14 FEASIBLE NASH IMPLEMENTATION OF SOCIAL CHOICE RULES WHEN THE DESIGNER DOES NOT KNOW ENDOWMENTS OR PRODUCTION SETS*

by Leonid Hurwicz, Eric Maskin, and Andrew Postlewaite

1. Introduction

The aim of the present paper is to analyze the problem of assuring the feasibility[1] of a mechanism (game form), implementing in Nash equilibrium[2] a given social choice rule abbreviated as (*SCR*) when the mechanism is constrained as to the way in which it is permitted to depend on endowments or production sets. A social choice rule is a correspondence specifying outcomes considered to be desirable in a given economy (environment). A mechanism is defined by (a) an outcome function and (b) a strategy domain prescribed for each player. Our outcome functions are not permitted to depend at all on the initial endowments or production possibility sets. As to strategy domains, the *i*th agent's strategy domain S^i is only permitted to depend on that agent's endowment (and/or production possibility set), but not on the endowments or production possibility sets of other agents. (For earlier results concerning endoment manipulation, see Postlewaite (1979) and Sertel (1990).)

* Earlier versions of this paper have been circulated since 1979.

A possible (but not necessary) interpretation is that those formulating the rules[3] of the game have no knowledge of the endowments; they may have no way of preventing the players from either understating or even destroying their own endowments, but they may formulate rules making an overstatement of their own endowments impossible, for instance, by requiring the players to "place the claimed endowments on the table." In that case, an agent's strategy domain is limited by his/her (true) endowment. As for the final allocations, these are determined by a formula based only on the agents' claims and hence are not directly dependent on the true values of the endowments.[4] The situation is similar with respect to production possibility claims.

In a pure exchange economy, whether or not the designer knows the individual endowments (as well as the traders' admissible consumption sets), suppose it is required that the outcome function be informationally decentralized, in the sense defined in part I. It is then seen, from Proposition 1 in part I, that feasibility out of equilibrium makes it unavoidable that each unit's strategic domain would depend on its initial endowment. It is furthermore to be noted that this result applies to all informationally decentralized mechanisms, regardless of the equilibrium concept[5] used. A stronger conclusion, at the expense of a stronger assumption is obtained in Proposition 2 of part I. We obtain: (i) certain conditions on the nature of game forms necessary for the implementability of SCRs; (ii) certain conditions that must be satisfied by an SCR in order that it be implementable; (iii) sufficient conditions for the implementability of an SCR, established by constructing an implementing game form.

When a mechanism is said to be feasible, all values of the outcome function, rather than only the equilibrium values, lie in the set of feasible outcomes. We shall denote by $A(e)$ the set of outcomes feasible in the environment e. This defines a correspondence $A(\cdot)$ from the space of environments (economies) into the space Z of outcomes (allocations).

Let us illustrate this in the situation of pure exchange private goods economies without free disposal with n traders. (Part II of the paper is devoted to this case.) Here, the ith trader's *characteristic* e^i is defined by his/her consumption set C^i, initial endowment ω^i, and preference relation R^i, written $e^i = (C^i, \omega_i, R^i)$.[6] The *environment* e is defined as the list of characteristics, i.e., $e = (e^1, \ldots, e^n)$. The space of feasible outcomes in this economy consists of all net trade lists $x = (x^1, \ldots, x^n)$, each x^i an element of the commodity space $\mathbb{R}^m$, satisfying the following two conditions: (a) *Individual feasibility*—every agent remains within his/her consumption set, i.e., $\omega_i - x^i \in C^i$; (b) *compatibility* or *balance*—the sum of all net trades is the null vector of the commodity space, written $\Sigma x^i = 0$.

In earlier mechanism design literature, the balance condition was observed, but not the individual feasibility. This contrasts with the conventional Walrasian auctioneer scenario where the reverse is the case. In the present paper, the emphasis is on mechanisms satisfying both conditions.

Looking at the problem of constructing a feasible game form implementing a given SCR over a class E of environments, we must distinguish situations in which the designer knows the feasible set $A(e)$ for each e in E, i.e., the feasibility correspondence $A(\cdot)$, from those in which the designer has no such information. Maskin's algorithm[7] (1977) for constructing a mechanism implementing a given SCR postulates a class E of environments with a common set A of feasible outcomes known to the designer.[8] In this paper we are interested in the situation where the feasible set is not known to the designer. Since the balance condition does not contain any unknown parameters, we are dealing in our illustrative example with a situation where the designer does not know the traders' initial endowments.

Part II is devoted to pure exchange economies without public goods. Part III (Theorems 8 and Corollary 8.1) deals with public goods, Part IV (Theorem 9) with production. In Part II, to gain insight into the problem, we start with the case where the designer does know preferences, but not the endowments. We then construct two types of endowment revelation games (involving, respectively, the withholding and destruction of endowments), each analogous to Maskin's algorithm for unknown preferences. The strategy space for each trader consists of n-tuples of claimed endowments. Thus, the ith trader claims that the endowment vector is $w_i\cdot = (w_i^1, \ldots, w_i^n)$ where w_i^j is j's endowment according to i's claim. It is assumed that i knows his/her true endowment ω_i. An important restriction imposed on the nature of the strategy space is that a trader cannot exaggerate his/her own endowment; i.e., $w_i i \leqslant \omega_i$. This means that the individual strategy domains depend on the true endowments. In Part I of the paper, it is shown that some such restriction is unavoidable.[9]

Two variants of an endowment game are considered: withholding (section II.A), and destruction (section II.B). When a trader is *withholding* a part of the endowment, he/she (falsely) claims some $w_i^i \leqslant \omega_i$ (so that $w_i^i \neq \omega_i$) as own endowment, but—in addition to the commodity bundle allocated by the outcome function—he/she can also consume the difference $\omega_i - w_i^i$. By contrast, when a trader is *destroying* the part $\omega_i - w_i^i$ of the endowment, this part is not available for consumption. In section II.C, we consider a mixture of withholding and destruction. Implementation under withholding is called W-implementation, that under destruction as Ɖ-implementation. When preferences are assumed known to the designer, they are dealt with respectively in Theorem 1 (section II.A.1) and Theorem 3 (section II.B). Under withholding, we assume individual rationality, under destruction, "non-confiscatoriness" of the social choice function.

The more interesting case is, of course, when the designer knows neither endowments nor preferences. Under withholding, this is referred to as W-R-implementation and is dealt with in Theorem 2, section II.A.2. It is shown there, for the case of withholding, how to deal with this situation. The proof involves combining the game form for the withholding game, for known preferences constructed in section II.A.1, with a Maskin type game form, for situations where endowments but not preferences are assumed known to the designer (see Maskin, 1977; Saijo, 1988; Hurwicz, 1986).

In the appendix to Part II, we exhibit a more specialized mechanism for implementing what we call the constrained Walrasian correspondence, which satisfies Maskin's conditions of no veto-power and monotonicity (Theorems 4, 5, and 6). It is shown (Theorems 5 and 6) that this implementation can be accomplished using a finite-dimensional strategy space, much smaller than the profile spaces used in our other results. We also give an example showing that there are cases in which the Walrasian correspondence (2) is not implementable, because it fails to satisfy the Maskin (necessary) condition of monotonicity.[10] In fact, a slight extension of Theorem 1 in Hurwicz (1979) implies that the constrained Walrasian correspondence is the smallest continuous social choice correspondence, satisfying the conditions of Pareto optimality and individual rationality, which can be Nash-implemented over a sufficiently rich class of economies. Indeed, Theorem 4 states that, for $n > 2$, any *PO*, *IR*, continuous and monotone correspondence contains the constrained Walrasian correspondence.[11]

In Part III, we consider economies with public goods, and in Part IV those with production. Significant results, going beyond ours, in feasible (weakly balanced) implementation, especially of the (constrained) Walrasian correspondence, in economies with production and externalities, are found in Nakamura (1989), and for balanced (i.e., without free disposal) implementation, in economies with production in Hong (1991, 1994).

As in the Groves–Ledyard (1977) treatment of public goods and in Maskin's 1977 algorithm, all our constructions assume that there are at least three agents ($n > 2$). Subsequent to the circulation of earlier versions of this paper, feasible game forms have been constructed for exchange economies with two agents in economies with free disposal[12] (see, in particular, Nakamura, 1989, 1990).

The mechanisms used in our existence proofs are far from informationally efficient. In fact, Page (1989) and Hong and Page (1994) show how the size of the message space can be substantially reduced. In the next section of this introduction, we provide a few additional comments concerning the contents of this paper.

While Theorem 1 only deals with social choice *functions*, it is indicated in the appendix to section II.A.1 how, for the endowment withholding game with preferences known to the designer, the result can be extended to the implementation of social choice *correspondences*. Analogous extensions from SCF's to SCR's (correspondences) seem to be possible for our other cases, but are not dealt with in the paper. However, the implementation results in the appendix at the end of part II deal with the implementation of constrained Walrasian correspondences; that is, it is not assumed that there is a unique constrained Walrasian allocation.

I. THE DEPENDENCE OF STRATEGY DOMAINS ON INITIAL ENDOWMENTS

In what follows, we show that, when the outcome function is privacy preserving with respect to endowments (but possibly "parametric" in the sense of Hurwicz (1972, pp. 310–313), the strategy domain of each person in a pure exchange economy must vary with that person's initial endowment. These results apply to noncooperative games in general and not merely to Nash equilibria. Proposition 1 and the corollary are valid whether or not the designer knows the initial endowments.

We consider a class E of pure exchange economies with the set of goods $L = \{1, \ldots, l\}$. The set of agents is denoted by $N = \{1, \ldots, n\}$. The ith person's true initial endowment is $\mathring{\omega}^i$, but sometimes the circle above ω is omitted. We write $\underline{\omega} = (\omega^1, \ldots, \omega^n)$ and $\omega^i = (\omega^i_1, \ldots, \omega^i_l)$ for each i in N. Each person's consumption set is contained in the nonnegative orthant $\mathbb{R}^l_+$.

Let $E = E^1 \times \ldots \times E^n$, with the generic element of E^i denoted by $e^i = (\omega^i, R^i)$, $\omega^i \in \mathbb{R}^l_{+0}$; here R^i denotes the ith agent's (weak) preference relation, assumed to be reflexive, transitive, and total.

Assumption 1: We assume that for every i in N, every r in L, and every positive number ε, there is $e^i \in E^i$, $e^i = (\omega^i, R^i)$, such that $0 < \omega^i_r < \varepsilon$.

Restricting ourselves, for the sake of simplicity, to single-valued social choice rules (performance correspondences), we denote a social choice function (performance function) by f: $E \to \mathbb{R}^{ln}$. The values of f specify net trades. Feasibility requirements are: for all $e \in E$ and all $r \in L$,

$$\textit{balance} \quad \sum_{i \in N} f_r^i(e) = 0 \tag{1}$$

$$\textit{individual feasibility}: \quad f_r^i(e) \geqslant -\omega_r^i \quad \text{for all} \quad i \in N. \tag{2}$$

where f_r^i denotes the net allocation of the rth good to the ith person, and ω_r^i the initial endowment of the ith person in the rth good.

To avoid triviality, we assume that there is at least one person $i \in N$, a good $r \in L$, and an economy $\bar{e} \in E$, such that, for a social choice rule f implementable on $\bar{e}$,

$$f_r^i(\bar{e}) \neq 0. \tag{3*}$$

From feasibility, it follows that there is at least one person $j \in N$, a good $r \in L$, and an economy $\bar{e} \in E$, such that

$$f_r^j(\bar{e}) < 0. \tag{3}$$

We shall write

$$f_r^j(\bar{e}) = -a, \quad a > 0. \tag{3'}$$

We now define a noncooperative game with the ith strategy domain denoted by S_i. Since the question is whether, or in what way, this domain depends on the initial endowments, we write $S_i = S_i(e^i) = S_i(\omega^i, R^i)$. (That is, the S_i may be "parametric," but must not depend on the characteristics of other agents.) This, of course, does not a priori preclude the possibility that $S_i(\cdot)$ is constant, i.e., that, for any two environments $\bar{e}$, $\bar{\bar{e}}$, we would have $S_i(\bar{e}^i) = S_i(\bar{\bar{e}}^i)$. However, the following proposition shows that, in fact, at least some persons' domains do vary with their own endowments.

Write $S = S(e) = S_1(e^1) \times \ldots \times S_n(e^n)$.

We shall permit the outcome functions to be "parametric," i.e., to depend on the initial endowments, but in a privacy-preserving way. That is, the ith individual's net allocation z^i is given by

$$z^i = h^i(s, e^i), \quad s \in S(e), \quad i \in N.$$

One could, of sourse, confine oneself to "nonparametric" outcome functions where $z^i = \mathring{h}^i(s)$. By permitting the dependence of h^i on $\mathring{\omega}^i$ (perhaps even on e^i), however, we strengthen the result.

We impose on the outcome functions the following feasibility restrictions for all $r \in L$, all $s \in S$, and all $e \in E$:

$$\text{balance} \quad \sum_{i \in N} h_r^i(s, e^i) = 0 \tag{1*}$$

$$\text{individual feasibility:} \quad h_r^i(s, e^i) \geqslant -\mathring{\omega}_r^i \quad \text{for all} \quad i \in N. \tag{2*}$$

We assume that the game form $(h, S(\cdot))$ implements f on E. By definition, this implies that for every e in E, there exists s^* in $S(e)$, such that for every i in N, and for every r in L, $h_r^i(s^*, e^i) = f_r^i(e)$.

Proposition 1. Assume Assumption 1 holds, let $e^* \in E$ and let f satisfying (3*) be implementable on e^*. Let further j, r, e^* and a be those specified in (3′), with $e^* = (\omega^{*i}, R^{*i})_{i \in N}$. Then there exists a strategy n-tuple $s = (s_i)_{i \in N}$ and an economy $e^{**} = (w^{**i}, R^{**i})_{i \in N}$, with $\omega_r^{**j} = \omega_r^{*j}$, while $\omega^{**k} = \omega^{*k}$ for all $k \in N/\{j\}$, such that $s_j \in S_j(e^{*j})$ but $s_j \notin S_j(e^{**j})$.

Proof. Since $(h, S(\cdot))$ implements f on E, there exists s in $S(e^*)$, $s = (s_1, \ldots, s_n)$, $s_i \in S_i(e^{*i})$ for all i in N, and such that, for some j,

$$h_r^j(s, e^{*j}) = f_r^j(e^*) = -a, \quad a > 0$$

and $s_j \in S_j(e^{*j})$. By Assumption (1), there is an environment e^{**} in E, such that

$$0 < \omega_r^{**j} < a,$$

while

$$\omega^{**k} = \omega^{*k} \quad \text{for all} \quad k \in N \backslash \{j\}.$$

By showing that $s_j \notin S_j(e^{**j})$, we shall complete the proof. Suppose, to the contrary, that s_j does belong to $S_j(e^{**j})$. Since the characteristics of others remain unchanged, it follows that $s \in S(e^{**})$. Using the individual feasibility requirement (2*) and previously established relations we obtain

$$h_r^j(s, e^{**j}) \geqq -\omega^{**j}_r > -\mathrm{a} = h_r^j(s, e_j^*),$$

while

$$\Sigma_{k \neq j} h_r^k(s, e^{**k}) = \Sigma_{k \neq j} h_r^k(s, e^{*k}).$$

Adding, we find that

$$\Sigma_{i \in N} h_r^i(s, e^{**i}) > \Sigma_{i \in N} h_r^i(s, e^{*i}),$$

which contradicts the balance requirement in (1*). Q.E.D.

Remark 1. Thus s_j depends on e^j. s_j need not depend on ω_j, but if it does not vary with ω_j, then it must vary with R_j.

Corollary: 1.1: If for *every* person $j \in N$, there exists a good $r \in L$ and an economy $\bar{e} \in E$, such that

$$f_r^j(\bar{e}) \neq 0,$$

then, for every $j \in N$, the domain correspondence $S_j(e^j)$ is non-constant; more specifically, there exists $s^* = (s_i^*)_{i \in N}$ and an economy $\bar{\bar{e}} = (\bar{\bar{\omega}}^i, \bar{\bar{R}}^i)_{i \in N'}$ with $\bar{\bar{\omega}}_r^j \neq \bar{\omega}_r^j$ while $\bar{\bar{\omega}}^k = \bar{\omega}^k$ for all $k \in N \backslash \{j\}$, such that $s_j^* \in S_j(\bar{e}^j)$ but $s_j^* \notin S_j(\bar{\bar{e}}^j)$.

Proof: Follows immediately from the preceding proposition.

Assume now that an agent's strategy is independent of preferences but may depend on his/her endowment, so that i's strategy domain can be written as $S_i(\omega^i)$. We shall next show that, under Assumption 2 on the social choice function (stated below), if, in environment e^* agent i has a greater endowment of a particular good than in environment e^{**}, while the other agents' endowments of all goods are the same, then i's strategy domain $S_i(\omega^{*i})$ must contain elements not present in $S_i(\omega^{**i})$.

To state (2), we first introduce a class of environments. We shall denote by $E/\underset{\rightarrow}{\bar{\omega}}$ the class of all environments in E whose endowment profile equals $\underset{\rightarrow}{\bar{\omega}}$, while preferences vary.

Hence, $f_r^i(E/\underset{\rightarrow}{\bar{\omega}})$ is the set of net allocations in the rth good to the ith agent produced by the performance function f, as environments trace out the class $E/\underset{\rightarrow}{\bar{\omega}}$. The additional assumption is as follows:

Assumption 2:

$$\forall i \in N, r \in L, \bar{\omega}_r^i \geqslant 0,$$
$$\inf f_r^i(E/\underset{\rightarrow}{\bar{\omega}}) = -\bar{\omega}_r^i.$$

Remark 2. It appears that, when the postulated class of environments is sufficiently rich, Assumption 2 is satisfied for social choice functions which always yield allocations that are Pareto optimal and individually rational.

Proposition 2. Assume Assumption 2 holds, and let e^*, e^{**} be two environments such that, for some agent i and a good r, $\omega^{*i}_r > \omega^{**i}_r$, while $w^{*j} = w^{**j}$ for all j not equal to i. Then there exists a strategy is available to i in e^* but not in e^{**}.

Proof. By Assumption 2, there is a sequence $\{e^{*k}\}$, $k = 1, 2, \ldots$ of environments[13] such that each e^{*k} belongs to the class E/ω^*, so that for each e^{*k} the endowment profile is ω^*, by individual feasibility $f^i_r(e^{*k}) \geqq -\omega^{*i}_r$, and, by Assumption 2, lim $f^i_r(e^{*k}) - -\omega^{*i}_k$ as k tends to infinity.

Write $c = \omega^{*i}_r - \omega^{**i}_r$. By hypothesis, $c > 0$. Then there exists a number c', with $0 \leqq c' < c$, such that, for a sufficiently large integer K, we have

$$f^i_r(e^K) = -\omega^{*i}_r + c'.$$

Write $)i(= (1, \ldots, i-1, i+1, \ldots, n)$. Since h implements f, there exists a strategy n-tuple $s^{*K} = \langle s^{*Ki}, s^{*K,)i(} \rangle$ such that (supressing in our notation the possible dependence of h^i on e^i) $h(s^{*K}) = f(e^{*K})$, and hence

$$h^i_r(s^{*Ki}, s^{*K,)i(}) = -\omega^{*i}_r + c'.$$

Hence,

$$s^{*Ki} \in S_i(\omega^{*i}).$$

But, since $c' < c$, it follows from the definition of c that

$$-\omega^{*i}_r + c' < -\omega^{**i}_r,$$

and hence $h^i_r(s^{*K}) < -\omega^i_r$, which violates the individual feasibility requirement for agent i in the environment e^{**}. Since s^{*K}, $)i($ was available members of $)i($ in e^*, and $S_j(\omega^{**j}) = S_j(\omega^{*j})$ for j in $)i($ (since, by hypothesis, $\omega^{**j} = \omega^{*j}$ for j not equal to i), we conclude that

$$s^{*Ki} \notin S_i(\omega^{**i}).$$

Q.E.D.

In what follows we, sketch the construction used in Theorem 1 (where endowments may be withheld, but preference profiles are known).

II. PURE EXCHANGE IN PRIVATE GOODS
II.A. WITHHOLDING
II.A.1. THE ENDOWMENT GAME (WITH ENDOWMENTS UNKNOWN BUT PREFERENCES KNOWN)

Notation and Assumptions

(i) VECTORS

Let m be a positive integer. Then

$\mathbb{R}^m = \{x | x = (x^1, \ldots, x^m), x^r \text{ a real number for all } 1 \leqslant r \leqslant m\}$.

Let $x, y \in \mathbb{R}^m$. Then $x \geqslant y$ means $x^r \geqslant y^r$ for all $1 \leqslant r \leqslant m$; $x \geq y$ means $x \geqslant y$, but $x \neq y$; and $x > y$ means $x^r > y^r$ for all $1 \leqslant r \leqslant m$. $\mathbb{R}^m_+ = \{x \in \mathbb{R}^m | x \geqslant 0\}$; $\mathbb{R}^m_{++} = \{x \in \mathbb{R}^m | x > 0\}$; $\mathbb{R}^l_{+0} = \mathbb{R}^l_+ \setminus \{0\}$, so that $x \in \mathbb{R}^l_{+0}$ means $x \geq$; 0 $\mathbb{R}^{ln}_{+0} = \mathbb{R}^l_{+0} \times \ldots \times \mathbb{R}^l_{+0}$ (n times). For $a, b \in \mathbb{R}^m$, $[a, b] = \{x \in \mathbb{R}^m | a \leqslant x \leqslant b\}$, $(a, b] = \{x \in \mathbb{R}^m | a \leqslant x \leqslant b, x \neq a\}$.

(ii) ENVIRONMENT

$N = \{1, \ldots, n\}$ = the set of agents; $n \geqslant 3$.

$L = \{1, \ldots, l\}$ = the set of goods.

$\mathring{\omega}_i$ = the true endowment of agent i; $\mathring{\omega}_i \in \mathbb{R}^l_{+0}$ for all i.

$\underset{\rightarrow}{\mathring{\omega}} = (\mathring{\omega}_1, \ldots, \mathring{\omega}_n)$ = the endowment profile.

