a complete characterization of the class of interim efficient mechanisms.

THEOREM 1: (q^*, a^*) *is an interim efficient mechanism if and only if* $\exists \lambda \gg 0$ *with* $\int_{\underline{v}^i}^{\bar{v}^i} \lambda(v^i) dF^i(v^i) = 1$ $\forall i$, *such that*:

  (a) $\forall v, q^*(v)$ *maximizes* $\{\sum_i w^i(v^i) - K\}$ *subject to:*

$$0 \le q(v) \le 1 \qquad \forall v \in V,$$
$$Q_i'(v^i) \ge 0 \qquad \forall i, v^i \in V^i, \qquad \text{where}$$

$$w^i(v^i) = v^i - \frac{1 - F_i(v^i)}{f_i(v^i)} + \frac{\int_{v^i}^{\bar{v}^i} \lambda_i(t^i) dF_i(t^i)}{f_i(v^i)};$$

and (b) $a^{*i}(v) = \int_{\underline{v}^i}^{v^i} t^i dQ_i^*(t^i) + \alpha^i(v)$ *where*

$$\sum_i \alpha^i(v) = Kq^*(v) - \sum_i \int_{\underline{v}^i}^{v^i} t^i dQ_i^*(t^i) \qquad \forall v \qquad \text{and}$$

$$\frac{\partial}{\partial v^i} \int \alpha^i(v) dF(v|v^i) = 0 \qquad \forall i, v^i.$$

PROOF: A sketch is given. For further details see Ledyard and Palfrey (1996).

Notice that the restriction of λ to $\int_{\underline{v}^i}^{\bar{v}^i} \lambda_i(v^i) dF(v^i) = 1$ $\forall i$ is without loss of generality. Since utilities are linear in the transfers, for some welfare weights total welfare can be made arbitrarily large simply by making *ex ante* transfers from one individual to another individual. That is, if, for two agents i and j, it were the case that

$$\int_{\underline{v}^i}^{\bar{v}^i} \lambda_i(v^i) dF_i(v^i) < \int_{\underline{v}^j}^{\bar{v}^j} \lambda_j(v^j) dF_j(v^j),$$

then total welfare could be made arbitrarily large by making *ex ante* transfers of the private good from i to j. Thus, a solution to the maximization problem only exists when the welfare weights are, in expectation, the same for all agents. Thus, without loss of generality, we restrict the welfare weights to satisfy

$$\int_{\underline{v}^i}^{\bar{v}^i} \lambda_i(s) dF_i(s) = 1 \qquad \forall i.$$

We can write (*) as

$$(**) \qquad \max_{\psi,\delta} \min \int_V \left[\sum_i \left(\lambda_i(v^i) - \frac{\psi_i'(v^i)}{f_i(v^i)} \right)(v^i q(v) - a^i(v)) - \frac{\psi_i(v^i)}{f_i(v^i)} q(v) \right] dF(v)$$

$$+ \int_V \delta(v) \left(\sum_i a^i(v) - Kq(v) \right) dv + \sum_i [\psi_i(\bar{v}^i)\hat{u}^i(\bar{v}^i) - \psi_i(\underline{v}^i)\hat{u}(\underline{v}^i)].$$

From the first order conditions with respect to $a^i(v)$, $\delta(v)$, and $\psi_i(v^i)$ we obtain, for $\underline{v}^i < v^i < \bar{v}^i$,

(1) $\qquad -(\lambda_i(v^i)f_i(v^i) - \psi_i'(v^i)) + \gamma(v)f_i(v^i) = 0,$

(2) $\qquad \sum_i a^i(v^i) - Kq(v) \begin{cases} = 0, & \text{if } \gamma(v) > 0, \\ \geq 0 & \text{if } \gamma(v) = 0, \end{cases}$

(3) $\qquad \dfrac{\partial}{\partial v^i} \int_{V^{-i}} a^i(v) \, dF(v|v^i) \equiv A_i'(v_i^i) = v^i Q_i'(v^i) \qquad \forall i, v^i$

where $\gamma(v) = \delta(v)/f(v)$.

From (1) it follows that γ is constant in v. Integration of (1) gives $\psi^i(v^i) = F_i(v^i)(\lambda_i^-(v^i) - \gamma) + C$ where $\lambda_i^-(v^i)$ is the expected value of λ_i conditional on i's valuation being less than or equal to v^i.

