

THE MANIPULABILITY OF THE SHAPLEY-VALUE

by

William Thomson

Discussion Paper No. 79-115, September 1979

Center for Economic Research
Department of Economics
University of Minnesota
Minneapolis, Minnesota 55455

The Manipulability of the Shapley-Value*

by

William Thomson

September 1979

In order to compute the value-allocations of market games, it is necessary to know the participants' preferences, which are unfortunately not directly observable; one has to rely on what the traders claim them to be. However, by misrepresenting his preferences, an agent may be able to affect value-allocations in his favor. If all agents attempt to so manipulate the value, the resulting outcomes should be expected to be different from the outcomes to which truthful behavior would have led.

The object of this paper is to study the vulnerability of the value to manipulative behavior. Manipulation by one individual only and simultaneous manipulation by all the players are successively considered, and equilibrium allocations are characterized. It is shown that under a smoothness requirement on the strategies available to the players, the set of equilibrium allocations of the manipulation game associated with the value coincides with the set of true Walrasian allocations (some qualifications have to be introduced concerning the boundary of the feasible space). This indicates that manipulative behavior does not necessarily lead to a violation of optimality, a (perhaps) unexpected conclusion. Although our starting point is transferable utility economies, the manipulation game associated to the value, if played as an abstract game

*Presented at the IMSSS Seminar, Stanford, August 1979.

This is a revised version of a paper entitled "On the Manipulability of the Shapley-Value," and dated June 1979.

I thank L. Hurwicz, T. Ito, and J. Jordan for several helpful conversations. Assistance from NSF grant SOC-7825734 is gratefully acknowledged.

in an economy without transferable utility, also yield as equilibria the Walrasian allocations.

The general study of the manipulability of economic mechanisms is pursued in Thomson [10], where a bibliography of the recent contributions to this topic is provided.

I. Private Good Case

1. Definitions

Environments: E is the class of transferable utility economies with n agents and m privately appropriable commodities, defined as follows: each agent is indexed by the subscript i in $I = \{1, \dots, n\}$ and characterized by a list $(X_i, \omega_i, \succsim_i)$ where $X_i = R \times R_+^{m-1}$ is his consumption space, ω_i his initial endowment, a point of X_i , and \succsim_i his preference relation defined over X_i , assumed to be continuous, convex, and to admit of a representation of the form " $x_i + v_i(y_i)$ " where $z_i = (x_i, y_i)$ with x_i in R and y_i in R_+^{m-1} is agent i 's consumption. The first commodity, called money, permits utility transfers among the agents. Because of the special structure of the preferences, in order to completely describe \succsim_i , it is sufficient to know either v_i (the normalization condition $v_i(\omega_i y) = 0$ will be imposed on v_i to simplify the exposition), or agent i 's indifference surface through ω_i , denoted m_i , since any other indifference surface can be obtained from m_i by a translation parallel to the x -axis. Given e in E , the set of feasible allocations of e , $F(e)$, is the set of lists $z = (z_1, \dots, z_n)$ in $(R \times R_+^{m-1})^n$ with $\sum z_i \leq \sum \omega_i$. In general, we will assume initial endowments to be known and dependence on ω will not be explicitly stated. z in $F(e)$ is Pareto-

optimal for e if there does not exist z' in $F(e)$ with z'_i at least as good as z_i for all i , and strictly preferred to z_i for at least one i . $P(e)$ denotes the set of Pareto-optimal allocations of e .

Value allocations: The value is introduced by Shapley in [7].

Its application to economics appears in Shapley and Shubik [8]. A coalition S is a subset of I . The worth w^S of S is the maximum aggregate utility it can achieve by redistributing its resources among its members:

$$w^S = \max \sum_{i \in S} [x_i + v_i(y_i)] \text{ st } \sum_{i \in S} z_i \leq \sum_{i \in S} \omega_i .$$

Denoting $w^S \equiv (w_x^S, w_y^S) = \sum_{i \in S} \omega_i$, we have

$$w^S = w_x^S + \max \sum_{i \in S} v_i(y_i) \text{ st } \sum_{i \in S} y_i \leq w_y^S .$$

The set of maximizers of the above bracket is denoted Y^S . It is a subset of $R_+^{|S| (m-1)}$.

The value ϕ_i of agent i is equal to a weighted average of his contributions to the various coalition to which he belongs, where by contribution of agent i to coalition S is meant the difference of the worths of the coalitions S and $S \setminus \{i\}$, (for simplicity $S \setminus \{i\}$ is written as $S \setminus i$): $\phi_i =$

$\sum_{S \in \mathcal{S}_i} k^S [w(S) - w(S \setminus i)]$, \mathcal{S}_i being the class of coalition that contains agent i ,

$k^S > 0$ for all S and $\sum_{S \in \mathcal{S}_i} k^S = 1$. We will lighten the notation by not

writing out the k^S explicitly, except for the two-person case that we will treat in greater detail for illustrative purposes, and because this case lends itself to somewhat more explicit results.

A value-allocation is a feasible allocation such that the utility of each agent be equal to his value. Value-allocations exist. Value-allocations are individually rational and Pareto-optimal. There is a unique value-allocation if and only if y^I contains a single element.

The value is what is often called a performance correspondence. A performance correspondence defined on an environment E associates to every element e of E a set of feasible allocations for e , that are in some sense desirable. The Shapley-value performance correspondence is appealing because it selects individually rational and Pareto-optimal outcomes, and satisfies a fairness requirement (the symmetry axiom).

The computation of value-allocations depend on the agents' preferences, which cannot be observed directly. One has to rely on what the agents claim them to be. Denoting by M_i the space of indifference surfaces through w_i consistent with the assumptions on E specified above, one defines the pair (M, V) where $M \equiv M_1 \times \dots \times M_n$ and $V: M \rightarrow X \equiv X_1 \times \dots \times X_n$ associates to every economy $m = (m_1, \dots, m_n)$ in M its set of value-allocations $V(m)$, as the direct mechanism associated with the value. The adjective "direct" refers to the fact that the message space of each agent can be interpreted as the space of his potential characteristics (by opposition to mechanisms with abstract message spaces), and that if agents are truthful (it is meaningful to speak of the truthfulness of an agent for such a mechanism), the desired allocations are directly obtained. To every performance correspondence Φ , can be similarly associated a direct mechanism (N, Φ) , through an appropriate definition of message spaces.

