A Note on the Characterization of Mechanisms for the Revelation of Preferences

Mark Walker

Econometrica, Volume 46, Issue 1 (Jan., 1978), 147-152.

Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28197801%2946%3A1%3C147%3AANOTCO%3E2.0.CO%3B2-W

Econometrica, Vol. 46, No. 1 (January, 1978)

A NOTE ON THE CHARACTERIZATION OF MECHANISMS FOR THE REVELATION OF PREFERENCES

BY MARK WALKER

Green and Laffont have characterized certain appealing dominant-strategy revelation mechanisms as precisely the mechanisms introduced by Groves, but they have established the characterization only for unstructured sets of public alternatives: if the set has some natural structure, their proof generally requires that pathological preferences be admissible. It is shown here that the same characterization holds on sets in $\boldsymbol{R}^n$, even when only nice preferences are admitted; this greatly extends the usefulness of the characterization.

T. GROVES [**4**] AND OTHERS have recently discovered a method whereby one can often obtain truthful revelation of individuals' preferences for public goods. Groves and Loeb [**5**] have shown, in particular, that in the case of transferable utility (or equivalently, one private good and no income effects), mechanisms of the Groves form *always* elicit truthful preference-revelation. Green and Laffont [**2**] have shown, moreover, that, in the transferable-utility case, Groves mechanisms are the *only* ones which elicit truthful revelation.

One must exercise some caution, however, in applying the characterization theorem of Green and Laffont: the proof provided in [**2**] fails to cover some of the most interesting possibilities for application. Specifically, the proof in [**2**] requires that we admit as possible many preferences which are often considered to be quite pathological. If individuals can be expected to have only relatively "nice" preferences (or if one wishes to restrict one's consideration to nice preferences in order to simplify the analysis of a problem), then one cannot (at present) make use of the Green-Laffont theorem.[1] In this note, the theorem is shown to hold over virtually any set of convex preferences; this result allows us, if we wish, to concentrate our attention upon nice preferences.

Suppose that we have a set Y of alternatives from which the "public" choice must be made, and a set $N=\{1, \ldots, n\}$ of n individuals, each with a utility function u_i: $\boldsymbol{R}^n \times Y \rightarrow \boldsymbol{R}$ of the form

$$u_i(\boldsymbol{x}, y)=x_i+v_i(y).$$

In other words, an outcome is an $(n+1)$-tuple $(\boldsymbol{x}, y)=(x_1, \ldots, x_n; y)$, and the utility that individual i associates with an outcome $(\boldsymbol{x}, y)$ is given by the amount x_i of "money" that he receives (a real number), plus a term $v_i(y)$ which depends only upon the choice of the public component y. Each individual's preference is thus completely described by a real-valued function v_i, which we call his *valuation*.

[1] For applications of the Green-Laffont theorem, see [**7, 1**] and (implicity) [**3**, Theorem 9, page 39]. All three use the theorem to derive impossibility results; it is first shown that there is no Groves mechanism with a given property, from which it follows (via the characterization theorem) that there can be no "incentive-compatible" mechanism with that property. One is perfectly happy, of course, to demonstrate such impossibility results in nice cases only—indeed, it is to some extent preferable, since then the result will not depend upon the inclusion of any pathologies. Of the three applications just mentioned, the second and third refer only to finite sets of alternatives, to which the proof in [**2**] applies; the first refers to a continuum (which includes the standard "public goods" model), and is accomplished with nice preferences; hence it requires the result established in this note.