Part (b) of the theorem[9] follows from (2) and (3).

Finally, the continuity of ψ_i along with the first order conditions for a^i at $\underline{v}^i$ and $\bar{v}^i$ imply that $\psi_i(\underline{v}^i) = \psi_i(\bar{v}^i) = 0$. So $C = 0$ and $\gamma \equiv \int_{\underline{v}^i}^{\bar{v}^i} \lambda_i(v^i) dF_i(v^i) = 1$. Substituting all of this into (**) implies that we must find q^* to solve

$$\max_q \int \left[\sum_i \left(v^i - \frac{\psi_i(v^i)}{f_i(v^i)} \right) - K \right] q(v) \, dF(v) \qquad \text{subject to:}$$

$$0 \leq q(v) \leq 1 \qquad \forall v \in V,$$
$$Q_i'(v^i) \geq 0 \qquad \forall i, v^i \in V^i. \qquad\qquad\qquad Q.E.D.$$

REMARK: The technique above applies equally to the case of general production functions. Let $C(q)$ be the cost of producing a public good level equal to q. First observe that the incentive compatibility constraints do not depend on $C(q)$. As a consequence, substitution of the incentive constraints results in the program:

$$\max_q \int \left[\sum_i \left(v^i - \frac{\psi_i(v^i)}{f_i(v^i)} \right) q(v) - C(q) \right] dF(v) \qquad \text{subject to:}$$

$$Q_i'(v^i) \geq 0 \qquad \forall i, v^i \in V^i.$$

---

[9] The existence of such an α for any given q was first shown by d'Aspremont and Gérard-Varet (1979). One α that satisfies (b) is

$$\alpha^i(v) = \frac{K}{N} \left[q^*(v) - Q_i^*(v^i) + \frac{1}{N-1} \sum_{j \neq i} Q_j^*(V^j) \right] - \frac{1}{N-1} \sum_{j \neq i} \int_{\underline{v}^j}^{v^j} s_j \, dQ_j^*(s_j).$$

The only regularity assumptions needed to guarantee a solution are that C is nondecreasing in q and that

$$\sum_i \left[\left(v^i - \frac{\psi_i(v^i)}{f_i(v^i)} \right) q(v) - C(q) \right]$$

has a solution for every valuation profile, v. The corresponding taxes are then constructed in a manner similar to part (b) of the theorem. Therefore, this general approach can be applied to problems with increasing or decreasing returns to scale, U-shaped average cost functions, and even lumpy public goods[10] with fixed costs and "jumps" in the cost function.

4. INTERPRETING THE CHARACTERIZATION

4.1. Virtual Cost-Benefit Criterion

Call

$$w_i(v^i) = v_i - \frac{F_i(v_i)}{f_i(v_i)}(\lambda_i^-(v_i) - 1),$$

type v^i of agent i's virtual valuation (à la Myerson). Suppose[11] $w_i'(v^i) \geq 0 \ \forall i, v^i$. Then, since $Q_i^*(v^i) = \text{prob}(\sum_{j \neq i} w_j(v^j) \geq K - w_i(v^i))$ it will be true that $Q_i'(v^i) \geq 0$ is never binding. So for (λ, F) such that $w_i'(v^i) \geq 0 \ \forall i, v^i$, interim efficient $q^*(v)$ satisfy

$$q^*(v) = 1 \quad \text{if } \sum_i w^i(v^i) \geq K,$$

$$= 0 \quad \text{otherwise.}$$

This is a *virtual cost-benefit* criterion.[12] The virtual utility has a familiar interpretation (see, for example, Myerson (1981)). It equals the "true" public good valuation of the v^i-type inflated[13] by a factor that depends on the distribution of types and on the welfare weights. The benchmark case is the one where $\lambda_i(v^i) = 1$ for all i and v^i. In this case the first best optimal level of public good is 1 or 0 depending only on whether or not $\sum_i[v^i - (K/N)] \geq 0$. That is, produce if and only if the sum of the marginal rates of substitution exceeds the marginal production cost. This is the Lindahl-Samuelson solution, precisely the solution investigated in most previous papers on the optimal provision of public good. (See d'Aspremont and Gérard-Varet (1979).) This simplification arises because the allocation of the private good (i.e., the incidence of the taxes on different types) does not affect social welfare. For this reason, incentive compatibility does not reduce social welfare relative to the first best solution. However, it must be emphasized

---

[10] These are sometimes referred to as threshold or step-level public goods. The simplest kind is just a binary public good, which is mathematically equivalent to the standard model presented in Section 2: constant returns with a maximum capacity.