$\vec{\mathbb{R}}^l_+$ is assumed to be the individually feasible consumption set for every agent.

$\mathring{R}_i$ = the true preference relation of agent i on $\mathbb{R}^l_+ \times \mathbb{R}^l_+$.

$\mathring{P}_i$ = the true strict preference of agent i (i.e., $x\mathring{P}_iy$ iff $x\mathring{R}_iy$ but not $y\mathring{R}_ix$).

$\mathring{R}_i$ is reflexive, transitive, and convex on $\mathbb{R}^l_+ \times \mathbb{R}^l_+$ (i.e., preferences are selfish); $\mathring{R}_i$ is assumed strictly increasing in all goods for all agents (i.e., $x \geq y$ implies $x\mathring{P}_iy$).

(iii) PERFORMANCE

$Z = \{\underset{\rightarrow}{z} \in \mathbb{R}^{ln} | \underset{\rightarrow}{z} = (z_1, \ldots, z_n); z_i \in \mathbb{R}^l, \forall i \in N; \Sigma_{i \in N} z_i = 0\}$ = the set of balanced *net* trades[14]. Given a configuration $z = (z_1, \ldots, z_n)$ of net trades, agent i's final (total) holdings are $\mathring{\omega}_i + z_i$.

f = the performance function[15] (social choice rule).

$f: \mathbb{R}^{ln}_{+0} \rightarrow Z$.

Let $\underset{\rightarrow}{v} = (v^1, \ldots, v^n) \in \mathbb{R}^{ln}_{+0}$; $v^i \in \mathbb{R}^l_{+0}, \forall i \in N$.

$f = (f_1, \ldots, f_n)$; if $z = (z_1, \ldots, z_n) = f(\underset{\rightarrow}{v})$, then $z_i = f_i(\underset{\rightarrow}{v})$; so, $f_i: \mathbb{R}^{ln}_{+0} \rightarrow \mathbb{R}^l$.

$f(\underset{\rightarrow}{\mathring{\omega}})$ is interpreted as the optimal[16] net trade configuration when the true endowment profile is $\underset{\rightarrow}{\mathring{\omega}}$; $f_i(\underset{\rightarrow}{\mathring{\omega}})$ is agent i's optimal net trade for the profile $\underset{\rightarrow}{\mathring{\omega}}$.

It is assumed that $v^i + f_i(\underline{v}) \geqq 0$, for all i and all $\underline{v} \in \mathbb{R}^{ln}_{+0}$.

(iv) STRATEGIES AND OUTCOME FUNCTIONS

For each $i \in N$, let T_i be an arbitrary nonempty set. It is assumed that the strategy space S_i of agent i is of the form

$$S_i = (0, \mathring{\omega}_i] \times T_i,$$

where T_i is independent of $\underline{\mathring{\omega}}$.
We also define $S = S_1 \times \ldots \times S_n$.
Generically, we write for the corresponding elements

$$s_i = (w_i^i, t_i),$$
$$s = (s_1, \ldots, s_n),$$

and[17]

$$s = (s_i; s_{)i(}),$$

where $t_i \in T_i$, $0 \leqslant w_i^i \leqslant \mathring{\omega}_i$, $s_i \in S_i$, $s_{)i(} \in \prod_{j \neq i} S_j$[18], $s \in S$.

If we interpret the component w_i^i of $s_i = (w_i^i, t_i)$ as a profession of agent i's endowment, the inequality $0 \leqslant w_i^i \leqslant \mathring{\omega}_i$ means that the agent cannot overstate his own endowment; on the other hand, the endowment can be understated (in one or more commodity components), but the claimed endowment w_i^i (like the true endowment $\mathring{\omega}_i$) must be semi-positive (i.e., different from the null vector and nonnegative in all commodity components).[19]

h = the outcome function (game form).

$h: S \to Z$.

$h = (h_1, \ldots, h_n)$; if $z = (z_1, \ldots, z_n) = h(s)$, then

$h_i(s) = z_i$; so, $h_i: S \to \mathbb{R}^l$.

$h(s)$ = then net trade configuration resulting from the strategic configuration s.

$h_i(s)$ = agent i's net trade resulting from the strategic configuration s.

Given s, agent i's final (total) holdings are

$$\mathring{\omega}_i + h_i(s).$$

For net trades $z_i', z_i'' \in z$, we shall sometimes write $z_i' \mathring{R}_i z_i''$ to mean $(\mathring{\omega}_i + z_i') \mathring{R}_i (\mathring{\omega}_i + z_i'')$, etc.

It will be assumed that, for all i, $s = (s_i, s_{)i(})$, $s_i = (w_i^i, t_i)$,

$$w_i^i + h_i(s) \geqq 0.$$

That is, the outcome function will never deprive the agent of goods in excess of his claimed endowment.

Since $w_i^i \leqq \mathring{\omega}_i$, a fortiori, the outcome function will never require the agent to give up more of any good than there was in the true initial endowment. Thus, individual feasibility is assured.

Furthermore, since h takes its values in Z, we have $\Sigma_{i \in N} h_i(s) = 0$ for all $s \in S$; hence, balance is also assured. Thus, feasibility is preserved at all points of the strategy space, out of equilibrium as well as at equilibrium.

On the other hand, since $w_i^i \leqslant \mathring{\omega}_i$ is permitted, the agent is able to *withhold* a part of the true endowment. Complete withholding is ruled out by the requirement $w_i^i \geqslant 0$.

We shall say that the outcome function h W-implements[20] (in Nash equilibrium (NE)) the performance function f for $\underline{\mathring{R}}$ of true preference profiles, if: for any true endowment profile $\underline{\mathring{\omega}}$, (1) an NE exists, and, further, (2) for any NE configuration s^* of strategies, $\underline{\mathring{\omega}} + h(s^*) = \underline{\mathring{\omega}} + f(\underline{\mathring{\omega}})$; i.e., every Nash outcome is f-optimal.

Definition 1. f is individually rational (IR) if, for all i in N, and all $\underline{\mathring{\omega}} \in \mathbb{R}^{ln}_{+0}$, $(\mathring{\omega}_i + f_i(\underline{\mathring{\omega}}))\mathring{R}_i\mathring{\omega}_i$.

Proposition 3: If preference $\underline{\mathring{R}}$ are continuous and nondecreasing, and if f is W-implementable (in NE) for $\underline{\mathring{R}}$, then f is individually rational (IR). ("W-implementable" stands for "withholding-implementable.")

Proof: Suppose f is implementable by h: $S \to \mathbb{R}^{ln}$, but is not IR. Then there exist $\underline{\mathring{\omega}} \in \mathbb{R}^{ln}_{+0}$ and $i \in N$ such that[21] $0 \, \mathring{P}_i f_i(\underline{\mathring{\omega}})$. Since h implements f, there exists an NE $s^* = (s_1^*, \ldots, s_n^*) \in S$ for $(\underline{\mathring{\omega}}, \underline{\mathring{R}})$, such that $h_i(s^*) = f_i(\underline{\mathring{\omega}})$. Hence $0 \, \mathring{P}_i h_i(s^*)$.

Then, by the assumed continuity of R_i, the semi-positivity of $\mathring{\omega}^i$, and the nondecreasing preferences, there exists a real number $\varepsilon > 0$ and an i-feasible net trade $b = (b_1, \ldots, b_l)$, where[22] $b \leqslant 0$, $\| b \| = \epsilon$, and, furthermore,

$$b\mathring{P}_i h_i(s^*).$$

But, for any $t_i \in T_i$ and[23] $s_i = (-b, t_i)$, we have $h_i(s_i, s^*_{)i(}) \geqq b$, since $w_i^i + h_i(s') \geqq 0$ for all s'. Hence, $h_i(s_i, s^*_{)i(})\mathring{R}_i b \mathring{P}_i h_i(s^*)$, which contradicts the supposition that s_i^* is an NE strategy.

Q.E.D.

Definition 2.[24] f is *non-confiscatory* (NC) if $\forall i \in N$, $\forall \underset{\rightarrow}{\mathring{\omega}} \in \mathbb{R}^{ln}_{+0}$, $\mathring{\omega}_i + f_i(\underset{\rightarrow}{\mathring{\omega}}) \geqslant 0$.

Remark 3. It may be noted that, when $\mathring{\omega}_i \geqslant 0$ and preferences are strictly increasing, IR implies NC. Clearly, however, f may be NC and not IR.

Theorem 1: (1) If f is IR, and if the assumptions (including strictly increasing[25] preferences) preceding the above proposition are satisfied, then f is W-implementable (in NE). ("W-implementable" stands for "withholding-implementable.")

(2) If preferences are continuous[26] and strictly increasing, f is W-implementable if and only if it is IR (individually rational).

Proof: The proof of (2) follows from (1) and the preceding proposition. To establish (1), we construct an outcome function h, which W-implements f.

For $i \in N$, let the strategy space of the ith agent be

$$S_i = \{(w_i^i, \ldots, W_i^n) \in \mathbb{R}^{ln}_{+0} | w_i^j \in \mathbb{R}^l_{+0}, 0 \leqslant w_i^i \leqq \mathring{\omega}_i, i, j \in N\}.$$

For $s_i \in S_i$, we shall sometimes write

$$s_i = w_i = (w_i^i, w_i^{)i(}),$$

where

$$w_i^{)i(} = (w_i^1, \ldots, w_i^{i+1}, w_i^{i+1}, \ldots, w_i^n)$$

and

$$w_k = (w_k{}^1, \ldots, w_k^n), \quad \text{with} \quad w_k^r \in \mathbb{R}^l_{+0} \quad \text{for all} \quad k \in N, r \in N.$$

We interpret w_i^j as agent i's statement about j's endowment. For all $i, j \in N$, it is assumed that $w_i^j \geqslant 0$; i.e., each agents's statement attributes to everybody, including himself, positive holdings of some commodity. In the spirit of informational decentralization (privacy-preserving property of the mechanism), it is assumed that an agent has no useable information about the other agents' endowments. Therefore, for $j \neq$ i, there is no upper bound on w_i^j. By contrast, an agent is assumed to know his own endowment. While he may conceal or destroy a part of it, he is not permitted to exaggerate it; hence, the requirement that $W_i^i \leqq \mathring{\omega}_i$ for all $i \in N$. (We might, for instance, imagine that the rules of the game require that the agent "put on the table" the reported amount w_i^i.)

Notice that this S_i has the structure of the strategy space $S_i = (0, \mathring{\omega}_i] \times$

T_i, introduced in the previous section. In $s_i = (w_i^i, w_i^{)i(})$, the component $w_i^{)i(}$ corresponds to t_i in $s_i = (w_i^i, t_i)$.

We will define the outcome function $h(w_1, \ldots, w_n)$, with $w_i \in \mathbb{R}_{+0}^{ln}$ for each $i \in N$, by the following rules:

(a) (*The case of unanimity*)

If, for some $\underline{v} \in \mathbb{R}_{+0}^{ln}$, $s = (s_1, \ldots, s_n) \in S$, $s_i = \underline{v}$ for all $i \in N$, then

$$h(s) = f(\underline{v}).$$

To state rules (b) and (c), we use the following notation.

Let $s = (s_1, \ldots, s_n) \in S$, $s_j = w_j = (w_j^1, \ldots, w_j^n)$, $w_j^k \in \mathbb{R}_{+0}^l$, $k, j \in N$. We define

$$M(s) = \{i \in N | w_i^i \geq w_j^i, \forall j \neq i, j \in N\};$$

$$w(s) = \sum_{i \in N} w_i^i;$$

$$\beta_i(s) = \sum_{\substack{j \in N \\ j \neq i}} \sum_{\substack{k \in N \\ k \neq i}} \| w_j^j - w_k^j \|, \; i \in N.$$

(When there is no danger of confusion, we suppress the argument s and write, respectively M, w, β_i.)

The second rule is then as follows:

(b) If $M(s) = \varnothing$, but there is no $\underline{v}$ such that $s_i = \underline{v}$ for all $i \in N$, then $\Sigma_{j \in N} \beta_j(s) > 0$, and we set

$$h_i(s) = \left[\frac{\beta_i(s)}{\sum_{j \in N} \beta_j(s)} \right] w(s) - w_i^i, \; i \in N.$$

The third rule is:

(c) If $M(s) \neq \varnothing$, we set

$$h_i(s) = \begin{cases} \dfrac{1}{\# M(s)} w(s) - w_i^i & \text{for } i \in M(s) \\ -w_i^i & \text{for } i \notin M(s). \end{cases}$$

We shall now prove three claims which together imply that this outcome function h does W-implement (in NE) the performance function f.

These claims are: (1) the unanimous announcement of the true endowment profile by all agents is a Nash equilibrium; (2) the unanimous announcement of a false endowment profile is not a Nash equilibrium; and (3) in the absence of unanimity, there is no Nash equilibrium.

Claim 1: The unanimous announcement of the true endowment profile by all agents is an NE. That is,

$$\text{if} \quad s_i = \underset{\to}{\overset{\circ}{\omega}}, \quad \forall i \in N,$$

$$\text{then} \quad s = (s_1, \ldots, s_n) \quad \text{is a NE for } \underset{\to}{\overset{\circ}{\omega}}.$$

Proof of Claim 1: For such a unanimous announcement s of the true endowment profile $\underset{\to}{\overset{\circ}{\omega}}$, by rule (a),

$$h(s) = f(\underset{\to}{\overset{\circ}{\omega}}).$$

Suppose s is not an NE. Then there is an agent j and some $\tilde{s}_j$, such that

$$(+++) \qquad h_j(\tilde{s}_j, s_{)j(})\overset{\circ}{P}_j h_j(s_j, s_{)j(})$$

where $(s_j, s_{)j(}) = s$. Necessarily, $\tilde{s}_j \neq \underset{\to}{\overset{\circ}{\omega}}$, and also, for $\tilde{s}_j = (\tilde{w}_j^j, \tilde{w}_{)j(})$, by the non-exaggeration rule,

$$\tilde{w}_j^j \leqslant \overset{\circ}{\omega}_j.$$

Writing $\tilde{s} = (\tilde{s}_j, \tilde{s}_{)j(}) = (\tilde{s}_j, s_{)j(})$, so that $\tilde{s}_r = s_r$ for all $r \neq j$, it follows[27] that

$$j \notin M(\tilde{s}).$$

Let k be any agent other than j; i.e., $k \neq j$. Since[28] $n \geqslant 3$, there exists a third agent m, with $m \neq j$ and $m \neq k$.

Now $\tilde{s}_r = s_r = \underset{\to}{\overset{\circ}{\omega}}$ for all $r \neq j$. Hence,

$$\tilde{s}_m = \underset{\to}{\overset{\circ}{\omega}} \quad \text{and} \quad \tilde{s}_k = \underset{\to}{\overset{\circ}{\omega}}.$$

Since $\tilde{s}_r = (\overset{\circ}{\omega}_1, \ldots, \overset{\circ}{\omega}_n) = (w_r^1, \ldots, w_r^n)$, $\forall\, r \neq j$, we have

$$w_k^k = w_m^k,$$

and hence,

$$k \notin M(\tilde{s}).$$

Since k was an arbitrary agent other than j, it follows that no agent other than j is in $M(\tilde{s})$, and we have seen above that j is not in $M(\tilde{s})$. So,

$$M(\tilde{s}) = \varnothing.$$

Thus, $M(\tilde{s}) = \varnothing$ but $\tilde{s}$ is not unanimous, so rule (b) applies to $\tilde{s}$.

Since $s_i = \tilde{s}_i$ for all $i \neq j$, we have

$$\beta_j(\tilde{s}) = \sum_{k \neq j} \sum_{i \neq j} \| w_k^k - w_i^k \| = 0,$$

and so

$$h_j(\tilde{s}) = 0 \cdot w(\tilde{s}) - \tilde{w}_j^j = - \tilde{w}_j^j.$$

Since f is IR,

$$h_j(s) = h_j(\underset{\rightarrow}{\mathring{\omega}}, \ldots, \underset{\rightarrow}{\mathring{\omega}}) = f_j(\underset{\rightarrow}{\mathring{\omega}}) \mathring{R}_j 0.$$

Because preferences are strictly increasing and $\tilde{w}_j^j \geqslant 0$,

$$0 \mathring{P}_j(-\tilde{w}_j^j).$$

Therefore,

$$h_j(s) \mathring{P}_j(-\tilde{w}_j^j),$$

and so

$$h_j(s) \mathring{P}_j h_j(\tilde{s}),$$

which contradicts the above supposition that $h_j(\tilde{s}) \mathring{P}_j h_j(s)$. Hence s is an NE.

Q.E.D.

Claim 2: The unanimous announcement of a false endowment profile is not an NE. That is, if $s = (\underset{\rightarrow}{v}, \ldots, \underset{\rightarrow}{v})$, $\underset{\rightarrow}{v} \in \mathbb{R}_{+0}^{ln}$ with $\underset{\rightarrow}{v} \neq \underset{\rightarrow}{\mathring{\omega}}$, then s is not an NE.

Proof of Claim 2: Since s is unaminous, rule (a) again applies, and so

$$h(s) = f(\underset{\rightarrow}{v}).$$

Suppose s is a Nash equilibrium.

Since $\underset{\rightarrow}{v}$ is not the true endowment profile, and agents cannot overstate their endowments, then there must be an agent i such that

$$w_r^i \leqslant \mathring{\omega}_i, \quad \forall r \in N.$$

(We have $\underset{\rightarrow}{v} = (v^1, \ldots, v^n)$, $v^k = w_r^k$, $\forall k, r \in N$. Since $\underset{\rightarrow}{v} \neq \underset{\rightarrow}{\mathring{\omega}}$, it must be that, for some i, $v^i \neq \mathring{\omega}_i$. $\therefore$ $w_i^i \neq \mathring{\omega}_i$, $\therefore$ $w_i^i \leqslant \mathring{\omega}_i$. But $w_r^i = v^i = w_i^i$, $\forall r \in N$. Therefore, $w_r^i \leqslant \mathring{\omega}_i$, $\forall r \in N$.)

Consider $\tilde{s} = (\tilde{s}_i, \tilde{s}_{)i(})$ such that

$$\tilde{s}_{)i(} = s_{)i(},$$
$$\tilde{s}_i^k = s_i^k \quad \text{for all} \quad k \neq i$$

while

$$\tilde{s}_i^i = \mathring{\omega}_i.$$

(That is, $\tilde{s}_i^i \neq s_i^i$ since $s_i^i = w_i^i \leqslant \mathring{\omega}_i$.)

Then

$$M(\tilde{s}) = \{i\},$$

and, by rule (c)

$$h_i(\tilde{s}) = w(\tilde{s}) - \tilde{w}_i^i = \sum_{k \neq i} w_k^k = \sum_{j \neq i} v^j$$

We shall show below that

$$(+) \qquad h_i(\tilde{s}) \geqslant h_i(s).$$

Since preferences are strictly increasing, the inequality (+) implies

$$h_i(\tilde{s}) \mathring{P}_i h_i(s).$$

Therefore, when (+) holds, agent i has an incentive to deviate from s_i, and so s is not an NE. That is, Claim 2 follows.

To establish (+), we note that, since, fo our outcome function $h(\cdot)$, $h_j(s') \geqq -w_j^j$, $\forall j$, and $\Sigma_{k \in N} h_k(s') = 0$, $\forall s' \in S$,[29] we have[30]

$$h_i(s') \leqq \sum_{j \neq i} w_j^j, \quad \forall\, s' \in S.$$

But $s = (\underset{\rightarrow}{v}, \underset{\rightarrow}{v}, \ldots, \underset{\rightarrow}{v})$, $\underset{\rightarrow}{v} = (v^1, \ldots, v^n)$ implies $v^k = w_k^k$ for all $k \in N$; hence

$$h_i(s) \leqq \sum_{j \neq i} v^j.$$

Suppose

$$(++) \qquad h_i(s) = \sum_{j \neq i} v^j.$$

We shall show that (++) cannot be true. Then, from the inequality in the preceding line, it will follow that

$$h_i(s) \leqslant \sum_{j \neq i} v^j.$$

But we have already shown above that $h_i(\tilde{s}) = \Sigma_{j \neq i} v^j$. Hence, $h_i(s) \leqslant h_i(\tilde{s})$, which is the inequality (+) above. It remains to show that (++) yields a contradiction.

Writing

$$h_j(s) = x_j, \quad j \neq i,$$

the balance requirement then yields

$$\sum_{j \neq i} x_j + \sum_{j \neq i} v^j = 0.$$

But, $x_j \geqq -v^j$, so that

$$x_j = -v^j + \varepsilon_j \quad j \neq i$$

$$\varepsilon_j \geqq 0.$$

Hence, the balance equation can be written as

$$\sum_{j \neq i} (-v^j + \varepsilon_j) + \sum_{j \neq i} v^j = 0,$$

and this implies $\varepsilon_j = 0$, $\forall j \neq i$; hence,

$$x_j = -v^j, \quad \forall\, j \neq i.$$

That is, if $h_i(s) = \sum_{j \neq i} v^j$,

then

(*) $$h_j(s) = -v^j, \quad \forall\, j \neq i.$$

As noted in Remark 3, before Theorem 1, under our assumption, IR (individual rationality) implies NC (non-confiscatority), so that

$$v^j + f_j(\underset{\sim}{v}) \geqslant 0.$$

But here

$$h_j(s) = f_j(\underset{\sim}{v}).$$

Hence

$$v^j + h_j(s) \geqslant 0$$

which contradicts (*).

Q.E.D.

Claim 3: In the absence of unanimity there is no NE. That is, if for some $i, j \in N$, $s_i \neq s_j$, then $s = (s_1, \ldots, s_n)$ is not an NE.

Proof of Claim 3: Let $s = (s_1, \ldots, s_n) = (w_1, \ldots, w_n)$ with $s_i \neq s_j$ for some $i, j \in N$. We consider three cases: (i) $M(s) = N$; (ii) $M(s) \neq \varnothing$, $M(s) \neq N$; (iii) $M(s) = \varnothing$.

