No.385

Incentive Compatible Mechanisms are not Individually Rational, I: The Public Good Economy Case

by

Tatsuyoshi Saijo*

November 1988

* The author wishes to thank Sen Eguchi and M. Miyakawa for their comments.

Abstract

Hurwicz's theorem states that Pareto optimality, Individual rationality and incentive compatibility are incompatible in a class of private good economies. A similar result is obtained by Ledyard-Roberts in a class of public good economies. This paper presents a new impossibility result in a class of public good economies. Natural individual-rationality, which strengthens individual rationality in a natural way, and incentive compatibility are incompatible. That is, the result can be interpreted that every participant's desire to be better off, i.e., natural individual-rationality, rather than Pareto optimality is a source of the impossibility.

Incentive Compatible Mechanisms are not Individually Rational, I:

The Public Good Economy Case

by

Tatsuyoshi Saijo

October 1988

1. Introduction

Hurwicz's theorem [1972] states that there does not exist any incentive compatible (IC) mechanism that attains Pareto optimality (PO) and individual rationality (IR) in pure exchange environments with private goods.[1] That is, the intersection of these three requirements is empty. A mechanism is incentive compatible if and only if truthful revelation of preferences is a dominant strategy for each participant.[2] The individual rationality condition requires that any allocation attained by a mechanism must be no worse than the initial endowment for each participant. Ledyard and Roberts [1974] reconfirmed the same result as Hurwicz's in public good environments.

I propose a similar impossibility result to Ledyard and Roberts' without requiring Pareto optimality. Instead, I slightly strengthen the IR condition, which I call the natural individual rationality (NIR) condition. Every participant's bundle in the final allocation must be no worse than a best bundle that can achieved solely by his endowment and technology without using any other participants' endowments. For example, every participant can install a streetlight by himself without using other participants' resources and can be happier than before. The new IR condition says that the final allocation should be no worse than a situation with a streetlight installed by himself. This modification of the IR condition is more natural than the previous one if rationality of participants is assumed. My claim is that the intersection of the IC and NIR requirements is empty.

This result is rather surprising. Among designers who construct Groves-type mechanisms (for example, see Walker [1980]), the PO requirement is an obstacle to find an IC mechanism presupposing a very weak condition, i.e., the IR condition. My finding here is that PO is not an obstacle to construct an IC mechanism, but rather that every participant's desire to be better off is a major source of impossibility to construct an IC mechanism.

2. IC Mechanisms are not NIR

Consider an economy with n participants, and two types of commodities x and y, where x is a k-dimensional vector of private goods and y is a ℓ-dimensional vector of public goods. Each participant i has a utility function $u_i(x_i,y)$. Let (w_{xi},w_y) be participant i's initial endowment vector. Since w_y is a public good endowment vector, there is no participant's name on it. Let $F^i(x,y)$ be knowledge of a technical transformation for participant i. An allocation $((x_i),y)$ is ***feasible*** iff there exists (x_j) such that $\Sigma_i\ x_i + \Sigma_j\ x^j = \Sigma_i\ w_{xi}$ and $F^j(x^j,y) = 0$ for all j.

Definition 1. An allocation $((x_i),y)$ is naturally individual rational (NIR) iff $u_i(x_i,w_y+y) \geq u_i(x_i',w_y+y_i)$ for all i, where (x_i',w_y+y_i) is a utility maximizer subject to $F^i(x^i,y_i) = 0$ and $x_i'' + x^i = w_{xi}$.

That is, if an allocation is NIR, every participant should be no worse than a commodity bundle that is a utility maximizer if he behaves like a Robinson-Crusoe. Note that in a pure exchange economy without production the NIR condition coincides with the IR condition.