Each individual knows his own valuation, of course, but the public choice y must be made (by some public agency, say) without a priori knowledge of the individuals' valuations. We could ask each individual to indicate his preference by reporting his valuation, and this would enable the public agency to identify those values of y which maximize the private benefits. The difficulty with this approach, however, is the "free-rider problem:" there is no obvious way to be sure that the individuals are reporting their *true* valuations, and hence the public agency cannot be sure that its choice of y is indeed optimal.[2]

Groves and Loeb have demonstrated that if we always choose the public component y so as to maximize the sum of the reported valuations, then there is a way to transfer money (utility) directly between the individuals so that each individual will always be led to report the preference that he actually holds, instead of some other (false) preference. More precisely, we define a *provision function* d, which takes its values in Y, and n real-valued *transfer functions* $t_1, \ldots, t_n$; the arguments of each of these $n+1$ functions are n-tuples $\boldsymbol{v} = (v_1, \ldots, v_n)$ of valuation functions. These transfer and provision functions define the outcome function of an n-person game in which the n private individuals will participate: $(\boldsymbol{x}, y) = (t_1(\boldsymbol{v}), \ldots, t_n(\boldsymbol{v}), d(\boldsymbol{v}))$. The strategies available to an individual are valuation functions. Groves and Loeb have shown that if d and $t_1, \ldots, t_n$ satisfy both

(1) for each n-tuple $\boldsymbol{v}$, $d(\boldsymbol{v})$ yields a maximum of the function

$$\sum_{i=1}^{n} v_i \colon Y \to \boldsymbol{R}, \qquad \text{and}$$

(2)[3]

$$t_i(\boldsymbol{v}) - \sum_{j \neq i} v_j(d(\boldsymbol{v})) = t_i(\boldsymbol{v}_{\sim i}, \tilde{v}) - \sum_{j \neq i} v_j(d(\boldsymbol{v}_{\sim i}, \tilde{v})),$$

for each $i \in N$ and each $\boldsymbol{v}$ and $\tilde{v}$,[4]

then they also satisfy

(3)
$$t_i(\boldsymbol{v}) + v_i(d(\boldsymbol{v})) \geqslant t_i(\boldsymbol{v}_{\sim i}, \tilde{v}) + v_i(d(\boldsymbol{v}_{\sim i}, \tilde{v})),$$

for each $i \in N$ and each $\boldsymbol{v}$ and $\tilde{v}$.

[2] It should be emphasized here that "optimal" does not mean "Pareto optimal," but merely that y yields a maximum of the function $\Sigma_{i=1}^{n} \mathring{v}_i \colon Y \to \boldsymbol{R}$, where $\mathring{v}_i$ is the true valuation of individual i. This condition, together with the condition that $\Sigma_{i=1}^{n} x_i = 0$, is both necessary and sufficient for Pareto optimality of a feasible allocation in the transferable-utility context being considered here; however it is shown in **[7]**, using the result in this note, that there is no way to elicit true responses and to always attain Pareto optimality, as well.

[3] For any n-tuple $\boldsymbol{v} = (v_1, \ldots, v_n)$ of functions, let $\boldsymbol{v}_{\sim i}$ denote the $(n-1)$-tuple formed by omitting v_i, and for any function v, let $(\boldsymbol{v}_{\sim i}, \tilde{v})$ denote the n-tuple formed by inserting $\tilde{v}$ into the ith component of $\boldsymbol{v}_{\sim i}$: $(\boldsymbol{v}_{\sim i}, \tilde{v}) = (v_1, \ldots, \tilde{v}_i, \ldots, v_n)$.

[4] Equivalently, for each $i \in N$, there is a function h_i on $(n-1)$-tuples $\boldsymbol{v}_{\sim i}$ such that

$$t_i(\mathbf{v}) = \sum_{j \neq i} v_j(d(\boldsymbol{v})) + h_i(\boldsymbol{v}_{\sim i}).$$

In other words, if the public agency chooses the provision and transfer functions so as to satisfy (1) and (2), then, in the resulting game, "telling the truth" will always be a dominant strategy for every individual. (Notice that when (3) is satisfied, (1) is both necessary and sufficient in order that the public choice y always be optimal.)