Once it is recognized that an agent can misrepresent his preferences, it is natural to think of the pair (M, V) as a quasi-game with the M_i as strategy

spaces and V as the outcome correspondence, multiplicities of value-allocations motivating the use of the term quasi-game instead of game.

The analysis of this quasi-game is the object of this paper. Before defining the notion of equilibrium that we will be using, a few extra pieces of notation have to be introduced: denoting $m_{-i} = (\dots, m_{i-1}, m_{i+1}, \dots)$, the list $m = (m_1, \dots, m_n)$ will also be written as (m_i, m_{-i}) . Given m_i in M_i , \succsim_{m_i} designates the preference relation corresponding to the map generated from m_i by a translation parallel to the money axis. For z_i in X_i , we define $W(m_i, z_i) \equiv \{z'_i \in X_i \mid z'_i \succsim_{m_i} z_i\}$, $\bar{W}(m_i, z_i) \equiv \{z'_i \in X_i \mid z'_i \succ_{m_i} z_i\}$. Truthful information is denoted by a circle (e.g., m_i^o is agent i 's true indifference surface through w).

We now elaborate on the difficulties created by the multiplicities of V . When value-allocations are not unique, they are indifferent to one another for all agents according to their announced preferences, but in general not according to their true preferences. This may make it difficult for an individual to evaluate his strategic opportunities: given an allocation z in $V(m)$ for some list m in M , it is conceivable that a strategy m'_i be available to agent i yielding two allocations z' and z'' in $V(m'_i, m_{-i})$ with $z' \succ_{m_i^o} z \succ_{m_i^o} z''$. Assuming that the other agents do not depart from m_{-i} , which of his two strategies m_i and m'_i will agent i prefer? An answer to this question would in general require that his expectations concerning the resolution of multiplicities be specified. An optimistic agent would expect z' to be selected over z'' , while a pessimistic agent would expect the opposite. However, the value has the following property: whenever, for agent i , some m'_i exists as just described, then some other strategy m''_i will also be available with $z' \succ_{m_i^o} z$ for all z' in $V(m''_i, m_{-i})$. This

justifies the use of the following equilibrium concept.

Definition: A pair $(m, z) \in M \times X$ is an equilibrium of the quasi-game (M, V) for the economy e in E , written as $(m, z) \in N(e)$, iff

$$(a) \quad z \in V(m)$$

$$(b) \quad \forall i \in I, \forall m'_i \in M_i, \forall z' \in X, z' \in V(m'_i, m_{-i}) \Rightarrow z \succsim_{m_i^0} z'.$$

If (m, z) is an equilibrium, z is an equilibrium allocation.

This is basically the concept of Nash-equilibrium.

Additional notation: $S^m = \{p \in R_+^m \mid \|p\| = 1\}$. For $m_i \in M_i$ and $z_i \in m_i$, $H(m_i, z_i) \equiv \{p \in S^m \mid \forall z'_i \in m_i, pz'_i \cong pz_i\}$. $h_i(m_i)$ is agent i 's offer curve if he announces $m_i \in M_i$, and $h(m_{-i})$ is the offer curve he faces if the other agents announce $m_{-i} \in M_{-i}$.

2. Unilateral Manipulation of the Value

The analysis of the manipulability of the value will be carried out in two steps. At first, we examine unilateral manipulation.

Definition: A direct mechanism (N, ϕ) defined on E with message space $N = N_1 \times \dots \times N_n$, and outcome correspondence $\phi: N \rightarrow X$ is incentive-compatible iff, for all z in $\phi(n^0)$, (n^0, z) is an equilibrium of the quasi-game (N, ϕ) , where n^0 denotes the list of truthful messages.

Theorem 1: The direct mechanism (M, V) is not incentive-compatible.

Proof: We will prove this result by showing that incentive-compatibility is violated in simple 2-person-2-good economies. This case lends itself to a particularly simple graphical analysis. In fact, we will do more than what is necessary for the proof of the Theorem, and carry out the following tasks:

A: characterize the set of allocations that an agent can make appear as value-allocations;

B: characterize the set of allocations that an agent can enforce as value-allocations;

C: determine when an agent has optimal strategies, and characterize them when they exist;

D: show that unilateral and optimal manipulation of the value by a given agent always leads to an allocation preferred by that agent to that obtained from unilateral and optimal manipulation of the price mechanism;

E: provide a counter-example showing that the gain from unilateral and optimal manipulation of the value may be smaller than the gain from unilateral and optimal manipulation of the price mechanism.

A - E will be done in the context of 2-person, 2-good economies.

F: indicate how these results would be affected by considering any number of agents and commodities.

Preliminaries Concerning 2-Person, 2-Good Economies

In the 2-person case, there are only three possible coalitions,

$S_1 = \{1\}$, $S_2 = \{2\}$ and $S_3 = \{1,2\}$, and we have

$$w^i \equiv w^{\{i\}} = w_{ix} + v_i(w_{iy}) \quad \text{for } i = 1,2 .$$

$$w^{12} \equiv w^{\{1,2\}} = w_x + v_1(y_1^*) + v_2(y_2^*)$$

where $y^* = (y_1^*, y_2^*)$ is a maximizer of $v_1(y_1) + v_2(y_2)$ among all the pairs (y_1, y_2) in R_2^+ satisfying $y_1 + y_2 \leq w_y$.

The value of agent i , denoted ω_i , satisfies

$$\varphi_1 = \frac{1}{2}w^1 + \frac{1}{2}(w^{12} - w^2) = w^1 + \frac{1}{2}[w^{12} - w^1 - w^2]$$

$$\varphi_2 = \frac{1}{2}w^2 + \frac{1}{2}(w^{12} - w^1) = w^2 + \frac{1}{2}[w^{12} - w^1 - w^2] .$$

The term $w^{12} - w^1 - w^2$ is a measure (in terms of the first commodity), of the gains from trade; the term w^i represents the initial utility of agent i ; the above formulae indicate that the agents should share the gains from trade equally.