Incentive compatibility is now formalized. A pair of (w_{xi},u_i) is called participant i's *characteristic* and denoted e_i and the list $e = (e_1,e_2,\cdots,e_n)$ is called an *environment*. Let E_i be the set of participant i's characteristics and let $E = \times E_i$. Assume that each participant's endowment is public knowledge, and hence e_i is regarded as a utility function for participant i. Let S_i be the set of all possible strategies for participant i, and set $S_i = E_i$ for all i. A *direct revelation mechanism*, simply called a mechanism later, is a function $g{:}E \rightarrow A$, where A is the set of all feasible allocations. Let $\hat{s}_i$ be participant i's true characteristic and let $g_i(s) = (x_i,y)$, where $g(s) = ((x_j),y)$.

Definition 2. A mechanism is incentive compatible iff

$\forall\hat{s}$, $\forall i$, $\forall s = (s_i,s_{-i})$, $u_i(g_i(\hat{s}_i,s_{-i})) \geq u_i(g_i(s_i,s_{-i}))$.

That is, incentive compatibility requires that truthful revelation of preferences is a dominant strategy for all participants.

Since Kolm's triangle (see Malinvaud [1971] and Thomson [1987]) will be used in Theorem 1's proof, a brief explanation of the triangle is given here. Let $n = 2$, $k = \ell = 1$, and let $F^j(x^j, y) = x^j - y = 0$ for $j = 1$ and 2. That is, it is assumed that both participants have the same knowledge on the production of public good. Therefore, the production function is written by $y = f(x) = x$. Assume that $w_y = 0$. Then the feasibility of an allocation is that $x_1 + x_2 + y = w_{x_1} + w_{x_2}$. Take any point in the triangle in Figure 2. The perpendicular distance from that point to line $O_1 - T$ measures the consumption of x_1, the perpendicular distance from that point to line O_2-T measures x_2, and the distance from the point to the bottom side measures y. Then, the sum of three distances is always constant and it measures the height of the triangle, which is also the total amount of endowments. The situation in Figure 1-a is translated to the triangle in the following manner: the 45 degree line A-F in Figure 1-a becomes A'-F' in Figure 1-b. Notice that participant *1* can choose any point on A-F or A'-F' without borrowing any help from participant *2*. The height of the triangle O_1-A'-F' is O_1-F and the triangle is equilateral. B becomes B' and w'_{x_1}-B' is parallel to O_1-F'. x_1^* in Figure 1-b corresponds to the distance O-x_1^* in Figure 1-a. Similarly, y^* in Figure 1-b corresponds to the distance O-y^* in Figure 1-a.

Theorem 1. Any incentive compatible mechanism is not naturally individual-rational.

<u>*Proof*</u>. Suppose by way of contradiction that there exists an incentive compatible mechanism g that is naturally individual-rational. Consider an

economy with two participants, one private good, one public good, $w_y = 0$, $w_{x_1} = w_{x_2} = 2$ and a public good production function $y = f(x) = x$. Assume that two participants have the following same utility function:

$$u_i(x_i,y) = \begin{cases} y + 3x_i & \text{if } \; y > \frac{1}{3}x_i + \frac{2}{3} \\ \frac{1}{3}(5 - \frac{8}{3y + x_i})(3y + x_i) & \text{if } \; y \le \frac{1}{3}x_i + \frac{2}{3} \end{cases}.$$

See Figure 2. w is the initial endowment point. Without any help from participant 2, participant 1 can choose any point on w–F. On this segment, participant 1 chooses A so as to maximize his utility. Similarly, participant 2 chooses B. Therefore, the set of naturally individual-rational allocations is the rectangle T-A-C-B including the boundary points. There are two cases: g chooses a point in T-A-C or chooses a point in T-C-B. Suppose that g chooses a point in T-A-C including the boundaries. Then participant 1 can announce the following utility function:

$$u_1'(x_1,y) = x_1.$$

Dotted lines in Figure 2 are the preference map for this utility function. Then g must choose B. Notice that B is strictly better than any point in T-A-C. A similar argument can be applied to T-C-B. □

w–F in Figure 2 can be regarded as a budget line when the Lindahl price for participant 1 is one supposing that the private good is a nemeraire. Notice that the budget line w-F rotates clockwise as the Lindahl price for participant 1 goes down. Therefore, if the Lindahl price, that is, the ratio of the burden to construct a public good is between zero and one, then the Lindahl allocation must be in T-A-C-B. That is, the Lindahl allocation is naturally individual-rational. Hence we have the following famous fact as a direct consequence of Theorem 1.