Let V be a collection of valuations $v: Y \to \boldsymbol{R}$, each of which attains a maximum, and let $d: V^n \to Y$ satisfy (1). Green and Laffont [2] have shown that if V is broad enough, then the only transfer functions $t_1, \ldots, t_n$ which everywhere satisfy (3) are those of the form (2). They have demonstrated this result, in particular, when V includes all functions $v: Y \to \boldsymbol{R}$ which attain a maximum, and also in the case in which Y is a subset of a Euclidean space and V includes all *continuous* functions which attain a maximum.

The proofs in [2] involve the construction of a valuation function v_i which, for a given $\boldsymbol{v}_{\sim i}$, approximates the function $-\Sigma_{j \neq i} v_j$ at two points of Y, and which is everywhere less than $\delta - \Sigma_{j \neq i} v_j(y)$, where δ is a given positive number. If Y is a linear space and the components of $\boldsymbol{v}_{\sim i}$ are concave, then such a construction will generally exhibit a considerable degree of non-concavity, as in Figure 1.

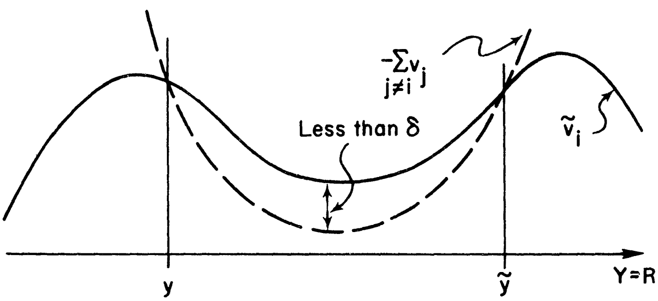

FIGURE 1

The proof to be given below, on the other hand, is valid for any collection V of concave, maximum-attaining functions (on a convex set $Y \subseteq \boldsymbol{R}^m$) which is rich enough to satisfy the following relatively mild condition (it is easy to show, for example, that the class of all analytic concave real functions (and therefore any larger class of concave real functions) satisfies this condition).[5]

CONDITION A: The set V satisfies Condition A if, for any linear function $f: \boldsymbol{R}^m \to \boldsymbol{R}$, any two members y and $\tilde{y}$ of Y, and any $\varepsilon > 0$, there is a $v \in V$ which satisfies both

(A1) $\nabla v(\tilde{y}) = \nabla f(\tilde{y})$, and

(A2) $f(y) - v(y) < f(\tilde{y}) - v(\tilde{y}) + \varepsilon.$

[5] Condition (A1) can in fact be replaced by the weaker

(A1*) $$\lim_{\lambda \downarrow 0} \frac{v(\tilde{y} + \lambda(y - \tilde{y})) - v(\tilde{y})}{\lambda |y - \tilde{y}|} = f(\tilde{y}) \cdot (y - \tilde{y}).$$

To simplify the analysis, we will consider only *strictly* concave valuation functions, so that π will be a uniquely determined, single-valued function. In order to apply the following theorem to the more general case (of functions that are concave, but not strictly so), one can follow [**2**, Theorem 6].

THEOREM: *Let Y be a convex subset of $\boldsymbol{R}^m$, let V be a set of strictly concave, maximum-attaining functions on Y, and let V satisfy Condition A. If the functions $d: V^n \to \boldsymbol{R}$ and $t_i: V^n \to \boldsymbol{R}$ $(i = 1, \ldots, n)$ satisfy* (1) *and* (3), *then they also satisfy* (2)—*i.e.*, $(d, \boldsymbol{t})$ *is a Groves mechanism.*

PROOF: Let $i \in N$ and $\boldsymbol{v}_{\sim i} \in V^{n-1}$, and let both be fixed throughout the proof; for each $y \in Y$, let $\psi(y) = \Sigma_{j \neq i} v_j(y)$. Functions denoted by v_i, $\tilde{v}_i$, and $\bar{v}$ will appear in the proof; in each case, the n-tuple with the same notation appended to it will be the one formed by inserting v_i, $\tilde{v}_i$, or $\bar{v}$ into $\boldsymbol{v}_{\sim i}$: $\boldsymbol{v} = (\boldsymbol{v}_{\sim i}, v_i)$, $\tilde{\boldsymbol{v}} = (\boldsymbol{v}_{\sim i}, \tilde{v}_i)$, and $\bar{\boldsymbol{v}} = (\boldsymbol{v}_{\sim i}, \bar{v})$.