Given y^* as previously defined, an allocation $z^* = ((x_i^*, y_i^*), i = 1, 2)$ with $x_i^* + v_i(y_i^*) = \varphi_i$ for $i = 1, 2$ is a value-allocation. Geometrically, $F(e)$ can be represented by an Edgeworth box, and $P(e)$ is a horizontal line (or band if y^* is not unique). In Figure 1, the social gain from trade is measured by the length of the contract curve $[z_1', z_2'']$. The value-allocation z^* is there unique and coincides with the middle of $[z_1', z_2'']$.

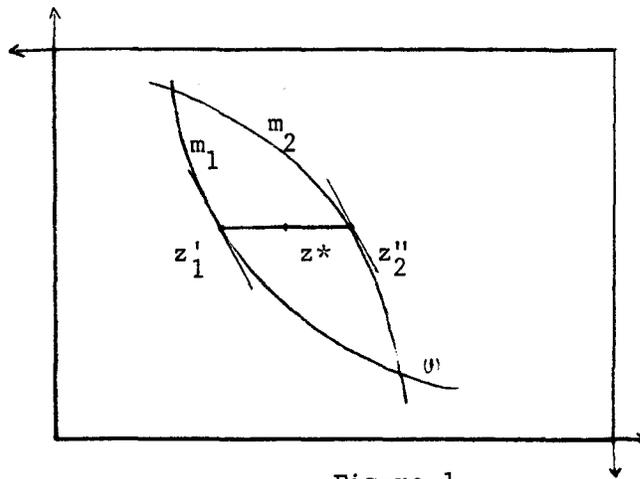


Figure 1

A: We now assume that agent 1 announces his true indifference curve m_1^0 and we determine the sets $A_2(m_1^0)$ and $A_2^*(m_1^0)$ of allocations that agent 2

can make appear or (resp) enforce as value-allocations through some appropriate strategy choice. Formally, $A_2(m_1^0) = \{z \in F(e) \mid z \in V(m_1^0, m_2) , m_2 \in M_2\}$, and $A_2^*(m_1^0) = \{z \in F(e) \mid \{z\} = V(m_1^0, m_2) , m_2 \in M_2\}$.

For each y_2 in $[0, \omega_y]$, we will determine the set $A_2(m_1^0) \cap \{z' \in F(e) \mid y_2' = y_2\}$. Let $y_1 = \omega_y - y_2$. By the assumptions made on the preferences, there exists a unique $x_1' \in R_+$ such that $z_1' = (x_1', y_1) \in m_1^0$. Let $p \in H(m_1^0, z_1')$ be given. If $y_1 > 0$, then $p_1 \neq 0$, otherwise agent 1's preference map could not be generated by horizontal translation of m_1^0 . Let $z_1'' = (\omega_{1x} - \frac{p_2}{p_1} (y_1 - \omega_{1y}) , y_1)$. Given $\lambda \in R$ with $0 \leq \lambda \leq 1$, we define $z_1(\lambda) = (\lambda \omega_{1x} + (1-\lambda)x_1'' , y_1)$; there exists $m_2(\lambda) \in M_2$ such that (a): $z_2(\lambda) \equiv \omega - z_1(\lambda) \in m_2(\lambda)$, (b): $p \in H(m_2(\lambda) , z_2(\lambda))$. Then, denoting $\tilde{z}_1(\lambda) = (\frac{1}{2}(x_1' + x_1(\lambda)) , y_1)$ and $\tilde{z}_2(\lambda) = \omega - \tilde{z}_1(\lambda)$, it is clear that $\tilde{z}(\lambda) \in V(m_1^0, m_2(\lambda))$.

If $y_1 = 0$, and $H(m_1^0, z_1')$ is a singleton with $p_1 = 0$, then no allocation with $y_1 = 0$ can be a value-allocation, no matter what $m_2 \in M_2$ is. If there exists $p \in H(m_1, y_1)$ with $p_1 \neq 0$, the same reasoning as above applies.

When y_2 describes $[0, \omega_y]$, and λ varies between 0 and 1 for each y_2 , $\tilde{z}_2(\lambda)$ describes $A_2(m_1^0)$. Note that the locus of $z_1(0)$ is $h_1(m_1^0)$.

Figure 2 presents two examples to illustrate this characterization of $A_2(m_1^0)$. In example (2a), m_1^0 is smooth (C^2) while m_1^0 is piece-wise linear, with a vertical segment, in example (2b). In (2a), $A_2(m_1^0)$ is represented by the shaded area, boundary included. In (2b), $A_2(m_1^0)$ is represented by the shaded area, boundary included, and the segment $[\beta', \gamma]$.