Corollary 1. The Lindahl mechanism is not incentive compatible.

Contribution is occasionally used to construct a public good. Each participant donates his endowment to build a public good so as to maximize his satisfaction given other participants' donation. The allocation by this contribution mechanism is not Pareto optimal, but is naturally individual rational (for further detail, see Saijo [1988]). Hence we have

Corollary 2. The contribution mechanism is not incentive compatible.

Footnotes

1. See also Postlewaite [1985].

2. It seems to be a strong requirement that truthful revelation of preferences is a dominant strategy for all participants, but it is actually equivalent to the condition that truthful revelation of preferences is a Nash strategy for all participants. See Dasgupta-Hammond -Maskin [1979].

References

Dasgupta, P., P. Hammond, and E. Maskin, 1979, The implementation of social choice rules: Some general results on incentive compatibility, *Review of Economic Studies* 46:185-216.

Hurwicz, L., 1972, On informationally decentralized system, in *Decision and organization: A volume in honor of Jacob Marschak*, eds. R. Radner and C. B. McGuire, Amsterdam: North-Holland, 297-336.

Ledyard, J., and J. Roberts, 1974, On the incentive problem with public goods, Discussion Paper 116, Center for Mathematical Studies in Economics and Management Science, Northwestern University.

Malinvaud, E., 1971, A planning approach to the public goods problem, *Swedish Journal of Economics* 73:96-112.

Postlewaite, A., 1985, Implementation via Nash equilibria in economic environments, in *Social goals and social organization*, eds., L. Hurwicz, S. Schmeidler and H. Sonnenschein, Cambridge University Press.

Saijo, T., 1988, Contribution mechanism for public good provision is not incentive compatible: a reexamination, Institute of Socio-Economic Planning Discussion Paper Series, No. 381, University of Tsukuba.

Thomson, W., 1987, Lecture on public goods, mimeo., University of Rochester.

Walker, M., 1980, On the nonexistence of a dominant strategy mechanism for making optimal public decisions, *Econometrica* 48:1521-1540.