For each $v_i \in V$, let

(4) $$h(v_i) = t_i(\boldsymbol{v}) - \psi(d(\boldsymbol{v}));$$

we must show that h is a constant function. Green and Laffont have shown (in their proof of Theorem 3 of [**2**]) that h depends only upon the provision $y = d(\boldsymbol{v})$—i.e., if $d(\boldsymbol{v}) = d(\tilde{\boldsymbol{v}})$, then $h(v_i) = h(v_i)$. Hence, for each $y \in Y$, we can write without loss of generality (we henceforth take Y to be $d(V^n)$):

(5) $$g(y) = h(v_i), \quad \text{where } v_i \in V \text{ is such that } d(\boldsymbol{v}) = y.$$

We will show that g is constant on Y.

Let y and $\tilde{y}$ be any two distinct members of Y, and let $z = y - \tilde{y}$. We will show that the directional derivative $g'_z(\tilde{y})$ defined by

$$g'_z(\tilde{y}) = \lim_{\lambda \downarrow 0} \frac{g(\tilde{y} + \lambda z) - g(\tilde{y})}{\lambda |z|}$$

exists and is zero. Because y and $\tilde{y}$ were chosen arbitrarily, and because Y is convex, it follows (via the mean-value theorem for functions of a single variable) that g is constant on Y.

Let v_i and $\tilde{v}_i$ be members of V for which, respectively, $d(\boldsymbol{v}) = y$ and $d(\tilde{\boldsymbol{v}}) = \tilde{y}$. Combining (3) and (4), we have

$$h(v_i) + \psi(d(\boldsymbol{v})) + v_i(d(\boldsymbol{v})) \geqslant h(\tilde{v}_i) + \psi(d(\tilde{\boldsymbol{v}})) + v_i(d(\tilde{\boldsymbol{v}})).$$

Combining this inequality with (5) yields both

(6) $$g(y) - g(\tilde{y}) \geqslant \psi(\tilde{y}) - \psi(y) + v_i(\tilde{y}) - v_i(y)$$

and

(7) $$g(y) - g(\tilde{y}) \leqslant \psi(\tilde{y}) - \psi(y) + \tilde{v}_i(\tilde{y}) - \tilde{v}_i(y).$$

Let $\psi_z'(\tilde{y})$ denote the directional derivative of ψ at $\tilde{y}$ in the direction z:

$$\psi_z'(\tilde{y}) = \lim_{\lambda \downarrow 0} \frac{\psi(\tilde{y}+\lambda z)-\psi(\tilde{y})}{\lambda |z|};$$

the existence of this limit is guaranteed by the concavity of ψ (see, e.g., [**6**, Theorem 23.1]). Let $s \in \boldsymbol{R}^m$ be a subgradient of $-\psi$ at $\tilde{y}$, and let s coincide with $-\psi_z'(\tilde{y})$ in the direction z; i.e., $s \cdot z/|z| = -\psi_z'(\tilde{y})$. Let $f: \boldsymbol{R}^m \to \boldsymbol{R}$ be the linear function defined by

$$f(y') = \tilde{v}_i(\tilde{y}) + s \cdot (y' - \tilde{y}).$$

For every $\varepsilon > 0$, according to Condition A, there is a $\bar{v} \in V$ such that

$$\text{(8)} \qquad \nabla \bar{v}(\tilde{y}) = s,$$

and

$$\text{(9)} \qquad f(y) - \bar{v}(y) < f(\tilde{y}) - \bar{v}(\tilde{y}) + \varepsilon.$$