(i) Suppose first that $M(s) = N$. Then consider $\tilde{s}$ with

$$\tilde{s}_k = s_k \quad \text{for all} \quad k \neq 1,$$
$$\tilde{s}_1^q = s_q^q \quad \text{for all} \quad q \in N.$$

(We shall sometimes write $\tilde{s}_p = \tilde{w}_p$, $p \in N$.)

For any agent $r \neq 1$,

$$r \notin M(\tilde{s}),$$

since $\tilde{s}_1^r = \tilde{s}_r^r$.

On the other hand, we shall show that

$$1 \in M(\tilde{s}).$$

Notice that $1 \in M(s)$, since $N = M(s)$ by hypothesis. Hence, by definition of $M(\cdot)$,

$$s_1^1 \geqslant s_r^1, \quad \forall r \neq 1.$$

Thus, by the construction of $\tilde{s}$,

$$\tilde{s}_1^1 \geqslant \tilde{s}_r^1,$$

and so

$$1 \in M(\tilde{s});$$

therefore, rule (c) applies to $\tilde{s}$.

Since it was shown previously that nobody else belongs to $M(\tilde{s})$, we have now established that

$$M(\tilde{s}) = \{1\}.$$

Rule (c) implies therefore

$$h_1(\tilde{s}) = 1 \cdot w(\tilde{s}) - \tilde{w}_1^1 = w(s) - w_1^1 = \sum_{k \neq 1} w_k^k.$$

But $h_1(s) \leqslant \Sigma_{k \neq 1} w_k^k$ because $M(s) = N$, so that, under s, part of $\Sigma_{k \neq 1} w_k^k$ was allocated to persons other than 1. [That is, $\beta_k(s) > 0$, $\forall k \neq 1$.]

Therefore,

$$h_1(\tilde{s}) \geqslant h_1(s)$$

and consequently, because of strictly increasing preferences,

$$h_1(\tilde{s}) \mathring{P}_1 h_1(s).$$

Hence, in case (i), s is not an NE.

(ii) Suppose now that $M(s) \neq \emptyset$, $M(s) \neq N$. Since $M(s) \neq \emptyset$, rule (c) applies to s. Because $M(s) \neq N$, there is an agent $j \notin M(s)$ who, by rule (c), gets

$$h_j(s) = -w_j^j.$$

Now consider $\tilde{s}$ where, for all k and all $i \neq j$,

$$\tilde{s}_i^k = s_i^k,$$

$$\tilde{s}_j^j = s_j^j,$$

and

$$\tilde{s}_j^k = s_k^k.$$

For any $r \neq j$, we have

$$\tilde{s}_j^r = \tilde{s}_r^r,$$

and so, by definition of $M(\cdot)$,

$$r \notin M(\tilde{s}) \quad \text{for all} \quad r \neq j.$$

Furthermore, since (by construction) $j \notin M(s)$ and $\tilde{w}_i^j = w_i^j$ for all $i \neq j$, we have $j \notin M(\tilde{s})$.

Thus,

$$M(\tilde{s}) = \emptyset,$$

and so either rule (a) or rule (b) applies to $\tilde{s}$. But rule (a) cannot be applicable because unanimity in $\tilde{s}$ is impossible: since $n \geqq 3$ and $M(s) \neq \emptyset$, there is a person $k \in M(s)$, $k \neq j$, and a person $i \neq j$, $i \neq k$ such that,

$$w_k^k \geq w_i^k;$$

hence $s_k \neq s_i$. But, w_k^k and w_i^k are unchanged in $\tilde{s}$, and so $\tilde{s}_k \neq \tilde{s}_i$. Hence, there is no unanimity in $\tilde{s}$ and rule (a) does not apply to $\tilde{s}$. Hence, rule (b) applies to $\tilde{s}$.

For agents j, k, and i just referred to, we have

$$\beta_j(\tilde{s}) \geqq \| w_k^k - w_i^k \| > 0.$$

Since $w(s) \geq 0$, it follows that

$$h_j(\tilde{s}) \geq -w_j^j = h_j(s),$$

and so, by the assumption of strictly increasing preferences,

$$h_j(\tilde{s}) \mathring{P}_j h_j(s).$$

Hence, in case (ii), s is not an NE.

(iii) Finally, suppose that $M(s) = \emptyset$ and s *is not unanimous*. Since, by the hypothesis of Claim 3, not all announced profiles are the same, there exist agents i and j, $i \neq j$, with

$$w_i^i \neq w_j^i,$$

We now distinguish two subcases, according to whether $\beta_j(s) = 0$ or $\beta_j(s) > 0$.

Subcase (iii.1): $\beta_j(s) = 0$.

Consider $\tilde{\tilde{s}}$ defined by

$$\tilde{\tilde{s}}_k = s_k \quad \text{for all} \quad k \neq j$$

$$\tilde{\tilde{w}}_j^j = \frac{1}{2} w_j^j$$

$$\tilde{\tilde{w}}_j^r = w_j^r \quad \text{for all} \quad r \neq j.$$

We note that, since s is not unanimous and $M(s)$ is empty, Rule (b) applies to s, and hence

$$h_j(s) = -w_j^j.$$

But, also, $\tilde{\tilde{s}}$ is not unanimous, because $\tilde{\tilde{w}}_i^i = w_i^i$, $\tilde{\tilde{w}}_j^i = w_j^i$ by construction,[31] and $w_i^i \neq w_j^i$ by the above hypothesis.

Also, $M(\tilde{\tilde{s}})$ is empty because $M(s)$ is empty, and the change from w_j^j to $\tilde{\tilde{w}}_j^j = \frac{1}{2} w_j^j$ (while $\tilde{\tilde{w}}_r^j = w_r^j$ for $r \neq j$) does not enlarge the set M. Hence, Rule (b) also applies to $\tilde{\tilde{s}}$. Now, since $\tilde{\tilde{s}}_k = s_k$ for $k \neq j$, $\beta_j(s) = 0$ implies $\beta_j(\tilde{\tilde{s}}) = 0$. Therefore,

$$h_j(\tilde{\tilde{s}}) = -\tilde{\tilde{w}}_j^j.$$

But,

$$-\tilde{\tilde{w}}_j^j = -\frac{1}{2} w_j^j \geqslant -w_j^j,$$

because, by assumptions on messages, $w_j^j \geqslant 0$. Hence, by the assumption of strictly increasing preferences,

$$h_j(\tilde{\tilde{s}}) \mathring{P}_j h_j(s).$$

So, $\tilde{\tilde{s}}$ is better than s for agent j, and hence s is not a Nash equilibrium.

Subcase (iii.2): $\beta_j(s) > 0$.

In this situation, consider $\tilde{s}$, such that

$$\tilde{s}_k = s_k \quad \text{for all} \quad k \neq j$$

and

$$\tilde{s}_j^r = s_r^r \quad \text{for all} \quad r.$$

By construction, $\beta_j(\tilde{s}) = \beta_j(s) > 0$ and $\Sigma_{k \neq j} \beta_k(\tilde{s}) < \Sigma_{k \neq j} \beta_k(s)$. Also, $M(\tilde{s}) = M(s) = \emptyset$, so Rule (b) applies to both $\tilde{s}$ and s. Therefore, $h_j(\tilde{s}) \geqslant h_j(s)$. And so, again by the assumption of strictly increasing preferences, s is not a Nash equilibrium.

Q.E.D.

[APPENDIX TO SECTION II.A.1] AN ENDOWMENT WITHHOLDING GAME FORM WHEN THE SCR IS A CORRESPONDENCE

Player i has a strategy "vector" of the form

$$s_i = (w_i, a_i, m_i),$$

where w_i is an endowment profile, a_i an element of the outcome space, and m_i an integer between 1 and n (inclusive). The SCR, denoted by F, is a correspondence from the space of endowment profiles into the outcome space. Define $m = (m_1, \ldots, m_n)$, and

$$R(m) = 1 + (\Sigma_{k \text{ in } N} m_k) \text{ modulo } n.$$

Rule (a.1~). There exist an endowment profile v, outcomes a and a', and integers $m_1, \ldots, m_n$ in N, and an agent j,

such that

$$s_i = (v, a, m_i) \quad \text{for all} \quad i \text{ in } N \backslash \{j\},$$

and

$$s_j = (v, a', m_j).$$

Then

$$h(s) = a \quad \text{if} \quad a \text{ is in } F(v),$$

while, for a not in $F(v)$,

$$h_i(s) = \Sigma w_r^r \quad \text{if} \quad i = R(m),$$

and

$$h_i(s) = -w_i^i \quad \text{if} \quad i \neq R(m).$$

Remark. Note that a and a' may but need not be equal.

Rule (a.2~). If there is a profile v, outcomes $a_1, \ldots, a_n$, and integers $m_1, \ldots, m_n$ from N such that at least three of the a_i's are distinct and

$$s_i = (v, a_i, m_i) \quad \text{for each } i \text{ in } N,$$

then

$$h_i(s) = \Sigma w_j^j \quad \text{if} \quad i = R(m),$$

and

$$h_i(s) = -w_i^i \quad \text{if} \quad i \neq R(m).$$

Rule (b~). Same as rule (b) in the proof of Theorem I.

Rule (c~). Same as rule (c) in the proof of Theorem I.

II. PURE EXCHANGE IN PRIVATE GOODS (c't'd)
II.A. WITHHOLDING (c't'd)
II.A.2. THE GAME WITH BOTH PREFERENCES AND ENDOWMENTS UNKNOWN TO THE DESIGNER

Notation and Assumptions

Here the performance correspondence (SCR) f associates elements of $\mathbb{R}^{ln}$ (net trades) with ordered pairs $(\underset{\sim}{\omega}, \underset{\sim}{R})$ consisting of endowment and preference profiles. The set of these elements is denoted by $f(\underset{\sim}{\omega}, \underset{\sim}{R})$. It is assumed that $f(\underset{\sim}{\omega}, \underset{\sim}{R})$ is nonempty for all $(\underset{\sim}{\omega}, \underset{\sim}{R})$ in its domain.

For the sake of simplicity, we shall assume in what follows that this correspondence is single-valued, i.e., a function. Subsequently, we shall indicate the modifications required to extend the results to the general case of correspondences.

We shall consider two games. The *main game*, in which both the endowments and preferences are unknown, and withholding (but not destruction) is permitted, is called the W-R game. In such a game, for any $i \in N$, a generic element of the ith strategy space S_i is denoted by s_i, with

$$s_i = (w_i, d_i).$$

$w_i \in \mathbb{R}^{ln}_{+0}$ as before,[32] $w_i = (w_i^1, \ldots, w_i^n)$, $w_i^j \in \mathbb{R}^l_{+0}$. $d_i \in D_i$ where D_i is an arbitrary set (the ith domain). The outcome function of this game is h: $S_1 \times \ldots \times S_n \to \mathbb{R}^{ln}$.

We shall also consider an *auxiliary game*, designed for situations where the endowment is given (though perhaps incorrectly) while preferences are unknown. Let the given endowment profile be $v = (v^1, \ldots, v^n)$, $v^i \in \mathbb{R}^l_{+0}$, $i \in N$. We denote by $A(v)$ the set of feasible net allocations in a pure exchange economy when v is the initial endowment profile and each consumption set is the nonegative orthant; i.e., $A(v) = \{(z^1, \ldots, z^n) \in \mathbb{R}^{ln}: z^i \in \mathbb{R}^l, \Sigma_{i \in N} z^i = 0, z^i \geqq -v^i, i \in N\}$.

We denote by g^v an outcome function, g^v: $D_1 \times \ldots \times D_n \to \mathbb{R}^{ln}$, for an auxiliary game when the set of feasible allocations is $A(v)$ and the strategic domains are D_i, $i \in N$. The mapping associating the outcome function g^v with the profile v is called the *auxiliary game form g*.

The set of Nash equilibria of this game (a subset of $D_1 \times \ldots \times D_n$) for the preference profile $\underset{\sim}{R}$ is denoted by $\upsilon_{g^v}(\underset{\sim}{R})$, and the corresponding set of Nash allocations (a subset of $\mathbb{R}^{ln}_{+0}$) by $N_{g^v}(\underset{\sim}{R})$.

Definition 3. f is R-implementable through the auxiliary game form g if, for every $v \in \mathbb{R}^{ln}_{+0}$, there exist domains $D_1, \ldots, D_n$ and an auxiliary outcome function g^v: $D_1 \times \ldots \times D_n \to \mathbb{R}^{ln}$, such that

$$N_{g^v}(\underset{\sim}{R}) = f(v, \underset{\sim}{R}) \quad \text{for all} \quad (v, \underset{\sim}{R}).$$

(That is, every Nash allocation generated by the auxiliary game is f-optimal for v and $\underset{\sim}{R}$, and every f-optimal allocation for v and $\underset{\sim}{R}$ is attainable as a Nash allocation of the auxiliary game.)

Definition 4. For each $i \in N$, let the ith person's strategy set be of the form

$$S_i = S_i(\mathring{\omega}_i) \subset \mathbb{R}^l_{+0} \times T_i,$$

where T_i is an arbitrary set. A generic element of S_i is denoted by $s_i = (w_i^i, t_i)$.[33] Write $S = S(\underset{\sim}{\mathring{\omega}}) = S_1 \times \ldots \times S_n$, and[34]

$$A(w_1^1, \ldots, w_n^n) = \{(z^1, \ldots, z^n) \in \mathbb{R}^{ln}: z^i \in \mathbb{R}^l, \sum_{i \in N} z^i = 0,$$

$$z^i \geqq -w_i^i, \forall i \in N\}.$$

An outcome function h: $S \to \mathbb{R}^{ln}$ is said to be $\underset{\sim}{\mathring{\omega}}$-feasible if

$$h(s) \in A(\mathring{\omega}_1, \ldots, \mathring{\omega}_n) \quad \text{for all} \quad s \in S,$$

where $s = (w_i^i, t_i)_{i \in N}$.

Definition 5. A SCR (performance correspondence) f is W-R-implementable (in NE) if, for every $\underset{\to}{\mathring{\omega}} \in \mathbb{R}_{+0}^{ln}$, and for every $i \in N$, there exist strategic domains

$$S_i = S_i(\mathring{\omega}_i) \subset \mathbb{R}_{+0}^{l} \times T_i,$$

where T_i is an arbitrary set, and an $\underset{\to}{\mathring{\omega}}$-feasible outcome function

$$h: \prod_{i \in N} S_i \to \mathbb{R}^{ln},$$

such that:

$$\forall \underset{\to}{\mathring{R}} \in R,$$

there is an NE s for $\underset{\to}{\mathring{\omega}}$ and $\underset{\to}{\mathring{R}}$ (i.e., $s \in \upsilon_h(\underset{\to}{\mathring{\omega}}, \underset{\to}{\mathring{R}})$)

such that

$$h(s) \in f(\underset{\to}{\mathring{\omega}}, \underset{\to}{\mathring{R}}).$$

Remark 4. In our applications, $S_i(\mathring{\omega}_i) = (0, \mathring{\omega}_i) \times T_i$ and $T_i = \mathbb{R}_{+0}^{l(n-1)} \times D_i$ where D_i is an arbitrary set.

Theorem 2.A: Let f be an IR social choice rule (performance function) which is R-implementable (in NE) through an auxiliary form $g: v \to g^v$. Then f is W-R-implementable in NE (by a "combination" of g with the endowment game of Sec. II.A.1).

Proof: We will construct an outcome function h as follows.

Let $\mathring{\omega}_i$ and the corresponding strategy spaces $S_i(\mathring{\omega}_i)$, $i \in N$, be given. By construction, $s_i = (w_i, d_i)$, $w_i = (w_i^1, \ldots, w_i^i, \ldots, w_i^n)$, and $w_i^i \leqslant \mathring{\omega}_i$.

Now we distinguish two types of situations according as to whether there exists $\underset{\to}{v} \in \mathbb{R}_{+0}^{ln}$ such that $\underset{\to}{v} = w_i$ for all $i \in N$.

If such $\underset{\to}{v}$ does not exist, we follow rules (b) and (c) above and conclude that s is not an NE (see Claim 3′ below).

On the other hand, suppose that $\underset{\to}{v}$ does exist. Then the outcome is dictated by the outcome function $g^{\underset{\to}{v}}$ generated through the mapping g for this $\underset{\to}{v}$. It then turns out (see Claims 1′ and 2′ below) that an NE obtains only if $\underset{\to}{v}$ coincides with the true endowment profile $\underset{\to}{\mathring{\omega}}$. But then, by the assumption on g, it follows that $N_h(\underset{\to}{\mathring{\omega}}, \underset{\to}{\mathring{R}}) = f(\underset{\to}{\mathring{\omega}}, \underset{\to}{\mathring{R}})$.

Formally, the rule (a) of the endowment game (W-game) described in the previous section is replaced by the following Rule (a′): if for some $\underset{\rightarrow}{\nu}$ such that, for all $i \in N$,

$$s_i = (\underset{\rightarrow}{\nu}, d_i)$$

for some $(d_1, \ldots, d_n) \equiv \underset{\rightarrow}{d}$, then, for $s = (s_1, \ldots, s_n)$, we set

$$h(s) = g^{\underset{\rightarrow}{\nu}}(\underset{\rightarrow}{d}).$$

The rules governing cases where there is no unanimity as to endowments are unchanged. The right hand sides of the definitions of $M(s)$ and $w(s)$ remain the same as in the W-game, although now $s_i = (w_i, d_i)$ rather than $s_i = w_i$. The two other rules ((b′) and (c′)) are the same as rules (b) and (c) for the W-game, again with $s_i = (w_i, d_i)$.

Theorem 2.B.B:[35] Let $n \geqq 3$, let endowments be semi-positive ($\omega_i \geq 0$), and preferences continuous and strictly increasing. Then, a social choice function f is W-R-implementable in NE if and only if it is monotone and individually rational (IR).

Proof: (i) Sufficiency. For $n \geqq 3$ and monotone f, Theorem 5 in Maskin (1977)[36] shows that there exists a function g which R-implements f in NE.[37] Hence, by Theorem 2.A above, the individually rational social choice function f is W-R-implementable.

(ii) Necessity. If f is R-implementable, it is monotone by Theorem 2 of Maskin (1977). If f is W-implementable, it is IR by Proposition 3 in Section II.A.1.

Claim 1′: *Correct unanimity with regard to endowments yields an NE.*

Let $s^* = (s_1^*, \ldots, s_n^*)$, and, for all $i \in n$, $s_i^* = (\underset{\rightarrow}{\mathring{\omega}}, d_i^*)$, such that $d^* = (d_1^*, \ldots, d_n^*)$ is an NE for $g^{\mathring{\omega}}$ given $\underset{\rightarrow}{\mathring{R}}$, i.e., $d^* \in \nu_{g\mathring{\omega}}(\underset{\rightarrow}{\mathring{R}})$. Then s^* is an NE for h given $(\underset{\rightarrow}{\mathring{\omega}}, \underset{\rightarrow}{\mathring{R}})$; i.e., $s^* \in \nu_h(\underset{\rightarrow}{\mathring{\omega}}, \underset{\rightarrow}{\mathring{R}})$.

Proof: Suppose s^* is not an NE. By the assumption concerning d^*, for any agent i, it would not help to depart from d_i^* while retaining $w_i = \underset{\rightarrow}{\mathring{\omega}}$.

Consider therefore $\tilde{s} = (\tilde{s}_1, \ldots, \tilde{s}_n)$, such that $\tilde{s}_j = s_j^*$ for all $j \neq i$, while $\tilde{s}_i = (\tilde{w}_i, \tilde{d}_i)$ with $\tilde{w}_i \neq \underset{\rightarrow}{\mathring{\omega}}$. ($\tilde{d}_i$ may or may not equal d_i^*.) Since, by the outcome rules, $\tilde{w}_i^i \leqq \mathring{\omega}_i$, it follows that $M(\tilde{s}) = \emptyset$ and so rule (b′) applies.

But

$$\beta_i(\tilde{s}) = 0,$$

since other agents remain unanimous with regard to endowments. Hence, rule (b′) prescribes

$$h_i(\tilde{s}) = -\tilde{w}_i^i.$$

By our assumptions on the auxiliary game form g and d^*,

$$h(s^*) = f(\underset{\to}{\mathring{\omega}}, \underset{\to}{\mathring{R}}).$$

Since f is assumed to be IR,

$$f_i(\underset{\to}{\mathring{\omega}}, \underset{\to}{\mathring{R}})\mathring{R}_i 0,$$

hence

$$h_i(s^*)\mathring{R}_i 0,$$

and therefore

$$-\tilde{w}_i^i \mathring{P}_i 0,$$

which contradicts the requirement of semi-positivity for endowment messages and strictly increasing preferences. Hence s^* is an NE for $(\underset{\to}{\mathring{\omega}}, \underset{\to}{\mathring{R}})$.

Claim 2′: Incorrect unanimity concerning endowments does not yield an NE.

Let $s = (s_l, \ldots, s_n)$, $s_i = (\underset{\to}{v}, d_i)\ \forall i \in N$, $\underset{\to}{v} = (v^l, \ldots, v^n)$, $v^i \in \mathbb{R}^{ln}_{+0}$, $\underset{\to}{v} \neq \underset{\to}{\mathring{\omega}}$. Then s is not an NE for $(\underset{\to}{\mathring{\omega}}, \underset{\to}{\mathring{R}})$.

Proof: Suppose that s is an NE for $(\underset{\to}{\mathring{\omega}}, \underset{\to}{\mathring{R}})$.

By the outcome rules, $\underset{\to}{v} \leqq \underset{\to}{\mathring{\omega}}$, and (since $\underset{\to}{v} \neq \underset{\to}{\mathring{\omega}}$ by hypothesis), $v^i \leqslant \mathring{\omega}_i$ for some i by virtue of the non-exaggeration requirement.

Then, by reasoning exactly like that in the proof of Claim 2, we show that Claim 2′ will have been established if [with $\tilde{s}_i = (\tilde{w}_i, d_i)$, $\tilde{w}_i = (\tilde{w}_i^i, w_{)i(})$, $\tilde{w}_i^i = \mathring{\omega}_i$, and $\tilde{s}_j = s_j\ \forall\ j \neq i$]

(+) $$h_i(\tilde{s}) \geqslant h_i(s),$$

and that if (+) fails then

(0) $$h_i(s) = \sum_{j \neq i} v^j$$

and

(*) $$h_j(s) = -v^j\ \forall\ j \neq i.$$

It will therefore suffice to show that the last two equalities yield a contradiction.