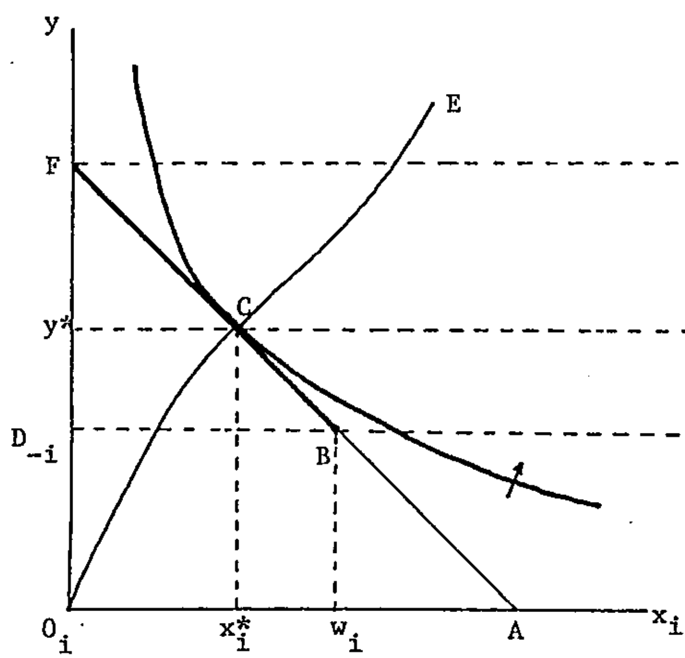

Figure 1-a

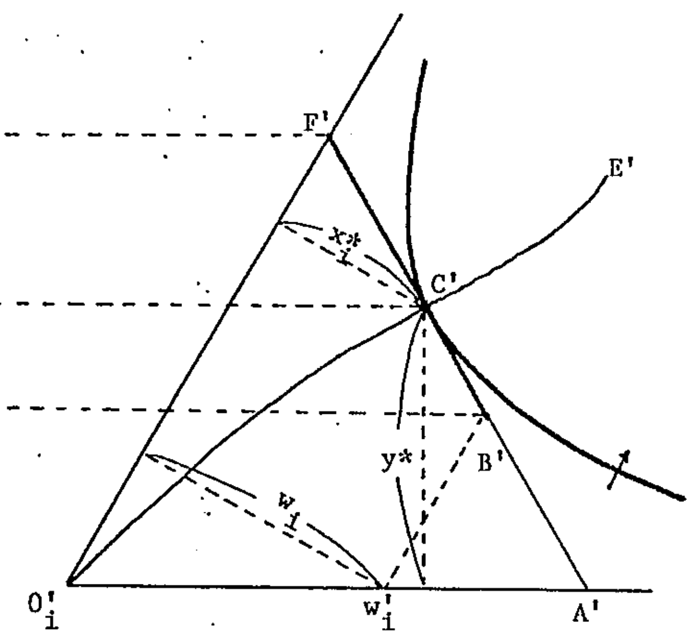

Figure 1-b

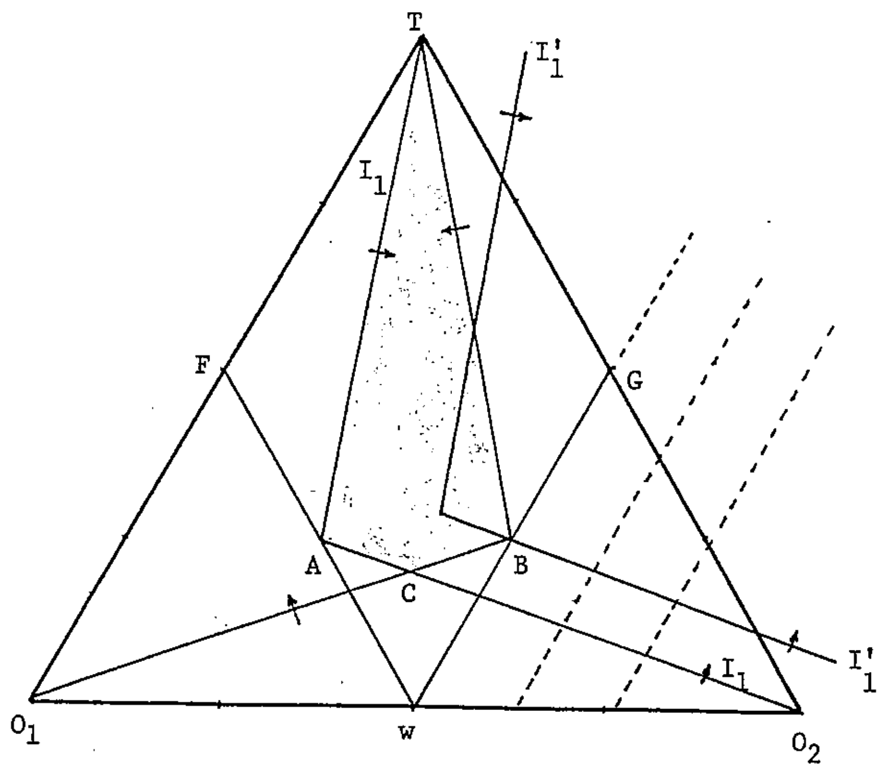

Figure 2: IC mechanisms are not IR in a public good economy.