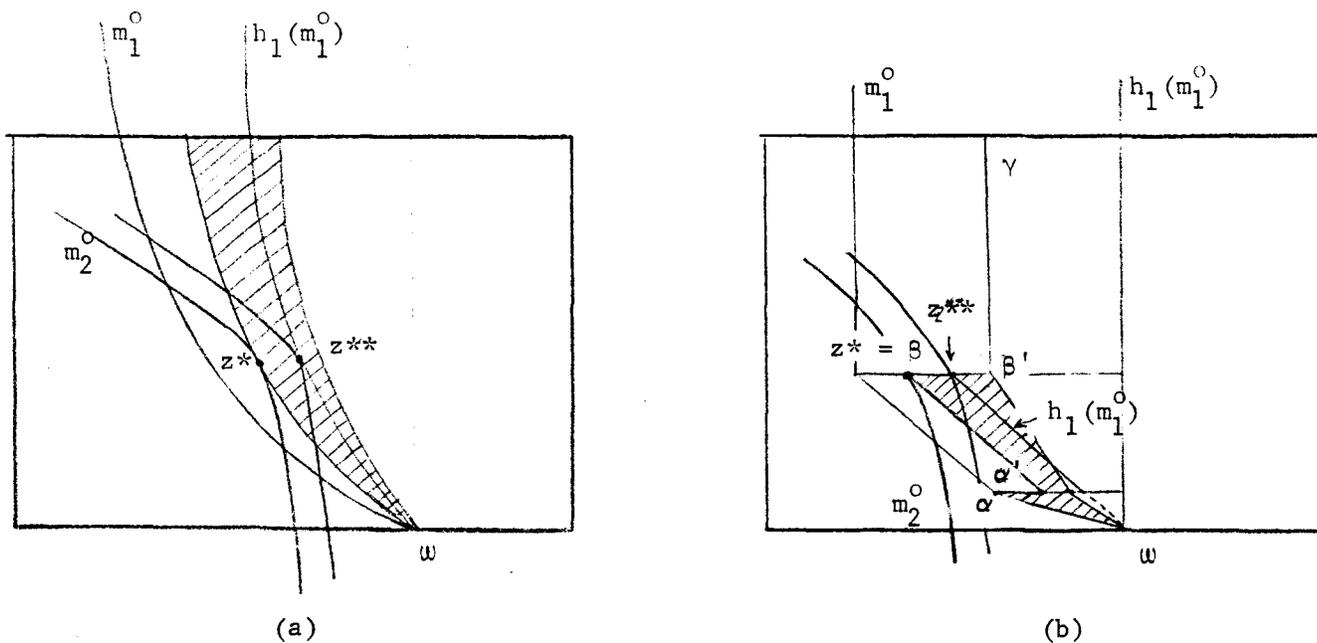


Figure 2

B. For any $y_2 \in [0, w_y]$, and any $0 < \lambda < 1$, $m_2(\lambda)$ can be chosen to satisfy conditions (a) and (b) stated above as well as (c): $m_2(\lambda)$ is strictly convex at $z_2(\lambda)$. This guarantees uniqueness of the value. Every allocation in the interior of $A_2(m_1^0)$ can therefore be enforced as a value-allocation, and belongs to $A_2^*(m_1^0)$. In addition, uniqueness is guaranteed whenever m_1^0 is strictly convex at z_1' . In (2a), m_1^0 is strictly convex everywhere, and $A_2^*(m_1^0) = A_2(m_1^0)$. In (2b), no allocation in $[w, \alpha]$, $[\alpha', \beta]$, and $[\beta', \gamma]$ can be enforced as unique value-allocation by agent 2.

C. Let now z^* be the element of $A_2(m_1^0)$ that agent 2 prefers according to his true preferences. If z^* belongs to $A_2^*(m_1^0)$, as in (2a), agent 2 has an optimal strategy. If z^* belongs to $A_2(m_1^0) \setminus A_2^*(m_1^0)$, agent 2 cannot enforce

z^* as a unique value-allocation, and has therefore no optimal strategy, but z^* can be approximated with any degree of accuracy by some enforceable allocation.

D. Note that by monotonicity of preferences in x , agent 2 is only interested in the allocations belonging to the left boundary of $A_2(m_1^0)$. Any such allocation is in the middle of a horizontal segment $[z_1', z_1(0)]$, where $z_1' \in m_1^0$ and $z_1(0) \in h_1(m_1^0)$. Since by manipulation of the price mechanism, agent 2 can only achieve allocations belonging to $h_1(m_1^0)$, it is clear that optimal manipulation of the value will be more profitable for him since the left boundary of $A_2(m_1^0)$ is to the left of $h_1(m_1^0)$. In (2a) and (2b), optimal manipulation of the price mechanism yields the monopoly allocations z^{**} , to which z^* is preferred. In (2b), if the preferences of agent 2 were drawn in such a way that α would maximize them over $h_1(m_1^0)$, then z^* would be equal to z^{**} . This is the only case where optimal manipulation of the value would not be strictly better than optimal manipulation of the price mechanism.

E. The gain from optimal manipulation of the value is measured by the difference between the true utilities at z^* as defined in D, and $z^0 \in V(m^0)$. The next example shows that this gain may be smaller than the gain correspondingly defined for the price mechanism.

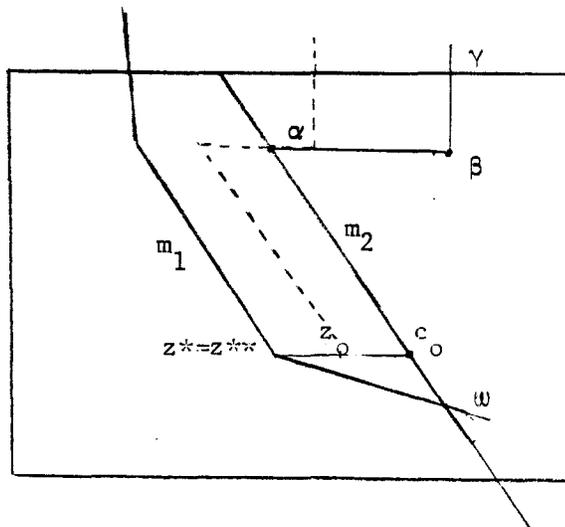


Figure 3

In this example, m_1^0 is piece-wise linear while m_2^0 is linear. $h_1(m_1^0)$ is the broken line $w z^* c_0 \alpha \beta \gamma$. $h_2(m_2^0)$ is the straight line $w z^* c_0$. z_0 (resp z_0) is a truthful competitive allocation (resp value-allocation). There are other such allocations (in both cases) but they are indifferent to c_0 and z_0 according to the true preferences.

Optimal manipulation of the price mechanism yields z^* , while z^* can be approximated by manipulation of the value. Therefore the gain from optimal manipulation of the price mechanism is greater than any gain obtainable by manipulation of the value.

F. In the 2-person, m-commodity case, offer curves would be replaced by "offer surfaces" but all the conclusions would be unaffected. In the n-person, m-commodity case, if agent i is the only one to cheat, the undominated boundary of the set of allocations that he could reach by unilateral manipulation of the value would dominate the net aggregate offer surface he would face, $h(m_{-i}^0)$, but would not bear the same simple geometrical relationship to it as in the 2-person case. The reasoning would not be otherwise affected.