[11] This is the so-called "regular" case, where the second order condition is never binding.

[12] Notice that similar (ex post) virtual cost-benefit conditions characterize the second-best optimum in the case of nonlinear production technologies, provided the second order condition is satisfied.

[13] This could be deflated if $\lambda_i^-(v^i) > 1$.

that this is a very special case. It is in fact the only system of welfare weights where incentive compatibility does not cause distortions relative to the first best solution.[14]

To better understand the intuition behind the virtual valuations, one can think of the mechanism operating in the following way. Each agent (truthfully) reports a valuation. If the public good is produced, then each agent pays the incentive tax, which equals a constant plus that agent's valuation minus his "informational rent," $(1 - F_i(v^i))/f_i(v^i)$. Recall from standard incentive theory that this is the amount that can be extracted from an agent, given incentive constraints. Of course, in this public good problem, the objective of the mechanism is not to extract rent from agents, so any excess incentive tax will be distributed lump sum back to the agents, by adjusting the incentive tax by a constant. Thus, if the good is provided, the government spends K to produce the public good and makes a lump-sum refund, which is formally captured by the constant (i.e. independent of v^i) that is added to each agent's incentive tax. The portion of this refund that comes from type v^i of agent i equals

$$\left(v^i - \frac{1 - F_i(v^i)}{f_i(v^i)} - \frac{K}{N} \right).$$

There are two other terms that complete the social cost/benefit picture, as it concerns type v^i of agent i. One is simply that producing the public good, produces a direct benefit of v^i to agent i, which is valued socially as $\lambda_i(v^i)v^i$. Last, but not least, is the fact that the incentive tax (before refund) is a social cost, and this social cost equals

$$\lambda_i(v^i)v^i - \frac{\int_{v^i}^{\overline{v}^i} \lambda_i(t^i) dF_i(t^i)}{f_i(v^i)}.$$

Collecting all these terms, gives us type v^i of agent i's contribution to the marginal net social value of producing the public good. Denoting this by $\overline{w}^i(v^i)$, gives us

$$\overline{w}^i(v^i) = \lambda_i(v^i)v^i - \left[\lambda_i(v^i)v^i - \frac{\int_{v^i}^{\overline{v}^i} \lambda_i(t^i) dF_i(t^i)}{f_i(v^i)} \right] + \left[v^i - \frac{1 - F_i(v^i)}{f_i(v^i)} - \frac{K}{N} \right]$$

$$= v^i - \frac{1 - F_i(v^i)}{f_i(v^i)} + \frac{\int_{v^i}^{\overline{v}^i} \lambda_i(t^i) dF_i(t^i)}{f_i(v^i)} - \frac{K}{N}$$

$$= w^i(v^i) - \frac{K}{N},$$

which is the *cost adjusted* virtual valuation of type v^i of agent i.

Notice that in the special case of neutral distributional weights, $\int_{v^i}^{\overline{v}^i} \lambda_i(t^i) dF_i(t^i) = 1 - F_i(v^i)$, so that

$$\lambda_i(v^i)v^i - \frac{\int_{v^i}^{\overline{v}^i} \lambda_i(t^i) dF_i(t^i)}{f_i(v^i)} = v^i - \frac{1 - F_i(v^i)}{f_i(v^i)}.$$

---

[14]Actually, this is the only system of welfare weights for which a first best solution exists. For any other weights, welfare can be arbitrarily increased by shifting the allocation of the private good to one particular type of some individual. Since we impose no feasibility bounds on the allocation of the private good, this means that the first best solution does not exist. Of course, with incentive compatibility constraints, the second-best problem is well defined.

and as a result there are no welfare costs associated with charging the incentive taxes in a type-dependent way and then redistributing them back in a lump sum fashion. Otherwise there is a cost to doing this.