To get this contradiction, we shall first prove the following:

Auxiliary Proposition: If s is an NE for $(\underset{\sim}{\mathring{\omega}}, \underset{\sim}{\mathring{R}})$, then s is also an NE for $(\underset{\sim}{v}, \underset{\sim}{\mathring{R}})$.

Proof: (1) Consider agent i. We know that our rules never give to an agent more than the others have "put on the table." That is, for all s_i',

$$h_i(s_i', s_{)i(}) \leqslant \sum_{j \neq i} w_j^j = \sum_{j \neq i} v^j.$$

But, by (0) above,

$$h_i(s) = \sum_{j \neq i} v^j.$$

Hence

$$h_i(s_i', s_{)i(}) \leqslant h_i(s) \quad \text{for all } s_i',$$

and so, by the monotonicity of preferences, s_i is a Nash equilibrium strategy for agent i.

(2) Now consider any agent j other than i. Suppose s_j is not a Nash equilibrium strategy for j in the economy $(\underset{\sim}{v}, \underset{\sim}{\mathring{R}})$.

Then there must exist a strategy s_j' for j with the characteristic $(v^j, \mathring{R}_j)$ such that

$$(\alpha) \qquad h_j(s_j', s_{)j(}) \; \mathring{P}_j h_j(s).$$

Now, since by the rules of the game $h_j(s^*) \geqq -w_j^j$ always, we have in particular

$$(\beta) \qquad h_j(s_j', s_{)j(}) \geqq -v^j = h_j(s)$$

where the last equality follows from (*) above.

Since replacing $\geqq$ by $=$ in (β) would contradict (α), if follows that $\geqq$ in (β) can be replaced by $\geq$, and (β) becomes

$$h_j(s_j', s_{)j(}) \geq h_j(s).$$

In view of the assumed strict monotonicity of preferences, the latter inequality implies

$$h_j(s_j', s_{)j(}) \; \mathring{P}_j h_j(s)$$

where j's characteristics are $(\underset{\sim}{\mathring{\omega}}_j, \underset{\sim}{\mathring{R}}_j)$, and so s is not an NE in the economy $(\underset{\sim}{\mathring{\omega}}, \underset{\sim}{\mathring{R}})$. This contradiction of our initial hypothesis completes the proof of the Auxiliary Proposition.

We now return to the proof of Claim 2′. By Rule (a′), since s is unanimous as to endowments, we have

$$h(s) = g^{\underset{\sim}{\nu}}(\underset{\sim}{d}),$$

and

$$(\gamma) \qquad h_j(s) = g_j^{\underset{\sim}{\nu}}(\underset{\sim}{d}) \ \forall j \in N.$$

Now, by the Auxiliary Proposition, $\underset{\sim}{d}$ constitutes an NE in the game $g^{\underset{\sim}{\nu}}$ for $\mathring{\underset{\sim}{R}}$, and, by hypothesis, g^{ν} R-implements f. Therefore

$$g^{\underset{\sim}{\nu}}(\underset{\sim}{d}) = f(\underset{\sim}{\nu}, \mathring{\underset{\sim}{R}}),$$

and so

$$(\delta) \qquad g_j^{\underset{\sim}{\nu}}(\underset{\sim}{d}) = f_j(\underset{\sim}{\nu}, \mathring{\underset{\sim}{R}}).$$

Using in turn (δ), (γ), and (*), we obtain

$$f_j(\underset{\sim}{\nu}, \mathring{\underset{\sim}{R}}) = g_j^{\underset{\sim}{\nu}}(\underset{\sim}{d}) = h_j(s) = -\nu^j, \quad \forall j \neq i;$$

hence,

$$f_j(\underset{\sim}{\nu}, \mathring{\underset{\sim}{R}}) = -\nu^j, \quad \forall j \neq i.$$

But this contradicts the hypothesis that f is NC, i.e., that

$$f_j(\underset{\sim}{\nu}, \mathring{\underset{\sim}{R}}) \geqslant -\nu^j, \quad \forall j \in N.$$

This contradiction implies that (+) holds, and hence that, by the strict monotonicity of preferences,

$$h_i(\tilde{s})\mathring{P}_i h_i(s).$$

So s is not an NE for $(\mathring{\underset{\sim}{\omega}}, \mathring{\underset{\sim}{R}})$. This completes the proof of Claim 2′.

Claim 3′: If there is no unanimity as to endowments, then there is no NE.

Proof: We proceed as in the proof of Claim 3 except for (iii), which is replaced by the following:

(iii)′ Finally, suppose that s is not unanimous as to endowments and $M(s) = \emptyset$. Since, by the hypothesis of Claim 3′, not all announcements in s are the same, there exist agents i and j, $i \neq j$, with

$$w_i^i \neq w_j^i,$$

We now distinguish two subcases according to whether $\beta_j(s) = 0$ or $\beta_j(s) > 0$.

Subcase (iii.1)': $\beta_j(s) = 0$.

Consider $\tilde{\tilde{s}}$ defined by

$$\tilde{\tilde{s}}_k = s_k \quad \text{for all} \quad k \neq j$$
$$\tilde{\tilde{w}}_j^j = \tfrac{1}{2} w_j^j$$
$$\tilde{\tilde{w}}_j^r = \tilde{\tilde{w}}_j^r \quad \text{for all} \quad r \neq j,$$

and the second component of $\tilde{\tilde{s}}_j$ arbitrary (e.g., $\tilde{\tilde{d}}_j = d_j$).

We note that since s is not unanimous as to endowments and $M(s)$ is empty, rule (b)' applies to s, and hence

$$h_j(s) = -w_j^j.$$

But, also, $\tilde{\tilde{s}}$ is not unanimous as to endowments because $\tilde{\tilde{w}}_i^i = w_i^i$, $\tilde{\tilde{w}}_j^i = w_j^i$ by construction,[38] and $w_i^i \neq w_j^i$ by hypothesis.

Also, $M(\tilde{\tilde{s}})$ is empty because $M(s)$ is empty, and the change from w_j^j to $\tilde{\tilde{w}}_j^j = \frac{1}{2} w_j^j$ (while $\tilde{\tilde{w}}_r^j = w_r^j$ for $r \neq j$) does not enlarge the set M. Hence rule (b)' also applies to $\tilde{\tilde{s}}$. Now, since $\tilde{\tilde{s}}_k = s_k$ for $k \neq j$, $\beta_j(s) = 0$ implies $\beta_j(\tilde{\tilde{s}}) = 0$. Therefore,

$$h_j(\tilde{\tilde{s}}) = -\tilde{\tilde{w}}_j^j.$$

But,

$$-\tilde{\tilde{w}}_j^j = -\tfrac{1}{2} w_j^j \geqslant -w_j^j,$$

because, by assumptions on messages, $w_j^j \geqslant 0$. Hence, by the assumption of strictly increasing preferences,

$$h_j(\tilde{\tilde{s}}) \mathring{P}_j h_j(s).$$

So, $\tilde{\tilde{s}}$ is better than s for agent j, and hence s is not a Nash equilibrium.

Q. E. D.

Subcase (iii.2)': $\beta_j(s) > 0$.

In this situation consider $\tilde{s}$, such that

$$\tilde{s}_k = s_k \quad \text{for all} \quad k \neq j$$

and

$$\tilde{s}_j^r = s_r^r \quad \text{for all} \quad r.$$

By construction, $\beta_j(\tilde{s}) = \beta_j(s) > 0$ and $\Sigma_{k \neq j} \beta_k(\tilde{s}) < \Sigma_{k \neq j} \beta_k(s)$. Also, $M(\tilde{s}) = M(s) = \varnothing$, so rule $(b)'$ applies to both $\tilde{s}$ and s. Therefore, $h_j(\tilde{s}) \geqslant h_j(s)$. And so, again by the assumption of strictly increasing preferences, s is not a Nash equilibrium.

Q.E.D.

II.B. DESTRUCTION OF ENDOWMENTS

In this section, we consider an alternative game, in which the agents may destroy a part of their endowment but are not able to withhold (conceal) any of it. D-implementability is defined analogously to W-implementability, with destruction replacing the withholding of endowments. We again assume pure exchange, with semi-positive initial endowments ($\mathring{\omega}_i \geqslant 0$) and strictly increasing preferences.

It then turns out that the outcome function introduced in Sec. II.A.1 above, with the modification indicated under Claim 3 below,[39] D-implements any non-confiscatory (NC)[40] performance function when preferences are known to the designer.[41] Similarly, when f is monotone as well as NC, outcome functions of the type considered in Sec. II.A.2 above implement f when neither endowments nor preferences are known to the designer.

In what follows we state the result for the case of known preferences and indicate the modifications in the proof for W-implementation needed to make it valid for D-implementation. The theorem on D-implementability when both endowments and preferences are unknown is the same as part (1) of the theorem on W-implementability, with NC replacing IR.

The notation for strategies remains the same as in Sec. II.A but the interpretation differs. In particular, given s, agent i's final (total) holdings $H^i(s)$ equal $w_i^i + h_i(s)$ where w_i^i denotes i's (true) endowment after destruction. Similarly, for $i \neq j$, w_i^j denotes i's estimate of j's endowment after destruction. It is still assumed that $w_i^k \geqslant 0$ (i.e., $w_i^k \in \mathbb{R}^l_{+0}$) for all i, k in N. Hence, an agent cannot destroy all of his endowment.

The result for the case of known preferences is given by the following:

Theorem 3: f is D-implementable (in NE) for $\underset{\sim}{\mathring{R}}$ if it is non-confiscatory (NC).

Proof. The proof is very much the same as that for W-implementability. In particular, in the former proof we used the fact (see Remark 3 in Section II.A.1 that IR implies NC, while here only NC is assumed. We shall therefore only spell out those parts of the proof of D-implementability which differ significantly from the proof of W-implementability, with page references to the former proof.[42]

First, for the destruction game, we replace rule (b) by the following rule (b*), consisting of two parts, (b_1^*) and (b_2^*).[43]

In order to state these rules we must define numbers t_i ($i = 1, \ldots, n$) as follows. Consider $s = (s_1, \ldots, s_n)$ where $s_i = (s_i^1, \ldots, s_i^n) = (w_i^1, \ldots,$

w_i^n), with w_i^j—as before—denoting the value of j's endowment claimed by i (called i's *estimate* of j's endowment). Denote by $t^i(s)$ the number of distinct commodity space points among the elements $w_1^i, \ldots, w_n^i$, to be called the *number of estimates* (*in* s) *of* i*'s endowment*, and define $t(s) = \max\{t^1(s), \ldots, t^n(s)\}$. We shall call $t(s)$ the *number of estimates in* s.

The rule (b^*) then reads as follows

If $M(s) = \varnothing$, and $t(s) = 2$, then (b_1^*)

$$h_i(s) = [\beta_i(s)/\Sigma_{j \in N}\beta_j(s)] \cdot w(s) - w_i^i, \quad i \in N. \qquad (\#)$$

If $M(s) = \varnothing$, and $t(s) > 2$, then (b_2^*)

$$h_i(s) = [\beta_i^*(s)/\Sigma_{j \in N}\beta_j^*(s)] \cdot w(s) - w_i^i, \; i \in N, \qquad (\#\#)$$

where

$$\beta_k^*(s) = 1 + \beta_k(s), \; k \in N.$$

The changes in the proof of the three claims, here labeled respectively with double primes, are indicated below.

Claim 1″: Here we must replace the part of the W-proof using the IR property of f by an argument using the NC property only. We therefore substitute for the last ten lines of the proof of Theorem 1[45,46] the following paragraph:

Since f is NC, and preferences are strictly increasing,

$$\mathring{\omega}_j + f_j(\underset{\sim}{\mathring{\omega}})\mathring{P}_j 0.$$

But here

$$\mathring{\omega}_j + h_j(s) = \mathring{\omega}_j + f_j(\underset{\sim}{\mathring{\omega}})$$

and

$$\tilde{w}_j^j + h_j(\tilde{s}_j, s_{)j(}) = \tilde{w}_j^j - \tilde{w}_j^j = 0.$$

Hence,

$$(\mathring{\omega}_j + h_j(s))\mathring{P}_j(\tilde{w}_j^j + h_j(\tilde{s}_j, s_{)j(}))$$

which contradicts our supposition (+++) in the proof of claim 1 and in the proof of Theorem 1.

Remark. This argument would not be valid for withholding where, under $\tilde{s}$, the total final holdings equal $\mathring{\omega}_j - \tilde{w}_j^j$ rather than 0.

Claim 2″: Replace the sentence after (+) in the proof of Claim 2 in the proof of Theorem 1 with:

Since preferences are strictly increasing and $\tilde{w}_i^i \geqslant w_i^i$, the inequality (+) implies

$$(\tilde{w}_i^i + h_i(\tilde{s}))\mathring{P}_i(w_i^i + h_i(s)).$$

Claim 3″: In the absence of unanimity there is no NE.

Proof: We consider three cases:

(i)″ $M(s) = N$; (ii)″ $M(s) \neq \varnothing$, $M(s) \neq N$; (iii)″ $M(s) = \varnothing$.

(i)″ Suppose first that $M(s) = N$. Then consider $\tilde{s}$ with

$$\tilde{s}_k = s_k \quad \text{for all} \quad k \neq 1,$$
$$\tilde{s}_1^q = s_q^q \quad \text{for all} \quad q \in N.$$

(That is, agent one accepts everyone's self-evaluation.)
Then

$$M(\tilde{s}) = \{1\}.$$

(This is proved exactly as in Theorem 1, Claim 3(*i*).)
Since $M(\tilde{s}) \neq \varnothing$, rule (c) applies. Therefore,

$$h_1(\tilde{s}) = 1 \cdot w(\tilde{s}) - \tilde{w}_1^1 = w(s) - w_1^1 = \sum_{i=1}^{n} w_i^i - w_1^1.$$

On the other hand, since $M(s) = N$, rule (c) also applies to s and yields

$$h_1(s) = \frac{1}{n}\sum_{i=1}^{n} w_i^i - w_1^1.$$

Since $\sum_{i=1}^{n} w_i^i \geqslant 0$ (by the rule $w_i \geqslant 0$), and $n > 1$, it follows that

$$h_1(\tilde{s}) \geqslant h_1(s).$$

Hence, since $\tilde{w}_1^1 = w_1^1$,

$$H_1(\tilde{s}) \geqslant H_1(s),$$

and, by strictly increasing preferences, $H_1(\tilde{s})\mathring{P}_1 H_1(s)$. So s is not an NE in case (i)″.

(ii)″ $M(s) \neq \varnothing$, $M(s) \neq N$.

Since $M(s) \neq \varnothing$ and there is no unanimity, rule (c) applies to s. Because $M(s) \neq N$, there is an agent $j \notin M(s)$ who, by rule (c), gets

$$h_j(s) = -w_j^j.$$

(Since this is the case of destruction, $H_j(s) = w_j^j + h_j(s) = w_j^j - w_j^j = 0$.)

Now suppose that agent j accepts everyone's self-evaluation. Thus

$$\tilde{s}_r = s_r \quad \text{for all} \quad r \neq j$$

and

$$\tilde{s}_j^q = s_q^q \quad \text{for all} \quad q.$$

Then (by the argument in Theorem 1)

$$M(\tilde{s}) = \varnothing.$$

Hence rule (c) does not apply. But neither does rule (a) because $\tilde{s}$ is not unanimous. (This is seen as follows: since $n \geqq 3$ and $M(s) \neq \varnothing$, there is a person $k \in M(s)$, $k \neq j$, and a person i, with $i \neq j$, $i \neq k$, such that

$$w_k^k \geq w_i^k;$$ [47]

hence $s_k \neq s_i$. But since $k \neq j$ and $i \neq j$, we have $\tilde{s}_k = s_k$ and $\tilde{s}_i = s_i$ by construction. Hence $\tilde{s}_k \neq \tilde{s}_i$, and so $\tilde{s}$ is not unanimous.)

Since $\tilde{s}$ is not unanimous and $M(\tilde{s}) \neq \varnothing$, rule (b*) applies to $\tilde{s}$.

For agents j, k, and i referred to above, we have

$$\beta_j(\tilde{s}) \geqq \| w_k^k - w_i^k \| > 0,$$

since $w_k^k \geq w_i^k$.

From $w(\tilde{s}) = w(s) \geq 0$, it follows that $h_j(\tilde{s}) = \dfrac{\beta_j(\tilde{s})}{\Sigma \beta_r(\tilde{s})} w(\tilde{s}) \geq 0$. On the other hand, $\beta_q^*(\tilde{s}) > 0$ by construction for all $q \in N$ and all $\tilde{s} \in s$, so that $\dfrac{\beta_j^*(\tilde{s})}{\Sigma \beta_r^*(\tilde{s})} w(\tilde{s}) \geq 0$. Hence, whether rule ($b_1^*$) or rule ($b_2^*$) applies, we have

$$h_j(\tilde{s}) \geq -w_j^j = h_j(s).$$

(The last equality was exhibited above.)

But $\tilde{w}_j^j = w_j^j$, so $H_j(\tilde{s}) \geq H_j(s)$, and, by strictly increasing preferences, $H_j(\tilde{s}) \mathring{P}_j H_j(s)$. Therefore, s is not an NE.

(iii)″ Finally, suppose there is no unanimity in s; hence the number of estimates $t(s)$ is at least 2, and $M(s) = \varnothing$. We distinguish two cases: case 1: The number $t(s)$ of estimates is 2; case 2: the number of estimates is at least three.

Consider first case 1 where the number of estimates is two, i.e., $t(s) = 2$. In this case, we distinguish two subcases, 1a, where all $\beta_k(s) > 0$, $k \in N$, and 1b, where not all $\beta_k(s)$ are positive (i.e., some are zero).

Subcase 1a: Here $t(s) = 2$, and $\beta_k(s) > 0$ for all k in N. Since there is no unanimity, there are agents i and j such that, in s, $w_i^i \neq w_j^i$. Let i change his strategy from s_i to $\bar{s}_i$, so that, in $\bar{s}_i$, $\bar{w}_i^i = w_j^i$, while other components of $\bar{s}_i$ are the same as in s_i. Then $\beta_i(\bar{s}) = \beta_i(s) > 0$, where the equality follows from the definition of $\beta_i(.)$ and the inequality holds by the hypothesis of case A. Also, $\beta_j(\bar{s}) = \beta_j(s)$. But, since our theorem assumes n > 2, there is at least one agent r other than i or j, and for all such agents $\beta_r(\bar{s}) < \beta_r(s)$. Clearly, $t(\bar{s}) = t(s) = 2$, so rule (b_1^*) applies. It follows from the above properties of the β's that $h_i(\bar{s}) > h_i(s)$, and hence s is not an NE.

Subcase 1b: Here, still, $t(s) = 2$, but there exists some agent i such that $\beta_i(s) = 0$. Here the argument depends on whether i has a strategy $\bar{s}_i$ such that $t(\bar{s}) > 2$, with $\bar{s}$ non-unanimous and leaving the set $M(\bar{s})$ empty.

Consider first the *sub-subcase 1b'* where such a strategy $\bar{s}$ is availailable to agent i. The situation with $\bar{s}$ qualifies then under rule (b_2^*). Now since $\beta_i(s) = 0$, it follows from (#) that $H_i(s) = 0$. On the other hand, since $\beta_i^*(\bar{s}) > 0$ by construction, it follows from (##) that $H_i(\bar{s}) \geqslant 0$. Again, s is not an NE.

But suppose (*sub-subcase 1b*, that i has no strategy $\bar{s}_i$ qualifying under rule (b_2^*). This can only happen if, under s, all agents other than i ("the crowd") are announcing identifical profiles but different from that announced by i (the only "dissident").[48]

Here again there are two possibilities:

(i) The dissident and the crowd agree about i's endowment; i.e., $w_i^i = w_j^i$ for all $j \neq i$. Then i can adopt the strategy $\bar{s}_i$ with $\bar{w}_i^i = w_i^i$ and $\bar{w}_i^j = w_j^j$ for all $j \neq i$. With others retaining their strategies from s, this will result in a unanimous $\bar{s}$, so that $h_j(\bar{s}) = f(\bar{s})$. Since f is, by assumption in the Theorem, NC (non-confiscatory), it follows that $\bar{w} + h_i(\bar{s}) \geqslant 0$. On the other hand, since $t(s) = 2$, so that (b_1^*) applies, and $\beta_i(s) = 0$, formula (#) yields $w_i^i + h_i(s) = w_i^i + (-w_i^i) = 0$. Hence $\bar{s}$ yields to agent i a bigger outcome, i.e., $\bar{w}_i^i + h_i(\bar{s}) \geqslant 0 = w_i^i + h_i(s)$, so—by the assumed monotonicity of preferences—s is not an NE.

(ii) The dissident and the crowd disagree about i's endowment; i.e., $w_i^i \neq w_j^i$ for all $j \neq i$. For any j in the crowd, $\beta_j(s) > 0$, $j \neq i$. Then any member of the crowd r (with $r \neq i$) can change from s_r to $\bar{s}_r$ such that

$\bar{w}_r^i = w_i^i$, while other components of s_i remain unchanged. This does not change the number of disagreements, so $t(\bar{s}) = 2$, continues to hold, $\bar{s}$ is not unanimous, and $M(\bar{s})$ is still empty. Hence formula (#) in (b_1^*) applies. Now $\beta_i(\bar{s}) = \beta_i(s) = 0$ and $\beta_r(\bar{s}) = \beta_r(s) > 0$.[49] But for any agent k other than i or r (i.e., any member of the crowd other than r) $\beta_k(\bar{s}) < \beta_k(s)$. Thus for agent r, in the expression for $h_r(\bar{s})$ in (#) the numerator is positive and the same as in $h_r(s)$ while the denominator is smaller; also, $w(\bar{s}) = w(s)$. Hence $h_r(\bar{s}) > h_r(s)$ and so s is not an NE.