It follows from (8) that $d(\bar{\boldsymbol{v}}) = \tilde{y}$, and from (9) that

$$\bar{v}(\tilde{y}) - \bar{v}(y) < -s \cdot (y - \tilde{y}) + \varepsilon,$$

and hence that

$$\text{(10)} \qquad \bar{v}(\tilde{y}) - \bar{v}(y) < |z| \psi_z'(\tilde{y}) + \varepsilon.$$

Since (7) holds for every $\tilde{v}_i$ for which $d(\tilde{\boldsymbol{v}}) = \tilde{y}$, it must hold, in particular, for each of the $\bar{v}$; hence, for each $\varepsilon > 0$, we can substitute (10) into (7), and we have

$$g(y) - g(\tilde{y}) \leqslant \psi(\tilde{y}) - \psi(y) + |z| \psi_z'(\tilde{y}) + \varepsilon.$$

We therefore have

$$\text{(11)} \qquad g(y) - g(\tilde{y}) \leqslant \psi(\tilde{y}) - \psi(y) + |z| \psi_z'(\tilde{y});$$

application of the same argument to (6) yields

$$\text{(12)} \qquad g(y) - g(\tilde{y}) \geqslant \psi(\tilde{y}) - \psi(y) + |z| \psi'_{-z}(y),$$

where $\psi'_{-z}(y)$ denotes the derivative of ψ at y in the direction $-z$.

For each positive number λ, of course, $\psi_{\lambda z}'(\tilde{y}) = \psi_z'(\tilde{y})$; if $0 < \lambda \leqslant 1$, then the inequality

$$\text{(13)} \qquad \frac{g(\tilde{y}+\lambda z) - g(\tilde{y})}{\lambda |z|} \leqslant \psi_z'(\tilde{y}) - \frac{\psi(\tilde{y}+\lambda z) - \psi(\tilde{y})}{\lambda |z|}$$

follows from (11), because Y is convex and y is arbitrary in (11). It is clear from (13) that $g_z'(\tilde{y}) \leqslant 0$. Similarly, inequality (12) implies that

$$\frac{g(\tilde{y}+\lambda z) - g(\tilde{y})}{\lambda |z|} \geqslant \psi'_{-z}(y + \lambda z) - \frac{\psi(\tilde{y}+\lambda z) - \psi(\tilde{y})}{\lambda |z|}$$

for $0<\lambda\leqslant 1$; concavity of ψ guarantees that $\lim_{\lambda\downarrow 0}\psi'_{-z}(\tilde{y}+\lambda z)=\psi'_z(\tilde{y})$ (see, e.g., [6, Theorem 24.1]) and hence that $g'_z(\tilde{y})\geqslant 0$, and we have the desired conclusion, $g'_z(\tilde{y})=0$.

State University of New York at Stony Brook

Manuscript received August, 1976; final revision received February, 1977.

REFERENCES

[1] BENNETT, E., AND D. CONN: "The Group Incentive Properties of Mechanisms for the Provision of Public Goods," Northwestern University, mimeographed, 1976.

[2] GREEN, J., AND J.-J. LAFFONT: "Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods," *Econometrica*, 45 (1977), 427–438.

[3] ———: "Revelation Des Preferences Pour Le Biens Publics," presented to the Seminaire D'Econometrie of C.N.R.S., Paris, January, 1976.

[4] GROVES, T.: "Incentives in Teams," *Econometrica*, 41 (1973), 617–631.

[5] GROVES, T., AND M. LOEB: "Incentives and Public Inputs," *Journal of Public Economics*, 4 (1975), 211–226.

[6] ROCKAFELLAR, R. T.: *Convex Analysis*. Princeton, New Jersey: The Princeton University Press, 1970.

[7] WALKER, M.: "On the Impossibility of a Dominant-Strategy Mechanism to Optimally Decide Public Questions," D. P. #160, Economic Research Bureau, S.U.N.Y., Stony Brook, New York, 1976.