The form of virtual utilities also makes it easy to see how distortions away from the classic optimum are related to the welfare weights. For example, if λ_i is decreasing in type then generally the interim efficient solution calls for *underproduction* relative to the Lindahl-Samuelson solution, since $\psi_i(v^i)$ is positive for all types. That is, the virtual valuations are always less than true valuations, so the sum of the true valuations must more than exceed the production cost in order for production to be optimal. Conversely, if λ_i is increasing in type, then there should be *overproduction* relative to the Lindahl-Samuelson solution.

4.2. *Second Order Conditions*

The discussion above assumes monotone virtual utilities, which ensures that maximization of the relaxed program, without the $Q^i(v^i) \geq 0$ constraint, automatically satisfies that constraint. It is straightforward to see what is required for virtual utilities to be monotone in type, and this provides a nice intuition for how our results differ from standard incentive problems of this type (e.g. Guesnerie and Laffont (1984)). From above,

$$w_i = v^i - \frac{1 - F_i(v^i)}{f_i(v^i)} + \frac{\int_{v^i}^{\overline{v}^i} \lambda_i(t^i) dF_i(t^i)}{f_i(v^i)}.$$

The first term, v^i, is clearly increasing in v^i. The second term, $(1 - F)/f$, the informational rent, is typically assumed to be monotone in v^i in adverse selection models in private goods environments, by requiring the distribution to satisfy a monotone hazard rate condition.

Since the incidence of incentive taxes can have welfare effects, there is a third term to worry about, indicating that one may need more (or sometimes less!) than the standard monotone hazard rate condition to guarantee that $Q_i'(v^i) \geq 0$ is automatically satisfied when one simply plugs in virtual utilities and maximizes subject only to production feasibility. These additional conditions will imply restrictions on the distribution of welfare weights, as we illustrate in the example below.

4.3. *Example*

Let v be distributed uniformly on $[0,1]$ for all i, so $F(v) = v$ and $f(v) = 1$. Then $w(v) = 2v - \int_0^v \lambda(t) dt$ and $w' = 2 - \lambda(v)$. Therefore, the second order condition is globally satisfied for uniform distributions of valuations if the maximum welfare weight is less than or equal to 2. Thus, if $\lambda(v) = 2(a + bv)/(2a + b)$, where $a \geq 0$ and $2a + b > 0$, then we are always in the "regular" case where virtual valuations are monotonic in type and the second order conditions are satisfied. If $b > 0$ (high valuation types receive more weight) then production will occur more often than in the Lindahl-Samuelson solution, while if $b < 0$, the reverse is true. However, there are λ such that virtual valuations are decreasing, even for the uniform distribution. For example, if $\lambda(v) = 3v^2$, then virtual valuations are decreasing for $v > \sqrt{2/3}$. The optimal solution in this case will involve

equal treatment of all types with valuations above some $v^* < 1$; $Q^i = 1$ for all such types.[15]

If $K \leq 1$, the solution above is straightforward, because the Q_i obtained from the relaxed problem is in fact weakly increasing, even though the virtual valuations are decreasing in some region.[16] The reason why Q_i obtained from the relaxed problem is nondecreasing is that virtual valuations are decreasing only for very high values of v, where $w(v) > 1$. So, when $K \leq 1$, the relaxed solution sets $Q_i = 1$ whenever $w(v) > 1$ for at least one agent, and all these high types are treated the same. But in general, for higher values of K, the relaxed solution may produce violations of the second order condition—i.e. Q_i decreasing in some region. In such cases, one applies a procedure called ironing (Rochet and Choné (1998)). The principle behind this procedure is to flatten out Q_i in the decreasing region (and for some adjacent types as well). The geometry is illustrated clearly in a series of figures in Guesnerie and Laffont (1984) for the single agent case.

It is also instructive to use this example to illustrate the range of public good provision rules (or cost-benefit criteria) that are interim efficient. Suppose $N = 2$, $K = 1$. The Lindahl-Samuelson efficient outcome is to produce if and only if the average valuation exceeds $1/2$, so the public good will be provided half the time.