We now proceed to case 2, with $t(s) > 2$, i.e., where the number of estimates in s is three or more. Hence formula (##) in rule (b_2^*) defines the outcomes under s.

Since $t(s) = \max\{t^1(s), \ldots, t^n(s)\} > 2$, there exist three agents i, j, and k such that among the three estimates w_i^i, w_j^i, and w_k^i no two are equal. Let now agent j change the endowment estimate profile from s_j to $\bar{s}_j$ so that $\bar{w}_j^p = w_j^p$ for all $p \neq i$, and $\bar{w}_j^i$ such that $\bar{w}_j^i$ is closer (in norm) to w_i^i than w_j^i was, while still $\bar{w}_j^i \neq w_i^i$, and $\bar{w}_j^i \neq w_k^i$. Hence formula (##) in rule (b_2^*) applies to $\bar{s}$ as well as to s. (All components of $\bar{s} = (\bar{s}_1, \ldots, \bar{s}_n)$, except $\bar{s}_j$, are the same as those of s.)

Note that, since the components of $\bar{s}$ other than $\bar{s}_j$ are unchanged, we have $\beta_j(\bar{s}) = \beta_j(s)$. Also, $\beta_i(\bar{s}) = \beta_i(s)$. However, for r other than i or j, it is the case that $\beta_r(\bar{s}) < \beta_r(s)$. The same relations hold respectively for the β^*'s. Hence in the quotient of formula (##) for $h_j(\bar{s})$, the numerator is the same as for $h_j(s)$ and positive, while the denominator is smaller. It follows that $h_j(\bar{s}) > h_j(s)$, and therefore s is not an NE. This completes the proof of Theorem 3.

Remark 6. If rule (b*) had not been substituted for rule (b), Claim 3″ section (iii)″, would no longer be true (when $M(s) = \emptyset$). This is shown by the following counterexample:

$$n = 3;\ l = 1;\ s = \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix} = \begin{pmatrix} w_1^1 & w_1^2 & w_1^3 \\ w_2^1 & w_2^2 & w_2^3 \\ w_3^1 & w_3^2 & w_3^3 \end{pmatrix} = \begin{pmatrix} 1 & 5 & 4 \\ 2 & 3 & 4 \\ 1 & 3 & 4 \end{pmatrix}.$$

Assume that $\mathring{\omega}_1 = 1$, $\mathring{\omega}_2 = 3$, $\mathring{\omega}_3 = 4$. (So $w_i^i = \mathring{\omega}_i$, for $i = 1, 2, 3$.)

This s is not unanimous, and $M(s) = \emptyset$. If the mechanism were generally rules (a), (b), and (c), then rule (b) would apply here to s. Contrary to Claim 3″, this s is a Nash equilibrium.

Proof: (1) No $\tilde{s}$ can be unanimous (because if one player changes, the other two still disagree). So rule (a) will not apply to $\tilde{s}$.

(2) For every $\tilde{s}$, we have $M(\tilde{s}) = \varnothing$. This is so because, by hypothesis, every agent is already telling the truth about himself (i.e., he is destroying nothing), so he cannot raise his w_i^i; therefore $M(s) = \varnothing$ implies $M(\tilde{s}) = \varnothing$. So rule (c) will not apply to $\tilde{s}$.

(3) Hence rule (b) applies to any $\tilde{s}$ (as well as to s).

(4) We have $\beta_1(s) = \beta_2(s) = 0$ and $\beta_3(s) > 0$. By rule (b), agent 3 gets everything (i.e., $H_3(s) = w_1^1 + w_2^2 + w_3^3$), while the other two agents get nothing (i.e., $H_1(s) = H_2(s) = 0$). Certainly, therefore, agent 3 cannot do any better under any change of his strategy $\tilde{s}_3$.

As for agent 2, $H_2(\tilde{s}) = 0$ for any change of his strategy $\tilde{s}_2$, because $\tilde{s}_2$ does not enter $\beta_2(\cdot)$, so that $\beta_2(s_1, \tilde{s}_2, s_3) = 0$ for all $\tilde{s}_2$. Hence, agent 2 cannot do any better under any change of his strategy $\tilde{s}_2$.

Agent 1 is in exactly the same situation as agent 2.

So, no agent can do any better by unilateral strategy change, and hence s is a Nash equilibrium.

It is of some interest to see why and how the situation differs in the withholding game, in contrast to the destruction game being considered here.

In the withholding game, comments (1), (2), and (3) of the above proof remain valid. It also remains true that $\beta_1(s) = \beta_2(s) = 0$ and $\beta_3(s) > 0$. It is still true that agent 3 cannot improve his situation, but either of the other two agents can. Thus, in the W-game, let agent 2 choose $\tilde{w}_2^2 = \frac{1}{2}w_2^2$. (Recall that $w_2^2 = \mathring{\omega}_2$.) Then[50] $H_2^W(\tilde{s}) = \mathring{\omega}_2 - \tilde{w}_2^2 = \mathring{\omega}_2 - \frac{1}{2}\mathring{\omega}_2 = \frac{1}{2}\mathring{\omega}_2 \geqslant 0$, which is better than $H_2^W(s) = 0$. On the other hand, $H_2^D(\tilde{s}) = \tilde{w}_2^2 - \tilde{w}_2^2 = 0$, which is no improvement.

Remark 7. If rule (b) must be modified (as seen in Remark 1), it is natural to ask why it cannot be replaced by rule (b_2^*), rather than the more complex rule (b^*), which distinguishes between disagreement situations depending on whether there are more than two distinct strategy profiles. The answer is that rule (b_2^*) would be inappropriate in the proof of Claim 1″, while rule (b_1^*) does work.

Remark 8. We may note that we need not distinguish the cases $\beta_j(s) = 0$ from $\beta_j(s) > 0$ when rule (b_2^*) applies, since in both cases $\beta_j^*(s) > 0$, and the derived conclusion is due to changes in the denominator of $\beta_j^*(s) / \Sigma\ \beta_j^*(s)$, while the positive numerator remains constant. On the other hand, as in Theorems 1 and 2, we must distinguish these two cases when rule (b_1^*), which is identical with rule (b), does apply.

Appendix to section II.B

The following is a sketch of the proof of (iii)″ using the rules suggested by Hong (see footnote (43) above).

(iii)″ Finally, suppose that not all announcements in s are the same and

[1] $M(s) = \varnothing$ [i.e., the set $M(s)$ is empty].

We must show that s is not a Nash equilibrium.

Since there is no unanimity, there exist agents i and j, such that

[2] $$w_i^i \neq w_j^i,$$

where w_j^i denotes j's statement about i's endowment. We distinguish two subcases, with [1] and [2] assumed to hold in both:

A#. There are in s precisely two distinct endowment profiles v' and v''.

B#. There are in s at least three distinct endowment profiles in s.

Subcase A#. Any player k in the subset of N containing more than one member[51] can so change its announced profile as to produce three distinct endowment profiles. Let then $\tilde{s}_k = \tilde{v}$, $\tilde{v} \neq v'$, $\tilde{v} \neq v''$, and

$$\tilde{m}_k = 1 + \tilde{m}_k + (\Sigma_{r \neq k} m_r)_{\text{mod } n}.$$

Then in $\tilde{s}$ there are three distinct endowment profiles: $\tilde{v}$, v', and v''. Hence, Hong rule (H-b2) applies and k gets "all", which is better for k than what it would have obtained in s under rule (H-b1), since in s, some players other than k would also have received something under rule (H-b1). So s is not an NE.

Subcase B#. Since there are three distinct profiles in s, Hong rule (H-b2) applies. Let $i = 1 + \Sigma m_q)_{\text{mod}\, n}$. (Such i exists, because the RHS is an integer in $\{1, \ldots, n\}$ – and is unique.) Consider a player $j \neq i$, and let $\tilde{s}_j$ be such that $\tilde{w}_j = w_j$, while $\tilde{m}_j$ is such that

$$j = 1 + (\Sigma_{r \neq j} m_r) \bmod n.$$

That is, j does not change its endowment profile (hence, rule (H-b2) still applies) but changes its integer to become a winner. Since j was not a winner under s, its situation is improved. Hence s is not an NE.

This completes the proof of (iii)″.

II.C. WITHHOLDING AND DESTRUCTION

When both withholding and destruction are permitted, the former always dominates the latter. Hence, this case reduces to that treated in section II.A.

Example: The following example shows that non-confiscatoriness is not a necessary condition for D-implementability. It is assumed that endowments are unknown but preferences known. There are three persons. The performance and outcome functions are as follows:

Performance function:

$$f_1(\mathring{w}_1, \mathring{w}_2, \mathring{w}_3) = \Sigma_{i=1}^{3}\mathring{w}_i$$

$$f_2 = f_3(\mathring{w}_1, \mathring{w}_2, \mathring{w}_3) = 0$$

$$S_i = \{(w_1^1, w_1^2, w_1^3) \mid 0 \leqslant w_i^i \leqq \mathring{w}_i\} \quad i = 1,2,3.$$

Outcome function:

$$H: S_1 \times S_2 \times S_3 \rightarrow \text{final holdings}$$

a) If $w_1^j \nleq w_j^j$, $j = 2, 3$, then

$$H_1 = w_1^1 + \left[\frac{1}{\| w_2^2 - w_1^2 \| + 1}\right]w_2^2 + \left[\frac{1}{\| w_3^3 - w_1^3 \| + 1}\right]w_3^3$$

$$H_2 = \left[\frac{\| w_3^3 - w_1^3 \|}{\| w_3^3 - w_1^3 \| + 1}\right]w_3^3$$

$$H_3 = \left[\frac{\| w_2^2 - w_1^2 \|}{\| w_2^2 - w_1^2 \| + 1}\right]w_2^2.$$

b) If $w_2^2 \geqslant w_1^2$ or $w_3^3 \geqslant w_1^3$

$$H_i = w_i^i \quad i = 1, 2, 3.$$

Note that H is balanced.

Claim 1: (w_1^*, w_2^*, w_3^*) is an NE $\Rightarrow w_i^{j*} = w_j^{j*}$, $j = 2, 3$. This follows from the fact that H_1 is maximized for given w_2^2, w_3^3 when $w_1^j = w_j^{j*}$ $j = 2, 3$.

Claim 2: (w_1^*, w_2^*, w_3^*) is an NE $\Rightarrow w_j^{j*} = \mathring{w}_j$, $j = 2, 3$. By Claim 1, at an NE we must have $w_1^j = w_j^j$ $j = 2,3$. But this yields $H_2 = H_3 = 0$ by rule (a). If $w_j^{j*} \leqslant \mathring{w}_j$ $j = 2$ or 3, then $\tilde{w}_j^i = w_j^{i*}$, $i \neq j$, $\tilde{w}_j^j = \mathring{w}_j$ yields $H_j = \mathring{w}_j \geqslant 0$, contradicting w_j^* being part of an NE.

Claim 3: $w_i = (\mathring{w}_1, \mathring{w}_2, \mathring{w}_3)$ for all i is an NE. We are in rule (a) and person 1 clearly can do no better with any other strategy. If person 2 or person 3 changes his/her strategy, only decreasing his/her own stated endowment is possible. But this will result in rule (a) still being applicable, and this implies he/she continues to get 0.

Proposition 4: f is implemented by H.

Proof: This follows from the three claims.

APPENDIX TO SECTION II ON THE IMPLEMENTATION OF WALRASIAN AND CONSTRAINED WALRASIAN CORRESPONDENCES

1. The following example shows directly that, in certain pure exchange economies, the Walrasian correspondence cannot be implemented (in NE) without violating the feasibility requirements. The same example demonstrates that, in such economies, the Walrasian correspondence lacks monotonicity (in Maskin's sense); Maskin's theorem (1977) then implies the non-implementability. It should be noted that the conclusion about non-implementability holds whether the initial endowments are or are not known to the designer.

The example is given graphically in Figures 1.1, 2.2, and 3.3, presenting, respectively, the preferences, endowments, and budget lines of agents 1, 2, and 3 in a two-good economy.

We consider two environments $\tilde{e} = (\tilde{R}^1, \tilde{R}^2, \tilde{R}^3)$ and $\tilde{\tilde{e}} = (\tilde{\tilde{R}}^1, \tilde{\tilde{R}}^2, \tilde{\tilde{R}}^3)$, which differ only with respect to the preferences of agent 1. That is, $\tilde{R}^j = \tilde{\tilde{R}}^j$ for $j = 2, 3$.

It is seen that (p^*, z^*) is a Walrasian equilibrium for $\tilde{e}$, but not for $\tilde{\tilde{e}}$. (However, (p^*, z^*) is a constrained Walrasian equilibrium for $\tilde{\tilde{e}}$, as well as for $\tilde{e}$.) In fact, z^* is not a Walrasian allocation for $\tilde{\tilde{e}}$ (for any price).

Suppose now that, for some outcome function h, z^* is a Nash allocation given $\tilde{e}$. Then, for this h, z^* is also a Nash allocation given $\tilde{\tilde{e}}$. For, agents 2 and 3 have unchanged preferences, and any commodity bundle preferred by agent 1 to z^{*1}, according to the new preferences $\tilde{\tilde{R}}^1$, was also preferred according to $\tilde{R}^1$, and hence must have been unavailable by the rules of h. Hence, for every outcome function h, either z^* is not a Nash allocation given $\tilde{e}$ (when z^* is Walrasian) or z^* is a Nash allocation given $\tilde{\tilde{e}}$ (when z^* is not Walrasian). Hence, no outcome function h can yield a set of Nash allocations coinciding with that of the Walrasian alloca-

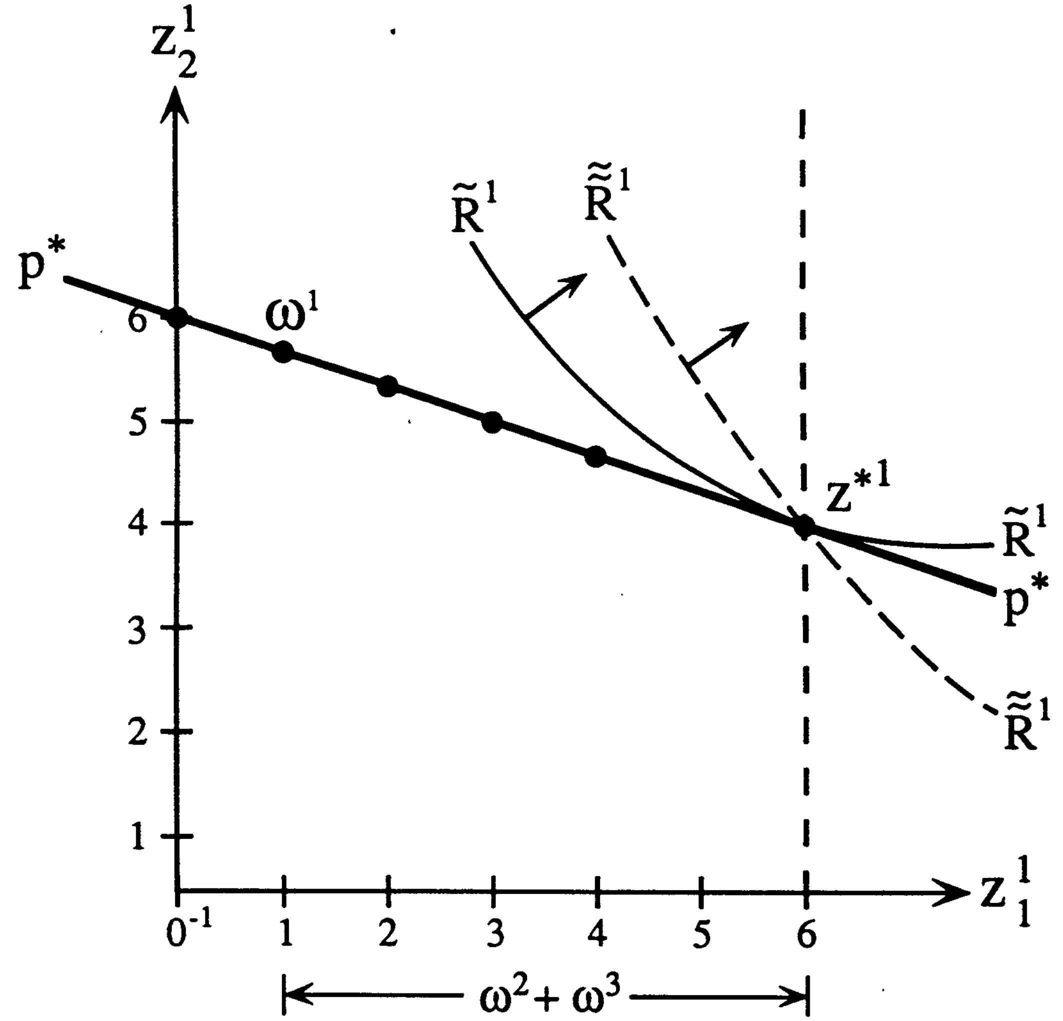

Figure 1.1.

tions; i.e., the Walrasian correspondence W is not implementable in Nash equilibria.

The fact that z^* is Walrasian for $\tilde{R}^1$, but not for $\tilde{\tilde{R}}^1$, shows that the Walrasian correspondence W is not monotone in Maskin's sense because, for all $i \in \{1, 2, 3\}$, $Z^{*i}\tilde{R}^iZ^i$ implies $Z^{*i}\tilde{\tilde{R}}^iZ^i$ within the feasible set.

2. It is convenient at this point to introduce the concept of a constrained Walrasian equilibrium.

We use the following notation:

$N = \{1, \ldots, n\}$ = the (finite) set of agents
$L = \{l, \ldots, l\}$ = the (finite) set of goods

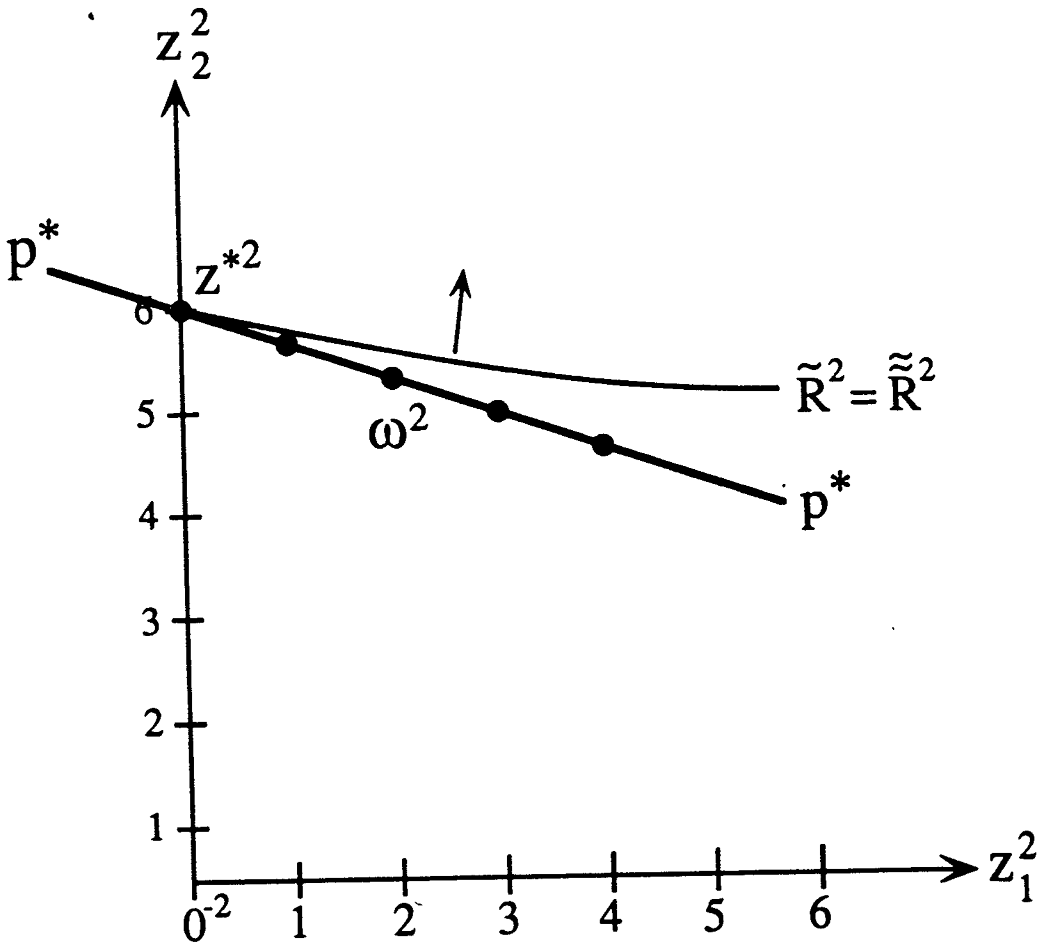

Figure 1.2.

P^l = the l-dimensional price simplex
ω_i = i's (true) initial endowment ($\omega_i \in \mathbb{R}^l_{+0}$, i.e., $\omega_i \geqslant 0$)

$\omega = \sum_{i \in N} \omega_i$

R_i = i's (true) weak preference relation
P_i = i's (true) strict preference relation.