It may be worthwhile pointing out the relationship between the Shapley-value and a scheme devised by Vickrey [12] for the truthful elicitation of demand functions as dominant strategies, in economies with transferable utility in the class E described above. This mechanism requires each of the agents to communicate a demand function on the basis of which certain side-payments are computed, along with the usual Walrasian net trades. The side-payment to each agent can be interpreted as the change in the aggregate welfare of the other n-1 agents caused by his presence. Under this scheme, agent i's final

utility turns out to be equal to the welfare of the grand coalition I minus the welfare of the coalition $I \setminus i$: $x_i^I + v_i(y_i^I) = w^I - w^{I \setminus i}$.

Since the value ϕ_i of agent i is computed as a weighted average of his contributions $w^S - w^{S \setminus i}$ to all the coalitions to which he belongs, and that in the evaluation of each of these terms, announcing one's true utility v_i is a dominant strategy, one could hope that the Shapley-value enjoyed the strong incentive property of the Vickrey mechanism. Theorem 1 shows this not to be the case. The reason is that the "Vickrey-allocations" $((x_i^S, y_i^S), i = 1, \dots, n)$ which achieve the "Vickrey utilities" in each coalition S are not feasible in their respective coalitions. Feasibility is made possible for the Shapley-value through a kind of averaging process in the commodity space that destroys incentives.

The unilateral manipulability of the Shapley-value can also be understood in the light of a theorem due to Hurwicz [1] and establishing that no mechanism selecting individually rational and Pareto-optimal allocations and defined over a sufficiently wide domain is immune to manipulation. The proof of this theorem is in the form of a counter-example involving a Cobb-Douglas economy. Since we are working here with economies having transferable utility, the theorem is therefore not applicable directly. However, adapting its proof to the restricted environment E would present no difficulty.

3. Joint Manipulation

In this section, we go one step further and we assume that every agent attempts to manipulate the value, and does so taking into account the announcements (strategies) of the other agents. In the preceding section, it was shown that unilateral manipulation of the value involved a "flattening"

of preferences, The next lemma shows that equilibria are obtained for lists of "flat" preferences.

Lemma 1: $\forall e \in E, \forall (m, z) \in M \times X, (m, z) \in N(e) \Rightarrow \omega \in P(m)$.

This means that at an equilibrium, there is no gain from trade in the apparent economy.

Proof: Given $(m, z) \in N(e)$, for all i , let $z'_i \in X_i$ such that $z'_i \in m_i$ and $y'_i = y_i$; z'_i exists and is unique. Since $z \in P(m)$, there exists $p \in S^m$ that supports agent i 's indifference surface through z_i at this point, for all i . Since z_i and z'_i differ only in their first coordinates, $p \in H(m_i, z'_i)$. Convexity of preferences implies that $pz'_i \leq p\omega_i$ for all i . We will show that in fact $pz'_i = p\omega_i$ for all i .

Suppose not. Then, for at least one agent, say agent 1, the inequality is strict: $pz'_1 < p\omega_1$. We now explicitly compute the value of an individual as a function of the list m , indicated as a subscript. The utility function that corresponds to the strategy m_i is denoted v_{m_i} . For all $S \subset I$, we have

$$w_m^S = w_x^S + \max \left\{ \sum_{i \in S} v_{m_i}(y_i) \text{ st } \sum_{i \in S} y_i \leq \sum_{i \in S} \omega_i \right\}.$$

The set of maximizers of the above bracket is denoted Y_m^S . It is a subset of $R_+^{|S| (m-1)}$. A typical element of Y_m^S is denoted y_m^S .

Also,

$$x_i + v_{m_i}(y_i) = \varphi_i \equiv \sum_{\substack{S \subset I \\ S \ni i}} k^S [w_m^S - w_m^{S \setminus i}].$$

Let $\varepsilon = p\omega_1 - pz'_1$. First we will show the existence of a strategy \tilde{m}_1 in

M_1 and $\tilde{z} \in V(\tilde{m}_1, m_{-1})$ with $\tilde{z} \succ_{m_1^0} z$. Given $z_1'' = \frac{1}{2}[z_1' + \bar{z}_1]$ where $p\bar{z}_1 = pw_1$ and $\bar{y}_1 = y_1$, let P be a hyperplane with normal p through z_1'' and let \tilde{m}_1 be defined as follows: for each y_1 in R_+^m , let $x_1(m_1)$ and $x_1(P) \in R$ be such that $(x_1(m_1), y_1) \in m_1$, $(x_1(P), y_1) \in P$. Then $\tilde{m}_1 \equiv \{(x_1, y_1) \in R \times R_+^{m-1} \mid x_1 = \max\{x_1(m_1), x_1(P)\} \text{ for each } y_1\}$. If $v_{\tilde{m}_1}$ represents the corresponding utility function, then $v_{m_1}(y_1') - v_{\tilde{m}_1}(y_1') < \epsilon/2$ for all $y_1' \in R_+^m$ and $v_{m_1}(y_1) - v_{\tilde{m}_1}(y_1) = \epsilon/2$. For each coalition S to which agent 1 belongs, and each y^S in Y_m^S , $v_{m_1}(y_1^S) - v_{\tilde{m}_1}(y_1^S)$ measures the amount by which agent 1's contribution would go down if no redistribution of w_y was carried out. In fact, some redistribution will in general be possible, and $y_{(\tilde{m}_1, m_{-1})}^S$ will differ from y_m^S .

$$\begin{aligned} w_m^S - w_{(\tilde{m}_1, m_{-1})}^S &= \sum_{i \in S} v_m(y_{i,m}^S) - [v_{\tilde{m}_1}(y_{1,(\tilde{m}_1, m_{-1})}^S) + \sum_{\substack{i \in S \\ i \neq 1}} v_{m_i}(y_{i,(\tilde{m}_1, m_{-1})}^S)] \equiv \\ &\equiv \sum_{i \in S} v_m(y_{i,m}^S) - [v_{\tilde{m}_1}(y_{1,m}^S) + \sum_{\substack{i \in S \\ i \neq 1}} v_{m_i}(y_{i,m}^S)] \leq \epsilon/2. \end{aligned}$$

Note that $w_m(\{1\}) = w_{(\tilde{m}_1, m_{-1})}(\{1\})$.