Next suppose one shifts welfare weight to the low valuation types, to the point where $\lambda(v) = 2$ for all $v < 1/2$ and $\lambda(v) = 0$ for all $v > 1/2$. This satisfies monotonicity of virtual valuations[17] and it is easy to see that the optimal mechanism is to produce if and only if the sum of valuations exceeds $3/2$. In other words, this weighting scheme effectively inflates the cost of the public good by 50 percent, so it should be produced only if the actual benefit/cost ratio exceeds 1.5. At first blush it seems as though this mechanism could be improved, since there are some states where both agents are "high" types (i.e. $v > 1/2$ for both of them), and the public good is not provided. Since all high types receive the same welfare weight, and since low types do not bear any of the cost of production in these states, it would seem to lead to an improvement in welfare. Why doesn't this lead to an improvement? The answer is that the mechanism is designed to achieve redistributive goals in addition to deciding on public good production. In this case, the welfare weights indicate that there should be a transfer from high valuation to low valuation types. Hence in the optimal mechanism there are some states where there is one low type and one high type, and the public good is not produced, but a private good transfer takes place between the low and high types. The extent of such transfers would be hindered by greater public good production due to incentive compatibility problems. We conjecture that this choice of welfare weights corresponds to the lowest possible expected output ($Q = .125$) of all interim efficient mechanisms for the uniform case with $N = 2$ and $K = 1$.

---

[15] Similarly, if $\lambda(v) = 3(1 - v)^2$ the optimal solution will involve equal treatment of all types with valuations below some $v^* < 0$; $Q_i = 0$ for all such types.

[16] In other words, this is an example demonstrating why monotonicity of w is not a necessary condition for the second order conditions to be satisfied.

[17] If one shifts the welfare weights even further downward, so that $\lambda(v) = A > 2$ for all $v < 1/A$ and $\lambda(v) = 0$ for all $v > 1/A$, then the virtual valuations are nonmonotonic and ironing must be done. Nevertheless, from the characterization in Theorem 1, it is easy to verify that standard ironing procedures can be used and will generate an optimal mechanism with the same property: produce if and only it the sum of valuations exceeds $3/2$.

At the other extreme, suppose the welfare weights are shifted in the opposite direction, with $\lambda(v) = 2$ for all $v > 1/2$ and $\lambda(v) = 0$ for all $v < 1/2$. In this case the optimal mechanism is to produce if and only if the sum of valuations exceeds $1/2$. In other words, the cost of the public good is effectively deflated by 50 percent, so that it should be produced if the actual benefit/cost ratio is at least .5. Again it would seem that efficiency would dictate that when both types are "low" types, the good should never be produced. However, redistributive goals implied by this welfare weighting scheme require the low types to subsidize the cost of the public good. The most efficient way to perform this subsidization requires some "overproduction" of the public good. We conjecture that this choice of welfare weights corresponds to the highest possible expected output ($Q = .875$) of all interim efficient mechanisms for the uniform case with $N = 2$ and $K = 1$.

5. CONCLUSIONS

In this paper, we have characterized the interim efficient public good allocation rules in a simple Bayesian public good environment. We find that the optimal mechanism involves either more or less production of the public good depending on whether the welfare weights are shifted in the direction of types with higher or lower valuations for the public good. Thus, compared to the classical optimal level of public good provision (the "Lindahl-Samuelson" solution), there should generally be some distortion. The reason for this distortion is that unless welfare weights are perfectly neutral, efficient allocations will depend in general on both the level of public good and the incidence of taxes to finance the public good. Because of incentive compatibility, the efficient way to reduce the tax burden on low-valuation (resp: high-valuation) consumers is to reduce (resp: increase) the level of provision of the public good. In the borderline case, the first-best solution is attainable only because the welfare function is independent of distribution of the private good.

There are several directions worth pursuing. One direction is to explore the use of simple mechanisms. The public good mechanisms proposed here involve complicated transfer schemes that can necessitate the use of very large taxes and subsidies. In a companion paper (Ledyard and Palfrey (1998)) we explore simple mechanisms in large populations, and show that for *any* interim efficient allocation rule there exists a simple dominant-strategy referendum mechanism that perfectly approximates the efficiency of that allocation rule. In a referendum, individuals simply submit a binary message (a "vote") either for or against production of the public good. If a sufficiently large fraction of the individuals vote in favor, then the public good is provided and the costs are distributed equally in the population. Otherwise, the public good is not produced. This provides an approximate "welfare theorem" for public goods: efficient allocation rules can be (approximately) decentralized by an appropriately chosen voting rule. Moreover, if there is a common value component to the distribution of preferences, then the optimal referendum is *unique*. We prove this by approximating the solution to the optimal mechanism where the second order condition is ignored. Thus a by-product is the result that the second order conditions are inconsequential in large economies.