Definition 6. An allocation $(\tilde{x}_i)_{i \in N}$, $\tilde{x}_i \in \mathbb{R}^l_+$, and a price p constitute a (pure exchange) *constrained Walrasian equilibrium* if

i) $\forall i \in N,\quad p \cdot \tilde{x}_i = p \cdot \omega_i;$

ii) $\forall i \in N,\quad \tilde{x}_i R_i x$ for all $x \leqslant \sum_{j \in N} w_j$ such that

$$p \cdot x \leqslant p \cdot \omega_i;$$

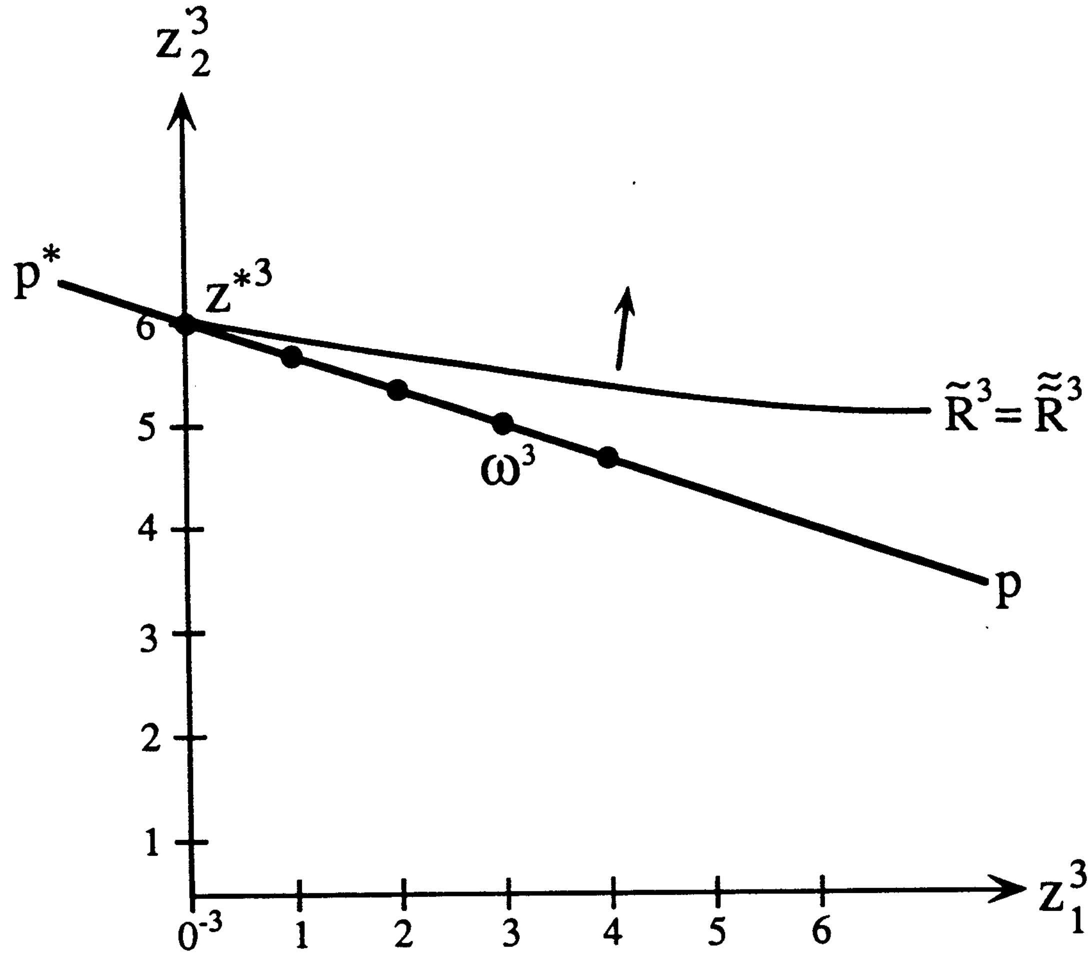

Figure 1.3.

iii) $\sum_{i \in N} \tilde{x}_i = \omega$

where $\tilde{x}_i \in \mathbb{R}^l_+$ denotes the total holdings (*not* the net trade) of the *i*th agent.

We shall denote the constrained Walrasian correspondence by W_c. Note that the Walrasian correspondence W is contained in W_c (i.e., $W \subseteq W_c$). In the absence of Edgeworth Box boundary Walrasian allocations, $W_c = W$.

3. A slight extension of Hurwicz's (1977) result implies that the constrained Walrasian correspondence is the smallest continuous social choice (= performance) correspondence satisfying the conditions of Pareto optimality and individual rationality that can be Nash-implemented over a sufficiently rich class of economies.

On the other hand, Maskin's (1977) result implies that, for three or more agents, any monotone performance correspondence can be Nash-implemented in pure exchange economies with strictly increasing preferences,[52] provided the feasible set is known.

Together these two results imply

Theorem 4: if a social choice correspondence f (defined over a sufficiently rich class of economies with $n \geqq 3$) is continuous, PO, IR, and monotone, then it contains the constrained Walrasian correspondence, i.e., $f \supseteq W_c$.

4. Our results in section 2 imply

Theorem 5: the constrained Walrasian correspondence W_c can be W-R-implemented (in NE).[53]

However, the outcome function in the proof of implementability uses a huge, in fact, infinite, dimensional strategic domain (message space). It is therefore of interest to see that

Theorem 6: *W_c can also be W-R implemented by a finite-dimensional strategic domain, similar to that used by Schmeidler (1976).*

It is sufficient here to demonstrate the R-implementability (when the initial endowments are known), since the outcome function in section II uses a finite-dimensional space for the revelation of endowments.

The strategic domain for agent $i \in N$ is given by

$$S_i = \{(p, x) \in \mathbb{R}^l_+ \times \mathbb{R}^l_+ : p \cdot x = p \cdot \omega_i\}.$$

x_i denotes the total holdings (*not* net trade) of the ith agent.

The outcome function H, for total holdings, is defined as follows:

1. If there exist $i, j, k \in N$, such that p_i, p_j, p_k are distinct, then[54]

$$H_t = \left[\frac{\| x_t \|}{\Sigma_{r\in N} \| x_r \|}\right]\omega, \quad \forall t \in N.$$

2. If there exist only two distinct announced prices p' and p'', and at least two agents announce each p' and p'', then

$$H_i = \omega_i, \quad \forall\, i \in N;$$

i.e., there are no trades.

3. If there is a $\bar{p}$, such that $p_i = \bar{p}$ for all $i \in N$ (unanimity as to price),

3.1) and $\sum_{i\in N} x_i \neq \omega$, then $H_i = \omega_i, \forall\, i \in N$;

3.2) and $\sum_{i\in N} x_i = x$, then $H_i = x_i, \forall\, i \in N$.

4. If there is a $\bar{p}$ and an agent $m \in N$, such that $p_m \neq \bar{p}$ but $p_j = \bar{p}$ for all $j \neq m$, then

$$4.1)\quad \begin{bmatrix} H_m = \dfrac{\bar{p}\cdot\omega_m}{\bar{p}\cdot x_m}x_m & \\ H_j = \dfrac{1}{n-1}(\omega - H_m) & \text{for} \quad j \neq m \end{bmatrix} \quad \text{if} \quad \frac{\bar{p}\cdot\omega_m}{\bar{p}\cdot x_m}x_m \leqslant \omega;$$

$$4.2)\quad H_i = \omega_i \quad \text{for all} \quad i \in N, \quad \text{if} \quad \frac{\bar{p}\cdot\omega_m}{\bar{p}\cdot x_m}x_m \nleqslant \omega.$$

Theorem 7: Let all initial endowments be semi-positive ($\mathring{\omega}_i \geqslant 0$) and known to the designer. Then the set of Nash equilibrium allocations (NA) for H coincides with the set of constrained Walrasian equilibrium allocations (CWA).

Proof: 1. CWA $\subseteq$ NA.

Let $[p, (\bar{y}_t)_{t\in N}] \in$ CWA.

Consider the strategies $(s_t)_{t\in N}$ where $s_t = (\bar{p}, \bar{y}_t)$ $\forall t$. By the definition of H, if any agent t unilaterally changes his strategy, he will receive a bundle y for which $p\cdot y = p\cdot\omega_t$ and $y \leqslant \omega$. By definition of a constrained Walrasian equilibrium, no such bundle is preferred to $\bar{y}_t$. Hence, $(\bar{y}_t)_{t\in N} \in$ NA.

2. NA $\subseteq$ CWA

Let $(s_t)_{t\in N}$, $s_t = (p_t, x_t)$, be Nash equilibrium strategies yielding the allocation $(\bar{y}_t)_{t\in N}$. Then we claim that there must exist a $\bar{p}$ with $p_t = \bar{p}$, for all t. Suppose, per absurdum, that there exist i, j with $p_i \neq p_j$; since the number of agents is at least three, there exists an agent $k \neq i, j$. This agent can choose $p'_k \neq p_i, p_j$ and x' such that $p'\cdot x' = p'\cdot\omega_k$ and $\|x'\|$ arbitrarily large. Then by rule 1, he/she can receive arbitrarily close to the entire endowment of this economy. Since, by the game rules, a Nash allocation would not have given agent k the whole endowment ω, the strategy (p'_k, x') is more advantageous to k, and so s is not an NE.

Given the common announced price $\bar{p}$, the bundle $\bar{y}_t$ is as good for agent t as any y in the set $\{y | y \leqslant \omega, \bar{p}\cdot\omega_t = \bar{p}\cdot y\}$ for every agent t. For if there were an agent t' and a bundle y such that $y \leqslant \omega$, $\bar{p}\cdot y = \bar{p}\cdot\omega_{t'}$, and $y p_{t'} \bar{y}_{t'}$, agent t' could change his strategy to $\left(p, \dfrac{p\cdot\omega'_k}{p\cdot y}\right)y$ where $p \neq \bar{p}$. [Here $p_{t'}$ represents the strict preference relations of agent t']. Then $H_{t'}$ would be calculated by rule 4.1 to be

$$\{(\bar{p}\cdot\omega_{t'})/(\bar{p}[(p\cdot\omega_{t}'/p\cdot y)y])\}\cdot[(p\cdot\omega_{t'}/p\cdot y)y] = y.$$

Since agent t'. strictly prefers y to $\bar{y}_t$, this would contradict the fact that $(s_t)_{t\in N}$ is a Nash equilibrium. Thus, for any t, we have $\bar{y}_t R_t y$, $\forall y \leqslant \omega$ such that $\bar{p}\cdot y \leqslant \bar{p}\cdot\omega_t$. Also, $\bar{p}\cdot\bar{y}_t = \bar{p}\cdot\omega_t$. Clearly $\Sigma_{t\in N}\bar{y}_t = \omega$ and, hence, $[\bar{p}, (\bar{y}_t)_{t\in N}] \in$ CWA. Q.E.D.

Note that the outcome function H is always individually feasible. Postlewaite and Wettstein (1983) have shown that the outcome function above can be modified so as to implement the constrained Walrasian equilibria with a <u>continuous</u> outcome function.

III. ECONOMIES WITH PUBLIC GOODS

Consider economies where private goods can be used as inputs to produce public goods.

III.A. To simplify exposition, start with the case where there is only one private good X and one public good Y and it takes one unit of X to produce one unit of Y. (In what follows, we shall always be assuming that the designer knows the production function for the public good. Hence the normalization used is legitimate.) As a further simplification we shall also assume zero initial endowments of Y and positive endowments of X for all agents.

The problem of implementability is posed as in the preceding sections. We begin, as in Section II, by considering the case of *known preference profiles* while the *initial endowments are unknown and may be withheld* (but not destroyed).

It turns out that, even with the assumptions made in Section II (semipositive endowments, strictly increasing preferences, and IR performance functions) we must impose a further restriction on the performance function in order to obtain W-implementability.[55]

Let $\underline{v} \in \mathbb{R}^n_{++}$ be an X-endowment profile and write

$$f(\underline{v}) = (f_i^X(\underline{v}),\ f_i^Y(\underline{v}))_{i\in N}$$

where $f_i^X(\underline{v})$ is the net trade in X received by agent i given $\underline{v}$, and $f_i^Y(\underline{v})$ is the net trade in Y received by i given $\underline{v}$.[56] Then the additional assumption we are adopting is that

$$(*)\qquad v_i + f_i^X(\underline{v}) > 0 \quad \text{for all} \quad i \text{ and all } \underline{v} \gg 0.$$

To W-implement f we shall use here a modified form of the game introduced in Section II.

First, we expand the strategy spaces used in the proof of Theorem 1, by adding to each agent's message a statement as to his/her desired level of the public good Y, to be denoted by y_i.[57] So the generic form of i's strategy is

$$s_i = (w_i^1, \ldots, w_i^n, y_i) = (w_i, y_i),$$

where w_i^j, is agent i's statment about j's X-endowment, and, as previously

$$s = (s_1, \ldots, s_n).$$

The outcome function $h = (h^X, h^Y)$ is as follows:

(a_{PB}) (Unanimity with regard to endowments)

If there is a $\underset{\rightarrow}{v} \in \mathbb{R}_{++}^n$ such that

$$s_i = (\underset{\rightarrow}{v}, y_i) \; \forall i \in N,$$

then

$$h(s) = f(\underset{\rightarrow}{v}).$$

(b_{PB}) (No unanimity with regard to endowments and[58] $M(s) = \emptyset$)

For every $i \in N$,

$$h_i^X(s) = \left[\frac{\beta_i(s)}{\Sigma_{j \in N} \beta_j(s)}\right] w(s) - w_i^i$$

where

$$\beta_i(s) = \sum_{\substack{j \in N \\ j \neq i}} \sum_{\substack{k \in N \\ k \neq i}} \| w_j^j - w_k^j \|, \quad i \in N$$

and

$$h^Y(s) = 0.$$

(c_{PB}) $M(s) \neq \emptyset$;

(c_{PB}^*) $\# M(s) > 1$:

$$h_i^X(s) = \left[\frac{\beta_i(s)}{\Sigma_{j \in M(s)} \beta_j(s)}\right] w(s) - w_i^i \quad \text{for} \quad i \in M(s),$$

$$h_i^X(s) = -w_i^i \quad \text{for} \quad i \notin M(s),$$

and

$$h^Y(s) = 0.$$

(c^{**}_{PB}) <u>$M(s)$ is a singleton, say $M(s) = \{i\}$</u>:

1. $h_j^X(s) = -w_j^j$ for all $j \neq i$;
2. if $\sum_{k\in N} w_k^k - y_i \geqq 0$, then

$$h_i^X(s) = \sum_{k\in N} w_k^k - w_i^i - y_i$$

and

$$h^Y(s) = y_i;$$

3. if $\sum_{k\in N} w_k^k - y_i < 0$, then

$$h_i^X(s) = \sum_{k\in N} w_k^k - w_i^i$$

and

$$h^Y(s) = 0.$$

The condition $\Sigma_{k\in N} w_k^k - y_i \geqq 0$ assures the non-negativity of the final holdings[59] for agent i, since when it is satisfied, we have

$$\begin{aligned} H_i^X(s) &= \mathring{\omega}_i + h_i^X(s) \\ &= \mathring{\omega}_i + \sum_{k\in N} w_k^k - w_i^i - y_i \\ &= (\mathring{\omega}_i - w_i^i) + \left(\sum_{k\in N} w_k^k - y_i\right). \end{aligned}$$

Since $\mathring{\omega}_i - w_i^i \geqq 0$ by the rules of our game, $\Sigma_{i\in N} w_k^k - y_i \geqq 0$ implies $H_i^X(s) \geqq 0$, i.e., individual feasibility for agent i.

We see that the outcome rules are essentially the same as in the absence of public goods, except when $M(s)$ is a singleton.

III.B. Consider now the following broader class of economies, E. There are r private goods $X^1, \ldots, X^r$ and $l - r$ public goods $Y^{r+1}, \ldots, Y^l$, $r \geqq 1$, $l - r \geqq 1$ where the private goods serve as consumer goods and possibly also as inputs for the production of the public goods. The generic input-output vector is written (x, y), $x \in -\mathbb{R}^r_+$, $y \in \mathbb{R}^{l-r}_+$. The production possibility set (assumed known to the designer) is denoted by $A \subset (-\mathbb{R}^r_+) \times \mathbb{R}^{l-r}_+$. It is assumed that $0 \in A$. For each agent $i \in N$, the initial X-endowment, denoted by ω_i, is semi-positive ($\omega_i \in \mathbb{R}^r_{+0}$, i.e., $\omega_i \geqslant 0$), while there are no initial endowments of the public goods. Also, every agent's preference relation $\mathring{R}_i$ (defined on $\mathbb{R}^l_+$) is strictly increasing in all goods, private and public. Thus $E = \{e | e = (\omega_1, \ldots, \omega_n, \mathring{R}_1, \ldots, \mathring{R}_n, A); \omega_i \geqslant 0 \text{ for all } i\}$.

For a given production possibility set A and all semi-positive X-endowments, the set of conceivable outcomes is

$$Z_A = \left\{(t_1, \ldots, t_n, y): t_i \in \mathbb{R}^r \ \forall\, i \in N, y \in \mathbb{R}^{l-r}_+, \left(\sum_{i \in N} t_i, y\right) \in A\right\},$$

where $t_i \in \mathbb{R}^r$ is the net transfer vector of private goods to agent i (with $\Sigma_{i \in N} t_i \leqslant 0$ the input vector used in the production of public goods) and $y \in \mathbb{R}^{l-r}_+$ is the product public goods vector. Then the performance function is $f: E \to Z_A$. Let $f(e) = (t_1, \ldots, t_n, y)$. We write

$$t_i = f_i^X(e), \quad y = f_i^Y(e).$$

The generic strategic message of agent i is $s_i = (w_i^1, \ldots, w_i^n, x_i, y_i)$ with $w_i^j \in \mathbb{R}^r_{+0}$, $(x_i, y_i) \in (-\mathbb{R}^r_+) \times \mathbb{R}^{l-r}_+$. Rules (a_{PB}), (b_{PB}), (c^*_{PB}), and rule 1 in (c^{**}_{PB}) defining h remain unchanged,[60] while rules 2 and 3 in (c^{**}_{PB}) are modified as follows:

$(c^{**}_{PB}.2'.)$ if $\sum_{k \in N} w_k^k + x_i \geqq 0$ and $(x_i, y_i) \in A$, then

$$h_i^X(s) = \sum_{k \in N} w_k^k - w_i^i + x_i,$$

and

$$h^Y(s) = y_i;$$

$(c^{**}_{PB}.3'.)$ if either $\sum_{k \in N} w_k^k + x_i \not\geqq 0$ or $(x_i, y_i) \notin A$, then

$$h_i^X(s) = \sum_{k \in N} w_k^k - w_i^i$$

and

$$h^Y(s) = 0.$$

This outcome function is used to prove the W-implementability Theorem VII (i) below.

For this broader class E of economics, the assumption $(*)$ above can be replaced by the following:

(**) For every economy $e \in E$, and for every agent $j \in N$,

$$\mathring{\omega}_j + f_j^X(e) \geqslant 0,$$

that is, everyone is left with some private goods.

Remark 9. Condition (**) would, in particular, be satisfied if f is individually rational, and if, for all j, $\mathring{\omega}_j \geqslant 0$, and

$$(***)_j \qquad (x, y)\mathring{P}_j(0, y')$$

for any $x \geqslant 0$ and any $y, y' \geqq 0$.

We shall see that (**) holds if $(***)j$ holds for all j.

Proof of the Remark: Suppose that $\mathring{\omega}_j + f_j^X(e) = 0$. Since $\mathring{\omega}_j \geqslant 0$, the above condition on j's preferences implies

$$(\mathring{\omega}_j, 0)\mathring{P}_j((\mathring{\omega}_j, 0) + f_j(e)),$$

which violates the IR property of f.

Theorem 8: Let $n \geqq 3$, $\underset{\rightarrow}{\mathring{\omega}} \in \mathbb{R}^{rn}_{+0}$, all preferences strictly increasing, and let f be $\mathbb{R}$ and satisfy assumption (**).[61] Then, for the class E of economies defined at the beginning of III B,

(i) f is W-implementable (when the preference profile is known to the designer but endowments are not);

and

(ii) f is W-R-implementable (when neither the preferences nor the endowments are known to the designer and the endowments can be withheld but not destroyed) if and only if f is monotone.

Proof: For (i) we use the strategy spaces and outcome function defined before Remark 9, on pp. 70–71, and again follow the pattern of proof of Theorem I (i) in Sect. II.A.1. It need only be noted that, in the proof of Claim 2, assumption (**) on f, together with the rule (c_{PB}^{**}), provide an incentive for agent i to break away from the unanimous agreement on a false endowment profile.

To prove (ii), we use a generic strategy space element $\sigma_i = (w_i, y_i, d_i)$ and proceed as in the proof of Theorems II.A. and II.B above.

For the necessity part of (ii), f must be monotone by Maskin's (1977) Theorem 2.[62]

The proof that the monotonicity of f is sufficient in (ii) follows the pattern of II.B except for the treatment of the "no veto power" (NVP) assumption. In II.B we applied Maskin's (1977) Theorem 5 on R-implementability[63] (which assumes that f satisfies NVP) by noting (in footnote 37) that, under the assumptions of II.B ($n \geqq 3$, pure exchange with private goods only, strictly increasing preferences) NVP holds vacuously. In the present section we show that a variant of Maskin's Theorem 5, applicable in E, permits us to dispense with NVP.[64]

The variant used differs from Theorem 5 only in that it presupposes an economy $e \in E$ as defined above and that it dispenses with NVP.

The proof of the new variant is the same as that of Theorem 5, except that the outcome function g is modified for the case with at least two different individual strategies. Property (5) of g in Maskin's Theorem 4 becomes

(5′) If, for $(s_1^*, \ldots, s_n^*) = s^* \in S$, there exists $i \in N$ such that it is *not* true that $s_1^* = \ldots = s_{i-1}^* = s_{i+1}^* = \ldots = s_n^*$, then the range of outcomes accessible to agent i consists of all outcomes $z = (t_1, \ldots, t_n, y)$ such that $(\Sigma t_q, y) \in A$ and $\mathring{\omega}_q + t_q > 0$ (i.e., $\mathring{\omega}_q + t_q \in \mathbb{R}^r_{++}$) for all $q \in N$.[65] (Such outcome functions do exist.)

Now, in the proof of Maskin's Theorem 5, NVP was only used (through Maskin's Theorem 4) to ensure that any Nash equilibrium without unanimity will be f-optimal. On the other hand, for economies in E (with strictly increasing preferences), with an outcome function satisfying the above condition (5′), a Nash equilibrium without unanimity cannot exist. For suppose that, in the R-game, $s^* = (s_1^*, \ldots, s_n^*)$ is such an NE, with $s_j^* \neq s_k^*$. Then any agent $i \in N/\{j, k\}$ satisfies the hypothesis of (5′). (It could be that $s_i^* = s_j^*$ or $s_i^* = s_k^*$, and that there are only *two* different individual strategies!)