It follows that

$$\begin{aligned} \phi_1 - \tilde{\phi}_1 &= \sum_{\substack{S \subset I \\ S \ni 1}} k^S [w_m^S - w_m^{S \setminus \{1\}}] - \sum_{\substack{S \subset I \\ S \ni 1}} k^S [w_{(\tilde{m}_1, m_{-1})}^S - w_{(\tilde{m}_1, m_{-1})}^{S \setminus \{1\}}] \\ &= \sum_{\substack{S \subset I \\ S \ni 1 \\ S \neq \{1\}}} k^S [w_m^S - w_{(\tilde{m}_1, m_{-1})}^S] + k^{\{1\}} [w_m^{\{1\}} - w_{(\tilde{m}_1, m_{-1})}^{\{1\}}] \equiv \end{aligned}$$

$$\cong \sum_{\substack{S \subset I \\ S \neq \{1\}}} k^S \epsilon/2 = [1 - k^{\{1\}}] \epsilon/2 .$$

Since $p \in H(\tilde{m}_1, z_1'')$, there exists $\tilde{z} \in V(\tilde{m}_1, m_{-1})$ such that $\tilde{y}_1 = y_1$.

Since $v_{\tilde{m}_1}(y_1) = v_{m_1}(y_1) - \epsilon/2$, it follows that $x_1 + v_{\tilde{m}_1}(y_1) = x_1 + v_{m_1}(y_1) - \epsilon/2 = \varphi_1 - \epsilon/2 = \tilde{\varphi}_1 - k^{\{1\}} \epsilon/2 < \tilde{\varphi}_1$. Therefore, $(\tilde{x}_1, \tilde{y}_1)$ will achieve agent i's value $\tilde{\varphi}_1$ only if $\tilde{x}_1 > x_1$, and consequently $\tilde{z} \succ_{m_1^0} z$.

If m_i is not strictly convex at z_i for one $i \neq 1$, there may be value-allocations different from \tilde{z} that are not as good as z for agent i. In order to ensure uniqueness of the first component \tilde{z}_1 of \tilde{z} , it is sufficient that the strategy \tilde{m}_1 announced by agent 1 be strictly convex at z_1'' . To achieve this, in the definition of \tilde{m}_1 , we replace P by P' , a strictly convex approximation to P defined as follows: given y_1 in R_+^m , let $x_1(P)$ and $x_1(P')$ be such that $(x_1(P), y_1) \in P$, $(x_1(P'), y_1) \in P'$, and $x_1(P) - x_1(P') < \frac{(1 - k^{\{1\}})\epsilon}{4n!}$. This leads to an adjustment in the computation of agent i's value that is no greater than $\frac{(1 - k^{\{1\}})\epsilon}{4}$. Now $\tilde{z} \in P(\tilde{m}_1, m_{-1}) = \tilde{y}_1 = y_1$, and by the same argument as above $\tilde{z} \in V(\tilde{m}_1, m_{-1}) \Rightarrow \tilde{z} \succ_{m_1^0} z$. This concludes the proof of the lemma.

Let $d_{m_i}: S^m \rightarrow R^m$ and $cd_{m_i}: S^m \rightarrow R^m$ be agent i's demand and respectively "constrained" demand correspondences associated to the preference map derived from m_i . The constraint in the definition of cd_{m_i} refers to the requirement that the agent's consumption should not exceed the aggregate endowment of the economy.

Lemma 2: $\forall (m, z) \in M \times X$, $(m, z) \in N(e) \Rightarrow \forall i \in I, h(m_{-i}) \cap W(m_i^0, z_i) = \emptyset$.

The net offer surface faced by an agent does not intersect his strict upper contour set at an equilibrium.

Proof: Let $(m, z) \in M \times X$ be given and suppose, by way of contradiction, that $\exists i \in I, \exists p \in S^m, [\exists z'_j \in X_j, \forall j \in I]$, s.t. (a): $[z'_j - \omega_j \in d_{m_j}(p), \forall j \in I, j \neq i]$, (b): $\Sigma(z_j - \omega_j) = 0$ and (c): $z'_i \succ_{m_i^0} z_i$. Then, denoting by a_1 the first unit vector, for $\epsilon > 0$ small enough, $z''_i = z'_i - \epsilon a_1$ satisfies $z''_i \succ_{m_i^0} z_i$. Let now $m'_i \in M_i$ be such that (a): $z''_i \in m'_i$, (b): $p \in H(m'_i, z''_i)$ and (c): m'_i be strictly convex at z''_i . Then, $\tilde{z} \in V(m'_i, m_{-i}) \Rightarrow \tilde{z}_i = z''_i + \eta \epsilon$ with $\eta \geq 0$ and consequently $\tilde{z} \succ_{m_i^0} z$, which implies $(m, z) \notin N(e)$, and proves the lemma.

Lemma 3: $\forall (m, z) \in M \times X, \forall i \in I, \forall \lambda \in R_+^n, (m, z) \in N(e)$ and $[0 \leq \lambda_j \leq 1, \forall j \in I, j \neq i] \Rightarrow \omega_i - \sum_{j \neq i} \lambda_j (z_j - \omega_j) \notin W(m_i^0, z_i)$.

Proof: Let $(m, z) \in N(e)$ and $i \in I$ be given. Let $p \in S^m$ be a price vector supporting the optimal allocation z ; p exists since $z \in V(m) \subset P(m)$. By lemma 1, $z_j \in m_j$ and $p z_j = p \omega_j$ and consequently $z_j - \omega_j \in d_{m_j}(p)$ for all $j \in I$. By convexity of preferences, for all $j \in I, j \neq i$, and for all $\lambda_j \in R_+, \lambda_j \leq 1, \lambda_j (z_j - \omega_j) \in d_{m_j}(p)$. Then $\omega_i - \sum_{j \neq i} \lambda_j (z_j - \omega_j) \in h(m_{-i})$. By lemma 2, $(m, z) \in N(e) \Rightarrow \omega_i - \sum_{j \neq i} \lambda_j (z_j - \omega_j) \notin W(m_i^0, z_i)$ which proves lemma 3.