There are several other directions. Participation constraints were not imposed in our solution for the optimum. It is fairly easy to show that when these constraints are binding, this implies a reduction in the level of the public good, since these constraints are necessarily binding on the low valuation types (Ledyard and Palfrey (1994)). It is also

true that, except in uninteresting cases, these constraints will imply $Q_N \to 0$ in large populations (Ledyard and Palfrey (1994), Mailath and Postlewaite (1990)). But for the case of large N, it would usually seem more realistic to assume that participation is generally obligatory to all members of the group under consideration, as we have assumed here. Related to the general issue of participation is the application of the general approach presented here to excludable public goods. In that case, participation constraints can be relaxed by the (no-cost) exclusion of low valuation types.

More involved extensions, such as relaxing the assumption of independent types, consideration of utility functions where the valuation parameter enters nonlinearly, or introducing multidimensional types, appear to be more difficult open questions. Finally, as we remarked at the end of Section 3, the analysis is easily extended to accommodate arbitrary production technologies.

Div. of Humanities and Social Sciences, California Institute of Technology, HSS 228-77, 1200 E. California Blvd., Pasadena, CA 91125, U.S.A.; jledyard@hss.caltech.edu
and
Div. of Humanities and Political Science, California Institute of Technology, HSS 228-77, 1200 E. California Blvd., Pasadena, CA 91125, U.S.A.; trp@hss.caltech.edu

REFERENCES

D'ASPREMONT, C., AND L.-A. GÉRARD-VARET (1979): "Incentives and Incomplete Information," *Journal of Public Economics*, 11, 25–45.

CRAMTON, P., AND T. PALFREY (1995): "Ratifiable Mechanisms: Learning from Disagreement," *Games and Economic Behavior*, 10, 255–283.

CRAWFORD, V. (1985): "Efficient and Durable Decision Rules: A Reformulation," *Econometrica*, 53, 817–836.

GUESNERIE, R., AND J.-J. LAFFONT (1984): "A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-Managed Firm," *Journal of Public Economics*, 25, 329–369.

HOLMSTRÖM, B., AND R. MYERSON (1983): "Efficient and Durable Decision Rules with Incomplete Information," *Econometrica*, 51, 1799–1819.

JACKSON, M., AND H. MOULIN (1992): "Implementing a Public Project and Distributing Its Cost," *Journal of Economic Theory*, 57, 125–140.

LEDYARD, JOHN O., AND THOMAS PALFREY (1994): "Voting and Lottery Drafts as Efficient Public Goods Mechanisms," *Review of Economic Studies*, 61, 327–355.

——— (1998): "Voting is Approximately Optimal," Working Paper, California Institute of Technology.

——— (1999): "Interim Efficiency in a Public Goods Problem," in *Social Organization and Mechanism Design*, Proceedings of the 1996 Francqui Prize Colloquium, ed. by Claude d'Aspremont. Belgium: De Boeck University, 229–251.

MAILATH, GEORGE, AND ANDREW POSTLEWAITE (1990): "Asymmetric Information Bargaining Problems with Many Agents," *Review of Economic Studies*, 57, 351–367.

MIRRLEES, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38, 175–208.

MYERSON, R. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58–73.

ROCHET, JEAN-CHARLES (1987): "A Necessary and Sufficient Condition for Rationalizability in a Quasi-Linear Context," *Journal of Mathematical Economics*, 16, 191–200.

ROCHET, JEAN-CHARLES, AND PHILIPPE CHONÉ (1998): "Ironing, Sweeping and Multidimensional Screening," ·Econometrica, 66, 783–826.

WILSON, ROBERT (1993): "Design of Efficient Trading Procedures," in *The Double Auction Market Institutions, Theories, and Evidence*, ed. by D. Friedman and J. Rust. Proceedings of the Workshop on Double Auction Markets Held June, 1991 in Santa Fe, New Mexico. Proceedings Volume XIV. Reading, MA: Addison-Wesley.