Denote the outcome function of the R-game by g. By (5′),

$$\mathring{\omega}_q + g_q^X(s^*) > 0 \quad \text{for all} \quad q \in N,$$

and i has available a strategy s_i such that $g_i^X(s_i, s^*_{)i(}) > g_i^X(s^*)$, while $g^Y(s_i, s^*_{)i(}) = g^Y(s^*)$. Say $t_q(s_i, s^*_{)i(})$ is slightly bigger than t_q^* for $q \neq i$ and t_i slightly smaller, with $\Sigma t_i = \Sigma t_i^*$. Then, by preferences strictly increasing in X, $g_i(s_i, s^*_{)i(})\ P_i\ g_i(s^*)$, and hence s^* is not an NE.

Using these facts, we establish that monotonicity of f is sufficient in part (ii) of Theorem VII.

Let E' be a subclass of the (above defined) class E of environments such that, in E', the constrained Lindahl Correspondence satisfies the condition (**) and is singleton-valued (i.e., is a function). Since this correspondence is IR and monotone, we obtain from Theorem 8 the following

Corollary 8.1. For $n \geqslant 3$, the constrained Lindahl performance (social choice) function is W- and W-R- implementable in Nash equilibria over E'.

(It appears that, with suitable modifications of the game forms used in the proofs, these results can also be shown to hold for correspondences.)

Example: This example shows that Condition (**), just before Remark 9, cannot be dispensed with in Theorem 8 (i), i.e., for the W-implementability of f.

In this example, there are two goods, one public (y), one private (x). It takes one unit of the private good to produce one unit of the public good. It is assumed that preferences are known but endowments are unknown and can be withheld. There are three persons with preferences given by the following utility functions and initial endowments: $U_1(x_1, y) = y = U_2(x_2, y)$, $\omega_1 = (1, 0) = \omega_2 = \omega_3$. $U_3(x_3, y)$ is such that $U_3(1, 3) > U_3(x, 4 - x)$ for all x satisfying $0 < x \leqslant 4$, $x \neq 1$.

We show that the following performance function f cannot be implemented:

$$f[(1, 0), (1, 0), (1, 0)] = (-1, -1, -1; 3)$$
(i.e., if $\omega_1^x = \omega_2^x = \omega_3^x = 1$, then $x_1 = x_2 = x_3 = 0$, and $y = 3$);

$$f[(1, 0), (1, 0), (2, 0)] = (-1, -1, -2; 4)$$
(i.e., if $\omega_1^x = \omega_2^x = 1$, $\omega_3^x = 2$ then $x_1 = x_2 = x_3 = 0$, and $y = 4$).

Claim: f cannot be implemented.

Proof: Suppose, per absurdum, that f can be implemented. Then for the economy with endowments $\omega_i = (1, 0)$ $i = 1, 2, 3$, there exists $s^* = (s_1^*, s_2^*, s_3^*)$ such that s^* is an NE and $h(s_1^*, s_2^*, s_3^*) = (-1, -1, -1, 3)$. Suppose for the economy with endowments $\{(1, 0), (1, 0), (2, 0)\}$ the same strategies s_i^* are also used. Clearly these are optimal for person 1 and person 2 since if they had a better strategy, s^* would not have been an NE for the original economy. By hypothesis $U_3(1, 3) > U_3(x, 4 - x)$ for all x satisfying $0 < x \leqslant 4$, $x \neq 1$. If $h_3(s_1^*, s_2^*, s_3^*) = (-1, 3)$ the game leaves a final holding for agent 3 of $(1, 3)$. By hypothesis this gives higher utility than any feasible bundle. Hence if h is feasible there cannot be any $\hat{s}_3$ such that $h_3(s_1^*, s_2^*, \hat{s}_3)$ is preferred to $h_3(s_1^*, s_2^*, s_3^*)$. Thus s^* is an NE for the second economy which gives rise to an allocation which does not agree with f for this economy. Q. E. D.

Note: Concerning the implementability of the Lindahl correspondence: In the following diagram there are indifference curves drawn for Mr 3 with the property that $U(1, 3) > U(x, 4 - x)$ $0 < x \leqslant 4$, $x \neq 1$. Furthermore, they are drawn so that f yields a Lindahl equilibrium for each of the two economies. To see this, note that the budget lines drawn through $x = 1$ has slope $= -3$ and is the budget set if $t_3 = 1/3$. The tangency is at $y = 3$. If $t_1 = t_2 = 1/3$ as well we see that Mr 1 and Mr 2 both demand $y = 3$ as well; thus f yields a Lindahl equilibrium for the first economy. If $w_3 = (2, 0)$ and $t_3 = 1/2$, the line through $x = 2$ is Mr 3's budget line. The tangency here is at $y = 4$. If $t_1 = t_2 = 1/4$ both Mr 1 and Mr 3 desire $y = 4$ and f yields a Lindahl for the second economy as well.

It appears that the preferences can be such that the Lindahl equilibria are unique. Thus we see that if we do not impose some conditions on preferences (or on the production possibility set, such as a private good not used for production) the Lindahl performance function cannot be implemented.

IV. PRODUCTION

The design of mechanisms for situations where production sets are not known to the designer is of interest because the problem of revealing true productivity does arise in practice.

For the sake of simplicity, we shall only deal with the case where initial endowments as well as preference profiles (but not the production possibility sets) are known to the designer. The extension to cases where these assumptions do not hold would be treated by methods analogous to those used in earlier sections.

Each participant is characterized by $e^i = (\omega_i, Y_i, R_i)$ where ω_i is the true endowment, R_i the true preference relation and Y_i the true production possibility set, assumed closed. (Subsequently we shall omit the circles over the ω's and R's since only true values will be entering the picture.)

The social choice function to be implemented f, assumed feasible, associates an allocation $z = (z_1, \ldots, z_n)$, a point of the outcome space Z, with a profile $(Y_1, \ldots, Y_n)$ of production possibility sets. For each i in

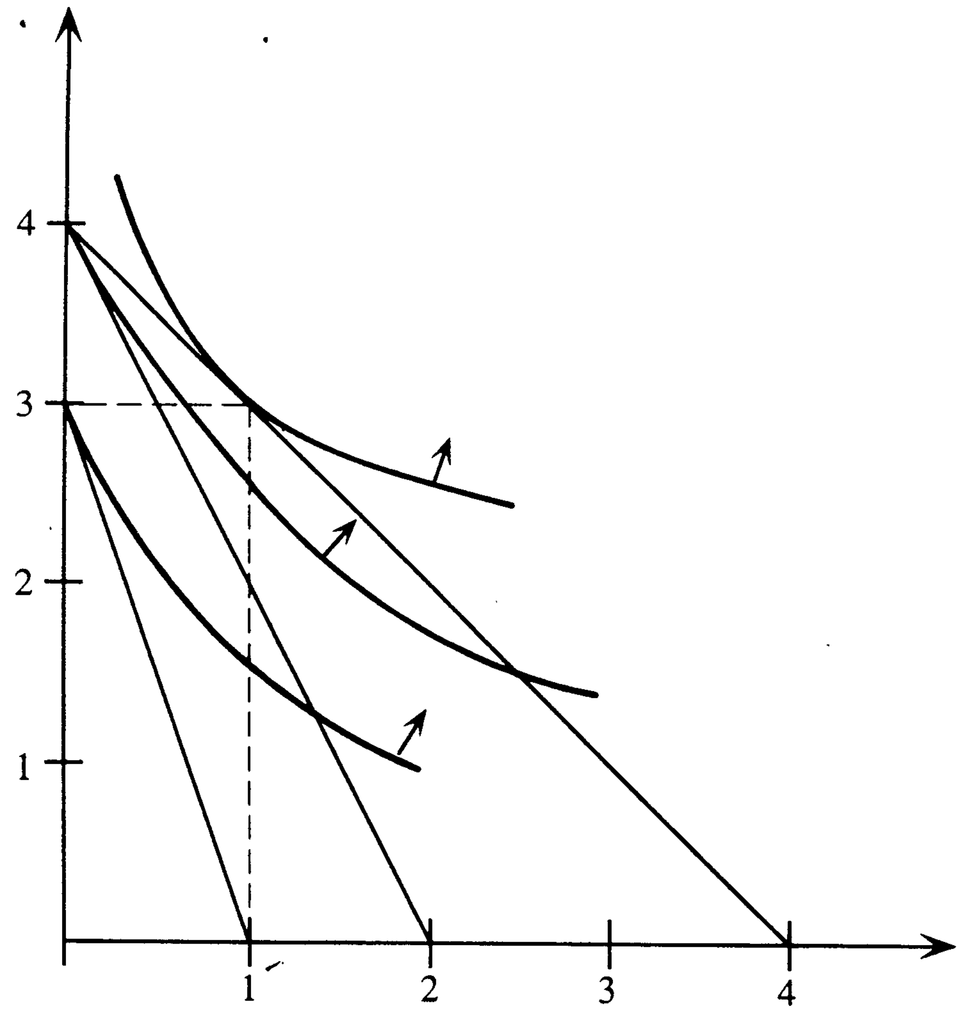

Figure 2.

N, $z_i = (x_i, y_i)$ where x_i is a net trade and y_i the input-output vector for agent i. (We are suppressing the dependence of f on endowments and preferences because these are assumed known to the designer.)

The implementability of f will be proved by constructing an implementing mechanism (S, h). The i-th agent's strategy will be of the form

$$s_i = (q_i, Y_i^1, \ldots, Y_i^n)$$

where Y_i^j is i's "estimate" of j's production possibility set and q_i is a point in the l dimensional non-negative orthant $\mathbb{R}^l{+}$ of the commodity space. Agents are not permitted to "exaggerate" their own production possibilities; i.e., $Y_i^i \subset Y_i$.

The implementing function h will have two components h^x, h^y, for net trades and production, respectively. It is somewhat similar to that for the destruction endowment game. There are, again, three rules, depending on whether there is unanimity, etc.

Notation:

$N = \{1, \ldots, n\}$ = the set of agents ($n \geqq 3$)
$L = \{1, \ldots, l\}$ = the set of commodities ($l \geqq 1$)
$\omega_i = \mathring{\omega}_i$ = the true initial endowment of agent i
$\omega = \Sigma_{i \in N} \omega_i$
$\mathring{Y}_i$ = the true production possibility set of agent i, $0 \in \mathring{Y}_i \subsetneqq \mathbb{R}^l$
$\underset{\rightarrow}{\mathring{Y}} = (\mathring{Y}_1, \ldots, \mathring{Y}_n)$
$\mathring{Y} = \Sigma_{i \in N} \mathring{Y}_i$
$R_i = \mathring{R}_i$ = the true preference relation of agent i, defined on $\mathbb{R}^l_+$ (= the consumption set)
$f\colon \underset{\rightarrow}{Y} \mapsto z = (z_1, \ldots, z_n)$ = the performance function[66] (= social choice rule)
$z_i = (x_i, y_i) \in \mathbb{R}^l \times \mathbb{R}^l$ where x_i is the net trade (increment) for agent i, y_i is the production plan (input-output vector) for agent i. (Outputs are the positive components of y_i.) If $f(\underset{\rightarrow}{Y}) = (x_1, y_1; \ldots; x_n, y_n)$,

we write

$$x_i = f_i^x(\underset{\rightarrow}{Y})$$

and

$$y_i = f_i^y(\underset{\rightarrow}{Y})$$

for each $i \in$ N.

Theorem 9:

Assume: $n \geqq 3$;

$\omega_i \geqq 0$ for all $i \in N$;
$0 \in \mathring{Y}_i \subsetneqq \mathbb{R}^l$ for all $i \in N$;[67]
($\underset{\rightarrow}{\mathring{Y}} = (\mathring{Y}_1, \ldots, \mathring{Y}_n)$ is called *admissible* if it satisfies the above conditions.)

R_i is strictly increasing in all components, and depends only on the agent's final allocation $\omega_i + x_i$ (not on production plans), for all $i \in N$.

Assume (ω_i, R_i) is known to the designer for all $i \in N$. [(ω_i, R_i) will be suppressed as an argument of f.]

Assume the performance function f is non-confiscatory (NC) and feasible. That is, for all admissible $\mathring{\underline{Y}}$,

$$\omega_i + f_i^x(\mathring{\underline{Y}}) \geqslant 0 \quad \text{for all} \quad i \in \mathrm{N},$$

and, for all admissible $\mathring{\underline{Y}}$,

$$\sum_{i \in N} f_i^x(\mathring{\underline{Y}}) \leqslant \sum_{i \in N} f_i^y(\mathring{\underline{Y}}).$$

Then f is implementable in Nash equilibria.

Proof: The proof is carried out by constructing strategy domains and an outcome function (game form) which implements a given f.

The strategy space for agent i is denoted S_i and is defined as follows:

$$S_i = \{(q_i, Y_i^1, \ldots, Y_i^n): q_i \in \mathbb{R}_+^l;$$
$$0 \in Y_i^i \subsetneqq \mathring{Y}_i;\ 0 \in Y_i^j \subsetneqq \mathbb{R}^l \quad \text{for all} \quad j \neq i\}.$$

We write $S = S_1 \times \ldots \times s_n$.

The n-tuple $Y_i \equiv (Y_i^1, \ldots, Y_i^n)$ is called the i-th *production profile*. We shall write $\omega = \sum_{i \in N} \omega_i$ and $Y = \sum_{i \in N} Y_i$.

We define the set

$$F(Y) = \{v \in \mathbb{R}_+^l: v \in \omega + Y\}$$

(Clearly, $F(Y)$ is non-empty because $\omega \in F(Y)$ since $\omega \geqslant 0$, and $0 \in Y$ by virtue of $0 \in Y_i^i$.)

For any s in S, we define the subset of agents

$$M(s) = \{i \in N \mid Y_j^i \subsetneqq Y_i^i, \quad \forall j \neq i\}.$$

A metric d on closed subsets of the Euclidean space $\mathbb{R}^p$, $p \geqq 1$, is defined as follows.

Given two closed sets, A and B, we first introduce the real-valued function $g(\cdot, B)$: $\mathbb{R}^p \to \mathbb{R}$, defined by

$$g(a, B) = \min_{b \text{ in } B} d(a, b)]/[1 + d(a, 0) + d(b, 0)]$$

where $d(x, y)$ is the Euclidean distance between x and y, and 0 is the origin of R^p. Note that $0 \leqslant g(a) < 1$.

In turn, let $r(A, B) = \max_{a \text{ in } A} g(a, B)$. Again, $0 \leqslant r(A, B) < 1$. The desired metric $d(\cdot, \cdot)$ is then defined by

$$d(A, B) = \max[r(A, B), r(B, A)].$$

Note that $0 \leqslant d(A, B) < 1$, $d(A, A) = 0$, $d(A, B) = d(B, A)$, and if $A \neq B$ then $d(A, B) \neq 0$.

The outcome function $h: S \to Z$, written

$$h = (h_1, \ldots, h_n),\ h_i(s) = (h_i^x(s), h_i^y(s)), \quad i \in N$$
$$h_i^x: S \to \mathbb{R}^l,\ h_i^y: S \to \mathbb{R}^l$$

is defined by the three rules that follow. ($h_i^x(s)$ and $h_i^y(s)$ are respectively the net trade (increment) and the production plan for agent i given the strategy n-tuple s.)

(a_{PR}) (*Unanimity with respect to production profiles.*)

If $Y_1 = \ldots = Y_n$, (i.e., if $Y_i^j = Y_j^j$ for all $i, j \in N$), then $h_i(s) = f_i(Y_1^1, \ldots, Y_n^n)$ for all $i \in N$.

That is, if agents agree on the production profiles, the net trades and production plans are those prescribed by the performance function.

(b_{PR}) (*There is no unanimity with respect to production profiles, and the set $M(s)$ is empty.*)

Two cases are distinguished, depending on whether there are more than two distinct estimates of any individual's production set. We further define, for any $k \in N$ and any $s \in S$,

$$\beta_k(s) = \sum_{i \neq k} \sum_{j \neq k} d(Y_i^j, Y_j^j).$$

The number of distinct estimates in s will be devoted by $t(s)$.

(b_{PR}.1) Let $M(s) = \emptyset$ and let there be exactly *two* distinct profiles among $Y_1, \ldots, Y_n$. I.e., $t(s) = 2$. Then $\Sigma_{j \in N} \beta_j(s) > 0$. The outcome function is then defined for every $i \in N$ by

$$h_i^x(s) = (\beta_i(s) / \sum_{j \in N} \beta_j(s))\, \omega - \omega_i$$

and

$$h_i^y(s) = 0.$$

That is, in this case there is no production and the aggregate of endowments is divided among the agents according to the accuracy of their estimates of other agents' production sets.

(b_{PR}.2) Let $M(s) = \varnothing$ and let there be *at least three* distinct estimates among $Y_1, \ldots, Y_n$. I.e., $t(s) \geqq 3$.

Define, for each $k \in N$ and each $s \in S$,

$$\beta_k^*(s) = 1 + \beta_k(s).$$

Clearly, $\Sigma_{j \in N} \beta_j^*(s) > 0$. Then, for this case, the outcome function is defined by

$$h_i^x(s) = (\beta_i^*(s) / \sum_{j \in N} \beta_j^*(s))\, \omega - \omega_i$$

and

$$h_i^y(s) = 0$$

for each $i \in N$ and each $s \in S$.

(c_{PR}) Let $M(s) \neq \varnothing$. Then the outcome function is defined by the following relations:

(c_{PR}.1) if $\sum_{i \in M(s)} q_i \notin F(Y)$,

then

$$\begin{cases} h_i^x(s) = 0 \\ h_i^y(s) = 0 \qquad \forall i \in N \end{cases}$$

(c_{PR}.2) if $\sum_{i \in M(s)} q_i \in F(Y)$,

then

$$h_i^x(s) = q_i - \omega_i \quad \text{for} \quad i \in M(s),$$

$$h_i^x(s) = -\omega_i \quad \text{for} \quad i \notin M(s),$$

and

$$h_i^y(s) = \hat{y}_i^q \qquad i \in N$$

where $\hat{y}_i^q$ is the i-th component of $\hat{y}^q$, and $\hat{y}^q$ is the element of the set Y^q, to be defined below, selected out of Y^q by a well-defined selections rule. The set Y^q is given by

$$Y^q = \{(y_1^q, \ldots, y_n^q);\ \sum_{i \in N} q_i = \omega + \sum_{i \in N} y_i^q,\ y_i^q \in Y^i \quad \text{for all} \quad i \in N\}.$$

We shall now show that the above outcome function implements the given f. This is accomplished by proving the following:

Proposition 5: $(s_i^*)_{i\in N}$ is a Nash equilibrium if and only if $Y_i^j = \overset{\circ}{Y}_j$ for all $i, j \in N$.

That is, an n-tuple of strategies is a Nash equilibrium if and only if every agent's production profile is truthful. In such a case there is, of course, unanimity as to production profiles.

The proof of the Proposition is contained in the following three claims.

Claim 1p: The truthful unanimous n-tuple $s^* = (\underset{\rightarrow}{\overset{\circ}{Y}}, \ldots, \underset{\rightarrow}{\overset{\circ}{Y}})$, $\underset{\rightarrow}{\overset{\circ}{Y}} = (\overset{\circ}{Y}_1, \ldots, \overset{\circ}{Y}_n)$ is a Nash equilibrium.

Proof: Suppose not. Then there must be an agent i and a strategy $\tilde{s}_i = (\tilde{\lambda}_i, \tilde{Y}_i)$ which yields agent i higher satisfaction. But $\tilde{Y}_i^i$ cannot strictly contain Y_j^{i*} for all $j \neq i$, since $Y_j^{i*} = \overset{\circ}{Y}_i$ for all $j \in N$. Furthermore, i (who is being truthful) cannot "overstate" his own production set. So $i \notin M(\tilde{s})$. In fact, because the others remain unanimous, $M(\tilde{s}) = \emptyset$. Thus any change from s_i^* results in the application of rule $(b_{PR}.1)$ with $\beta_i(\tilde{s}) = 0$; hence, in $\tilde{s}$, agent i receives nothing: $H_i^x(\tilde{s}) = 0$.

On the other hand, since s^* is unanimous, rule (a_{PR}) applies and $H_i^x(s^*) = \omega_i + f_i^x(\underset{\rightarrow}{\overset{\circ}{Y}})$. Since f is assumed NC, it follows that $H_i^x(s^*) \geqslant 0$, hence $H_i^x(s^*) \geqslant H_i^x(\tilde{s})$. Since preferences are strictly increasing, s^* is strictly preferred to $\tilde{s}$. This contradiction shows that s^* is an NE.

Claim 2p: A non-truthful unanimous n-tuple $s = (s_1, \ldots, s_n)$, $s_1 = s_2 = \ldots = s_n = \underset{\rightarrow}{Y}^*$, $\underset{\rightarrow}{Y}^* \neq \underset{\rightarrow}{\overset{\circ}{Y}}$, is not a Nash equilibrium.

Proof: To see this note that there exists an agent i who is "underreporting," i.e., $Y_i^{i*} \subsetneqq \overset{\circ}{Y}_i$. Suppose that this agent i switches to $\tilde{s}_i = (\tilde{z}_i, \tilde{Y}_i^i)$ with $\tilde{Y}_i^i = \overset{\circ}{Y}_i$. Then $M(\tilde{s}) = \{i\}$, i.e., $Y_i^{i*} \subsetneqq \overset{\circ}{Y}_i$. Suppose that this agent switches to $\tilde{s}_i = (\tilde{q}_i, \tilde{Y}_i^i)$, with $\tilde{Y}_i^i = \overset{\circ}{Y}_i$ while others retain their strategies without change. Then $M(\tilde{s}) = \{i\}$, and, provided that $\tilde{q}_i \in F(\tilde{Y}_i^i + {}_{j\neq i} Y_j^{i*})$, by rule $(c_{PR}.2)$, $h_i^x(\tilde{s}) = \tilde{q}_i - \omega_i$.