Given e in E , let $\Lambda(e)$ be defined by

$$\Lambda(e) = \{z \in F(e) \mid \forall i, z_i = \omega_i + k_i t_i, \text{ with } 0 \leq k_i \leq 1, t_i \in d_{m_i^0}(p_i) \text{ for } p_i \in S^m; \forall i, p z_i = p \omega_i \text{ for } p \in S^m.\}$$

In the 2-person, 2-commodity case, $\Lambda(e)$ is the area delimited by the true offer curves. It is lens shaped under a smoothness assumption on the preferences.

Then we can prove:

Corollary 1: Suppose $n = 2$, $m = 2$. For all e in E , $N(e) = \Lambda(e)$.

Claim 1: $\forall (m, z) \in M \times X$, $(m, z) \in N(e) \Rightarrow z \in \Lambda(e)$.

Proof: Let $(m, z) \in M \times X$ be given, and assume by way of contradiction, that $z \notin \Lambda(e)$. Then, for at least one agent, say agent 1, at the price p defined by $[w, z]$, for all t_1 in $d_1^0(p)$, $\|t_1\| < \|z_1 - w_1\|$. Therefore, for $0 < \lambda_2 < 1$, $w_1 - \lambda_2(z_2 - w_2) \in W(m_1^0, z_1) \neq \emptyset$, in contradiction with Lemma 3.

Claim 2: $\forall z \in X$, $z \in \Lambda(e) \Rightarrow \exists m \in M$ s.t. $(m, z) \in N(e)$.

Proof: Given $z \in \Lambda(e)$, let P_1 and P_2 be lines of support to the agents' true indifference curves through z at that point with p_1 and $p_2 \in S^2$ as non-negative normals. Since $z \in \Lambda(e)$, $p_1 z_1 \geq p_1 w_1$ and $p_2 z_2 \geq p_2 w_2$. Let then $m_1 \in M_1$ be such that (a): $m_1 \supset [w_1, z_1]$ and (b): $p_2 \in H(m_1, z_1)$; m_2 is similarly defined. Clearly $z \in V(m_1, m_2)$. Since $z' \succ_{m_1} z \Rightarrow \omega_2 \sim_{m_2} z \not\prec_{m_2} z'$ and since the value is individually rational, agent 1 has no better strategy against m_2 than m_1 (similarly for agent 2). This proves the claim.

Corollary 1 is a direct consequence of lemmas 1 and 2.

This result should be compared to the one obtained by Hurwicz [3] in his study of the manipulability of the price-mechanism: to the Walrasian performance correspondence can be associated a quasi-game by defining strategy spaces to be spaces of offer curves, and the outcome correspondence to be the Walrasian correspondence. Hurwicz showed that if offer curves are required to be consistent with smooth (C^2) strictly convex preferences, and if one eliminates from consideration strategy pairs yielding multiplicities of Walrasian allocations,

the equilibrium set of the generalized game so defined would coincide with the interior of $\Lambda(e)$. The boundary of the lens is added to the equilibrium set by removing the smoothness requirement. If multiplicities are dealt with in the way they are being analyzed in the present paper, one reaches the conclusion that the equilibrium set of the Walrasian manipulation quasi-game coincides with $\Lambda(e)$.

It can be argued that the allocations on the boundary of $\Lambda(e)$ are more likely equilibrium allocations. Given $(m, z) \in N(e)$ such that $z \in \text{int } \Lambda(e)$, by the previous lemmas, we know that $m_i \supset [w_i, z_i]$, $i = 1, 2$. Let p be the price implicitly defined by $[w, z]$. If $z'_i \in d_{m_i^0}(p)$, then $\|z'_i - w_i\| > \|z_i - w_i\|$. Assume now that the equilibrium (m, z) has somehow been reached and that both agents know it to be an equilibrium. For $i = 1, 2$, consider the strategy $m'_i \in M_i$ defined by (a): $m'_i \supset [w_i, z'_i]$ with $z'_i \in d_{m_i^0}(p)$, (b): m'_i is extended from z'_i on by the agent's true indifference curve through that point. By changing his strategy from m_i to m'_i , agent i gives the other agent opportunities for further trading at prices p , with the knowledge that no rational response of the other agent will lead to value allocations to which he (agent i) would prefer z . If both agents perform this change, a Pareto-improving move will take place that will lead the economy to the boundary of $\Lambda(e)$. It is because of this special stability property that allocations on the boundary of $\Lambda(e)$ are more prominent equilibrium allocations.

Corollary 2: Let $n = 2$. For all e in E , $N(e) = \Lambda(e)$.

Proof: The proof that no allocation outside of $\Lambda(e)$ is an equilibrium allocation is identical to that of Claim 1 of Corollary 1. To prove the analogue of Claim 2, let z in $\Lambda(e)$ be given, P_1 and P_2 as in the proof of

Claim 2. Then let $m_1 \in M_1$ satisfy (a): $m_1 \supset [\omega_1, z_1]$, (b): $p_1 \in H(m_1, z_1)$, (c): $p_2 \in H(m_1, z_1)$. For m_2 similarly defined, we have $z \in V(m)$. The proof ends as that of Claim 2*.

The set of equilibrium allocations of the Walrasian mechanism was proved by Otani and Sicilian [] to coincide with $\Lambda(e)$ for any number of commodities.

Corollary 3: Let $(m, z) \in N(e)$ be given, and let $p \in S^m$ be a price supporting z . Suppose that for all i in I the cone generated by the list $\{z_j - \omega_j, j \neq i\}$ spans the linear space defined by p . Then $z \in CW(e)$.