Suppose now that

$$\tilde{q}_i = \sum_{j\in N} [h_j^x(s) + \omega_j].$$

Clearly, this is feasible since

$$\sum_{j\in N} [h_j^x(s) + \omega_j] \in Y^* \subset \tilde{Y}_i^i + \sum_{j\neq i} Y_j^{i*}.$$

Also, since f is NC and in the case of (false) unanimity, $h^x(s) = f^x(\underset{\rightarrow}{Y}^*)$, we have

$j \in M(s)$

$$\tilde{s}_j = s_j \quad \text{for all} \quad j \neq 1$$

and

$$\tilde{s}_1 = (\tilde{q}_1, \tilde{Y}_1), \quad \text{such that} \quad \tilde{Y}_1^k = Y_k^k, \quad \forall\, k \in N,$$

and

$$\tilde{q}_1 = \sum_{k \in N} [\omega_k + h_k^x(s)]. \quad \text{Then} \quad M(\tilde{s}) = \{1\}$$

$$\text{and} \quad h_1^x(\tilde{s}) = \tilde{q}_1 - \omega_1 = h_1^x(s) + \sum_{j \neq 1} [h_j^x(s) + \omega_j] \geqslant h_1^x(s).$$

Hence again, by strict monotonicity of preferences, s is not an NE.

(ii^P) Suppose now that $M(s)$ is a singleton, say $M(s) = \{1\}$. There are two cases, depending on whether q_1 is or is not in $F(Y)$.

(1) Suppose that $q_1 \notin F(Y)$. Then

$$h_1^x(s) = 0.$$

Consider $\tilde{s}$ with

$$\tilde{s}_j = s_j \quad \forall j \neq 1,$$
$$\tilde{s}_1 = (\tilde{q}_1, \tilde{Y}_1), \quad \text{such that} \quad \tilde{Y}_1 = Y_1,$$

and

$$\tilde{q}_1 = \sum_{k \in N} \omega_k.$$

Then $M(\tilde{s}) = \{1\}$, and

$$h_1^x(\tilde{s}) = \tilde{q}_1 - \omega_1 = \sum_{j \neq 1} \omega_j \geqslant 0 = h_1^x(s).$$

Hence by strict monotonicity of preferences, s is not an NE.

(2) Suppose that $q_1 \in F(Y)$. Since $M(s) \neq \emptyset$, rule (c_{PR}.2) applies to s. Because $M(s) \neq N$, there is an agent $j \notin M(s)$ who, by rule (c_{PR}.2), gets

$$h_j^x(s) = -\omega_j.$$

Therefore

$$H_j^x(s) = \omega_j + h_j^x(s) = \omega_j + (-\omega_j) = 0.$$

Now suppose that agent j accepts everyone's self-evaluation, while other agents remain unchanged, i.e.,

$$\tilde{Y}_r = Y_r \quad \text{for all} \quad r \neq j$$

and

$$\tilde{Y}_j^k = Y_k^k \quad \text{for all} \quad k \in N.$$

Then (by the argument in Theorem 1),

$$M(\tilde{s}) = \varnothing.$$

Hence rule (c_{PR}) does not apply; nor does rule (a_{PR}), since $\tilde{s}$ is not unanimous, by the counterpart of the argument in Theorem 3, Claim 3′, (iii)″. Specifically, there exists a person $k \in M(s)$, $k \neq j$, and a person i, $i \neq j \neq k \neq i$, such that

$$Y_k^k \neq Y_i^k,$$

hence $\tilde{Y}_k^k \neq \tilde{Y}_i^k$, hence $\tilde{Y}_k \neq \tilde{Y}_i$.

Now, since $\tilde{s}$ is not unanimous as to production profiles and $M(\tilde{s}) = \varnothing$, rule (b_{PR}) applies to $\tilde{s}$.

Suppose the applicable part of rule (b_{PR}) is $(b_{PR}.2)$. Then

$$\begin{aligned} h_j^x(\tilde{s}) &= \omega_j + h_j^x(\tilde{s}) \\ &= \omega_j + \Big(\beta_j^*(\tilde{s}) / \sum_{k \in N} \beta_k^*(\tilde{s})\Big)\,\omega - \omega_j \\ &= \Big(\beta_j^*(\tilde{s}) / \sum_{k \in N} \beta_k^*(\tilde{s})\Big)\,\omega \geq 0, \end{aligned}$$

since $\omega \geq 0$ by hypothesis and $\beta_q^*(\tilde{s}) > 0$ for all $q \in N$ and all $\tilde{s} \in S$. On the other hand, let the applicable part of rule (b_{PR}) be $(b_{PR}.1)$. We note that, for agents j, k, and i referred to above,

$$\beta_j(\tilde{s}) \geqq d(\tilde{Y}_k^k, \tilde{Y}_i^k) > 0$$

since $\tilde{Y}_k^k \neq \tilde{Y}_i^k$. Hence in this case

$$\begin{aligned} H_j^x(\tilde{s}) &= \omega_j + h_j^x(\tilde{s}) \\ &= \omega_j + \Big(\beta_j(\tilde{s}) / \sum_{k \in N} \beta_k(\tilde{s})\Big)\,\omega - \omega_j \\ &= \Big(\beta_j(\tilde{s}) / \sum_{k \in N} \beta_k(\tilde{s})\Big)\,\omega \geq 0, \end{aligned}$$

as before. So, in either case

$$H_j^x(\tilde{s}) \geq 0.$$

On the other hand, as seen above, $H_j^x(s) = 0$. Hence $H_j^x(\tilde{s}) \underset{\circ}{\geq} H_j^x(s)$. By the assumption of strictly increasing preferences, $H_j^x(\tilde{s}) \mathring{P}_j H_j^x(s)$, and so s is not an NE.

(iii^P) In this case there is no unanimity and $M(s) = \varnothing$.

The proof in this section is essentially the same as in section (iii)″ for the case of destruction of endowments, with ω_i and ω respectively replacing w_i^i and $w(s)$, and the metric for sets $d(\cdot, \cdot)$ defined above replacing the Euclidean norm of the difference of two endowment profiles. Q.E.D.

Acknowledgment

We are indebted to Professor Lu Hong of Syracuse University for pointing out (see Hong, 1990) errors in an earlier version of the proof of Theorem III, as well as suggesting a way of repairing it. The game form suggested by Hong for the proof is described in footnote 43 below.

Notes

1. Earlier models of tatonnement and of proposed mechanisms designed to implement social choice rules (e.g., Walras or Lindahl) were criticized for not guaranteeing the feasibility at disequilibrium points. Some, like the Walrasian auctioneer, were not balanced (1), others failed to assure individual feasibility. (See Wilson, 1976.)

2. From now on "implementation" is to be understood in the sense of Nash non-cooperative equilibria. Let n be the number of players, Z the outcome space (the space of allocations), S the joint strategy space, i.e., $S = S^1 \times \ldots \times S^n$, where S^i is the strategy domain of the ith player, and let h: $S \rightarrow Z$ be the outcome function. An *SCR*, denoted by F, is a correspondence from the space E of environments into Z, specifying for each environment (economy) e in E a nonempty set in the outcome space Z. An environment (economy) is defined as an n-tuple of characteristics $e^i = (C_i, \omega^i, R^i)$, where, for the ith agent, C^i is the admissible consumption set, wi the initial endowment, and Ri the (weak) preference relation. I.e, $e = (e1, \ldots, en)$ and E is the class of a priori admissible environments. A possible interpretation is that the designer believes (correctly) that an environment (economy) outside of E will not occur.

We say that a mechanism (S, h) Nash *implements* an *SCR F over* a class of environments E if it is the case that, for every e in E, (1) the set of Nash equilibrium outcomes $N_{S,h}(e)$ generated by the mechanism (S, h) is nonempty, and (2) this set $N_{S,h}(e)$ is a subset of $F(e)$. (The term sometimes used in the literature for this concept is "weakly implements.") The mechanism (S, h) is said to *fully* implement F over E if, for every e in E, $N_{S,h}(e) = F(e)$. In most of the present paper we actually deal with a singleton-valued correspondence F, i.e., one equivalent to a function. In that case the two concepts of implementation coincide and we simply say that (S, h) implements the social choice function f, abbreviated *SCF*, the function equivalent to the singleton-valued correspondence F. (A method for extending our results to correspondences is illustrated in the Appendix to section II.A.1.)

3. Those formulating the rules are often collectively referred to as "the designer." Hence the title of this paper.

4. Of course, because of the non-exaggeration requirement, an agent's claim as to his/her own endowment provides partial information as to the true endowment, namely that the true endowment is at least as high as that claimed.

5. For example, maximin, Nash, etc.

6. Preferences do not affect feasibility.

7. Maskin's construction is an algorithm in the sense that it is a 'recipe' for constructing implementing mechanisms for a class of SCR's (by inserting the SCR F in an outcome function schema), rather than a single mechanism. The same remark applies to our results except for those in the Appendix to Part II where a specific mechanism is constructed.

8. On the other hand, the designer does not know which preference profile (from a known family of profiles) will prevail.

9. When the goods are physical their existence (and ownership) might have to be shown. Similarly, proof might be required for claimed rights or entitlements, or ever claimed skills. See discussion in Hong and Page (1994).

10. The example given is for the case of $n = 3$, but can be constructed in an analogous manner for any number of traders greater than one. The reason for using $n = 3$ is to show its relevance for other results in which we assume that there are at least three traders.

11. Using a more direct proof than that in the present paper it can be shown that this result also holds for $n = 2$.

12. I.e., where the balance condition is in the form of a weak inequality rather than equality (called "weak balance").

13. The environment e^{ik} have the endowment profile $\bar{\omega}$ but may differ with respect to preferences.

14. The amount received by i is a positive component z_i.

15. To simplify exposition, we confine ourselves in this section to single-valued social choice rules; subsequently, we shall extend our treatment to correspondances.

16. The term "optimal" is always used in the sense of the given performance function f.

17. We use, here and elsewhere, the somewhat imprecise notation which identifies $(S_i, S_1, \ldots, S_{i-1}, S_{i+1}, \ldots, S_n)$ with $(S_1, \ldots, S_{i-1}, S_i, S_{i+1}, \ldots, S_n)$.

18. $\Pi_{j \neq i} S_j = S_1 \times \ldots \times S_{i-1} \times S_{i+1} \times \ldots \times S_n$.

19. It would be possible to relax our assumptions by replacing the requirement $\mathring{\omega}_i \geqslant 0$ by $\mathring{\omega}_i \geqq 0$ and, at the same time weaken $w_i^i \geqslant 0$ to: $w_i^i \geqslant 0$ if $\mathring{\omega}_i \geqslant 0$. But we cannot permit an agent to claim $w_i^i = 0$ when $\mathring{\omega}_i \geqslant 0$. For let all agents claim zero endowments while in fact $\sum_{i \in N} \mathring{\omega}_i \geqslant 0$. Then, since the possibility of withholding means that $w_i^i + h_i(s) \geqq 0$ for all $i \in N$, the net Nash allocation would have to be 0 for everyone, and this might be non-optimal. If the assumptions were relaxed along the indicated lines, a minor modification would have to be made in the outcome function.

20. Here *W*- is mnemonic for withholding, as distinct from strategies to be labeled *D*-, in which an agent may not withhold but only destroy his endowment, and from those labeled WD-, where the agent may do both.

21. Here 0 is a net trade (the l-dimensional null vector), strictly preferred by i to the net trade $f_i(\mathring{\omega})$.

22. With $\|x\|$ denoting the norm of the vector x; any norm can be used.

23. That is, $w_i^i = -b$.

24. When the requirement $\mathring{\omega}_i \in \mathbb{R}^l_{+0}$ is relaxed to $\mathring{\omega}_i \in \mathbb{R}^l_+$, the above definition is generalized as follows: f is non-confiscatory (NC) if $\forall i \in N$, $\forall \underline{\mathring{\omega}} \in \mathbb{R}^{ln}_+$, $\mathring{\omega}_i \geqslant 0$ implies $\mathring{\omega}_i + f_i(\underline{\mathring{\omega}}) \geqslant 0$.

25. But not necessarily continuous.

26. Note that the continuity of preferences is only needed for the necessity part of Theorm 1.2.

27. Because $\tilde{w}_i^j = w_i^j = \mathring{\omega}_j$ for all $i \neq j$.

28. By assumption, $\#N = n \geqq 3$.

29. These properties of h(·) can be verified directly.

30. *Proof*: (omitting reference to s'):

$$h_j \geqq -w_j^j \quad \text{implies} \quad \sum_{j \neq i} h_j \geqq - \sum_{j \neq i} w_j^j.$$

But balance implies $\Sigma_{j \neq i} h_j = - h_i$. Hence, the previous inequality can be written as $-h_i \geqq -\Sigma_{j \neq i} w_j^j$ which is equivalent to $h_i \leqq \Sigma_{j \neq i} w_j^j$.

31. Since $i \neq j$.

32. $\mathbb{R}^m_{+0} = \{x \in \mathbb{R}^m: x \geqq 0, x \neq 0\}$.

33. The w_i^i component can be interpreted as the i-th agent's claim concerning his own initial endowment.

34. That is, $A(w_1^i, \ldots, w_n^i)$ would be the set of feasible net allocations if $(w_1^i, \ldots, w_n^i)$ were the true endowment profile.

35. Note that the continuity of preference is only needed for the necessity part of this theorem.

36. See also the theorem in Saijo (1988), p. 698, and theorem M^1 in Hurwicz (1986), p. 86; in the latter the assumptions of transitivity and completeness are dispensed with. The latter paper follows Maskin's original schema, with lemmas 1 (p. 88) and 2 (p. 90) corresponding to Maskin's theorems 4 and 5, respectively.

37. This is so because, for $n \geqq 3$, in a pure exchange economy with strictly increasing preferences, the "no veto power" (NVP) requirement in Maskin's Theorem 5 is necessarily satisfied.

38. Since $i \neq j$.

39. It may be that this same modification would also work in Sec. II.A.1.

40. f is *non-confiscatory* (NC) if $\forall i \in N$, $\forall \underline{\mathring{\omega}} \in \mathbb{R}^{ln}_{+0}$, $\mathring{\omega}_i + f_i(\underline{\mathring{\omega}}) \geqslant 0$.

41. NC is however, not a necessary condition for D-implementability.

42. Note, bowever, that for purposes of this section $z_i'\ R_i\ z_i''$ should be interpreted as $(w_i^i + z_i')\ R_i\ (w_i^i + z_i'')$.

43. Because of errors present in an earlier version of the present paper and pointed out in Hong (1990), we are using here a rule (b^*) somewhat different from the rule (b^*) in the earlier version. Actually, the values of the outcoem function in (b^*1) and (b^*2), as given respectively by formulae (#) and (##), are the same as before, but the class of situations covered by (b^*1) as against (b^*2) is different here as compared with the earlier version.

In her (1990) note, Hong suggested the following alternative (rules (H-a, b, c): Let a strategy message of agent i be of the form $(w_i^1, \ldots, w_i^n, m_i)$ where m_i is an integer from the set $\{0, 1, \ldots, n-1\}$. Rules (H-a) and (H-c) are respectively the same as our (both old and new) rules (a) and (c), so that for these rules the integer does not affect the outcome. Rule (H-b1) states that if there is no unanimity, $M(s) = \varnothing$ and there are two distinct endowment formula (#), i.e., it is the same as in our earlier version. The innovation comes in rule (H-b2): if there is no unanimity, $M(s) = \varnothing$, and there are at least three distinct profiles in s, then $h_i(s) = w(s) - w_i^i$ if $i = r$, but $h_i(s) = -w_i^i$ if $i \neq r$, where

$$r = 1 + (\Sigma_{j \in N}\, m_j)_{\text{modulo } n}.$$

[The respective classes of situations covered by (H-b1) and (H-b2) are the same as the corresponding classes in our earlier version of (b_1^*) and (b_2^*)].] A sketch of the proof using the Hong rules is provided in an appendix to this section.

44. I.e., the formula of rule (b) for W-implementation applies.

45. The paragraph starting with the words "Since f is IR . . ."

46. Ending with "Hence s is an NE."

47. In fact $k \in M(s)$ means that $w_k^k \geqslant w_r^k$ for all r in $N/\{k\}$.

48. For suppose that among agents other than i there are present at least two distinct

profiles, say for agents j and k. If j and k disagree as to i's endowment, so that $w_j^i \neq w_k^i$, then i can choose $\widetilde{w}_i^i \geqslant 0$, so that $\widetilde{w}_i^i$ is simultaneously different from w_j^i and w_k^i and not higher than ω_i. On the other hand if j and k agree about i's endowment, then they must disagree about the endowment of some agent r other than i (since, by hypothesis, they are in disagreement). In that case agent i can choose $\widetilde{w}_i^r$ that is different both from w_j^r and w_k^r (without removing any existing disagreements). In either case, the result is that $(\bar{s}) > 2$, contrary to the hypothesis of 1.B".

49. $\beta_r(\bar{s}) = \beta_r(s)$ because $\beta_r(\cdot)$ does not depend on r's statements concerning the others' endowments.

50. The superscript refers to the game (W or D).

51. Such a subset exists since $n \geqq 3$ and there are only two distinct profiles.

52. Under these assumptions the "no veto power" (NVP) condition would be satisfied.

53. It can also be D-R-implemented.

54. We omit the argument $s \in S_1^{\times \cdots \times S_n}$ in $H_r(s)$.

55. See counterexample, preceding Corollary 8.1 below.

56. We suppose the preference profile (assumed known) is an argument of f.

57. In a more general model, with r private goods $X^1, \ldots, X^r$ (used as consumer goods and/or inputs for producing Y), $l - r$ public goods $Y^{r+1}, \ldots, Y^l$, and a production set $A \subsetneqq \mathbb{R}^l$ with $0 \in A$, we would replace y_i by the vector (x_i, y_i) specifying both the desired level y_i of Y and an input vector $x_i = (x_i^1, \ldots, x_i^r)$ and a desired public goods output vector $y_i = (y_i^{r+1}, \ldots, y_i^l)$.

58. Here, as on p. 17. we defined $M(s) = \{i \in N | w_i^i \geqslant w_j^i, \forall_j \neq 1, j \in N\}$; $W(s) = \Sigma_{i \in N} w_i^i$; $\beta_i(s) = \Sigma_{j \neq i}^{j \in N} \Sigma_{k \neq i}^{j \in N} \| w_j^i - w_k^i \|$, $i \in N$.

59. With $\mathring{\omega}_i$ and w_i^k referring to the X-endowments.

60. Except that in (a_{PB}) we require $\underline{Y} \in \mathbb{R}_{+0}^m$.

61. Because here we are assuming f to be IR and $\underline{\mathring{\omega}} \in \mathbb{R}_{+0}^m$, (**) is satisfied if prefeences satisfy $(***)_j$ for each j.

62. In what follows, Theorems 4 and 5 are also from Maskin (1977).

63. R-implementability and R-game refer to the case where the feasible set (hence endowments) is known to the designer but preferences are not.

64. This approach could also havé been used in II.A.2.

65. Property (5) used in Theorem 4 would specify a broader range of outcomes, viz. The whole feasible set; hence, the condition $\omega_i + t_i > 0$ would be weakened to $w_i + t_i \geqslant 0$.

66. Given the (suppressed) endowments and preferences, here assumed known to the designer.

67. The assumptions $\omega_i \geqslant 0$ and $0 \in \mathring{Y}_i$, $\forall i \in N$, imply that the set $\mathbb{R}_+^l \cap (\Sigma_{i \in N} \omega_i + \Sigma_{i \in N} \mathring{Y}_i)$ is non-empty. In fact it contains that semi-positive point $\Sigma_{i \in N} \omega_i$.

References

Groves, T., and Ledyard, J. 1977. "Optimal Allocation of Public Goods A Solution to the 'Free Rider' Problem." *Econometrica* 45, 783–811.

Hong, L. Endowment Game with Destruction. Unpublished note, dated September 27, 1990.

Hong,. L. 1991. "Balanced Nash-implementation of Social Choice Correspondences in Production Economies." Doctoral dissertation, University of Minnesota.

Hong, L. 1994. "Nash Implementation in Production Economies." *Economic Theory*, forthcoming.
Hong, L., and Page, S. 1994. "Reducing Informational Costs in Endowment Mechanisms." Mimeo. *Economic Design*, forthcoming.
Hurwicz, L. 1972. "On Informationally Decentralized Systems." In McGuire, C.B., and Radner, R. (eds.), *Decision and Organization*. Amsterdam: North-Holland, chapter 14.
Hurwicz, L. 1973. "The Design of Mechanisms for Resource Allocation." *American Economic Review* 63, 1–30.
Hurwicz, L. 1979. "On Allocations Attainable through Nash Equilibria." *Journal of Economic Theory* 21, 140–165.
Hurwicz, L. 1986. "On the Implementation of Social Choice Rules in Irrational Societies." In Heller, W. P., Starr, R. M., and Starrett, D. A. (eds.), *Essays in Honor of Kenneth J. Arrow*, Vol. 1. Cambridge University Press, chapter 4.
Maskin, E. 1977. "Nash Equilibrium and Welfare Optimality." Mimeo, MIT.
Nakamura, S. 1989. "Efficient Feasible Nash Mechanisms with Production and Externalities." Doctoral dissertation, University of Minnesota.
Nakamura, S. 1990. "A Feasible Nash Implementation of Walrasian Equilibria in the Two-Agent Economy." *Economics Letters* 34, 5–9.
Page, S. 1989. "Reducing the Dimension of the Message Space in HMP." Unpublished note, dated July 1989.
Postlewaite, A. 1979. "Manipulation via Endowments." *Review of Economic Studies* 46, 255–62.
Postlewaite, A., and Wettstein, D. 1983. "Implementing Constrained Walrasian Equilibria Continuously." CARESS Working Paper No. 83–24, University of Pennsylvania.
Saijo, T. 1988. "Strategy Space Reduction in Maskin's Theorem: Sufficient Conditions for Nash Implementation." *Econometrica* 56(3), 693–700.
Sertel, M. 1990. "Manipulating Lindahl Equilibrium via Endowments." Bôgaziçi University Research Paper, ISS/EC 90–07. *Economics Letters*, forthcoming.
Schmeidler, D. 1976. "A Remark on Microeconomic Models of an Economy and on a Game Theoretic Interpretation of Walras Equilibrium." Mimeo, University of Minnesota, March.
Wilson, R. 1976. "Competitive Processes of Price Formation: A Survey of Several Models." Mimeo, Stanford University (IMSSS), July.