Proof: First note that p exists since $z \in V(m) \subset P(m)$. The conclusion is a straightforward consequence of lemma 3.

Corollary 4: Assume $n > m$, and let $(m, z) \in N(e)$ with $p \in S^m$ as a price supporting z . Suppose that there exists a subset J of I such that the cone generated by the list $\{z_j - \omega_j, j \in J\}$ spans the linear space defined by p . Then, for at least $n-m$ agents, $z_i \in cd_i^0(p)$.

Proof: Straightforward consequence of lemma 3, since in an m -dimensional commodity space, m agents will suffice for the spanning condition to be satisfied.

In a large economy, if n is much greater than m , almost all agents will be getting their constrained Walrasian net trades, and equilibrium allocations could then be said to be approximate constrained Walrasian allocations.

*While revising this paper, the following results, independently obtained by Sobel [9], were communicated to me: 1) with or without transferable utility competitive allocations are always equilibrium allocations of the manipulation game associated with various solution concepts in a class that contains the Nash and the Raiffa-Kalai-Smorodinsky solutions. This result is in perfect agreement with Corollary 2 since the Nash solution coincides with the Shapley-value in 2-person economies, and competitive allocations belong to $\Lambda(e)$. In addition, 2): under a smoothness assumption, all and only the constrained Walrasian allocations are equilibrium allocations; the counterpart of this result for the value is formulated in Theorems 2 and 3 of the present paper.

If the spanning condition is not satisfied, none of the agents may be getting his constrained Walrasian net trade relative to the prices p . However, let B be the subspace of the largest dimensionality spanned by $\{z_j - w_j, j \in J\}$ for some subset J of I . A constrained B -Walrasian bundle for an agent is obtained by maximization of his preference subject to the requirement that his net trade belong to B . Then we have that if k_B is the dimensionality of B , J need not contain more than k_B agents and $n - k_B$ agents will be getting their constrained B -Walrasian net trades.

This comment is illustrated by the following examples:

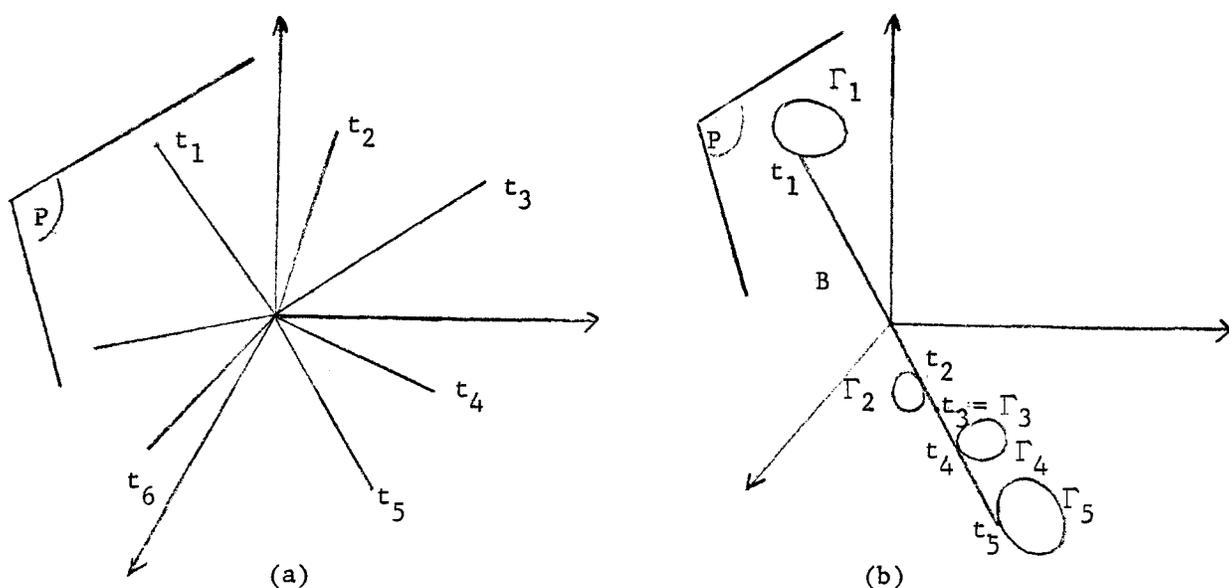


Figure 4

In a 3-dimensional space, we let $(m, z) \in N(e)$ be given and p be a supporting price for z . $P = \{z \in R^3 \mid pz = 0\}$. $\Gamma_i = \{t'_i \in R^3 \mid w_i + t'_i \sim_{m_i}^0 w_i + t_i\} \cap P$.

In (4a), $\{t_i\} = \Gamma_i$ for all i and $z \in CW(e)$.

In (4b), $\{t_i\} \neq \Gamma_i$ for all i except $i = 3$. B is a one-dimensional subspace of the 3-dimensional subspace P . Note that all of the net trades except t_1 are utility maximizing in the subspace B , although z is not in $CW(e)$.

Corollary 5: For all e in E , $N(e) \subset \Lambda(e)$.

Proof: It is a direct consequence of lemma 3.

We will now consider the quasi-game (M^S, V) , where M^S is the subset of M corresponding to C^2 utility functions. This smoothness requirement can be interpreted in two ways: first, one may actually be working in the subset E^S of E where agents actually have smooth preferences; then (M^S, V) remains a direct quasi-game since strategy spaces still coincide with spaces of possible characteristics. Second, the quasi-game (M^S, V) can be seen as an abstract quasi-game defined on E : truthful behavior is not possible anymore for an agent whose preferences are not smooth. In both cases, however, one can prove that the set of equilibrium allocation shrinks to the set of constrained Walrasian allocations. The constrained Walrasian performance correspondence, denoted CW , is an extension of the Walrasian performance correspondence introduced by Hurwicz, Maskin, and Postlewaite [4], which coincides with it in the interior of the feasible space, and always selects individually rational and Pareto-optimal outcomes. The constrained demand correspondence of an agent is obtained by maximization of his preferences subject to his budget constraint and the requirement that his consumption should not exceed the aggregate amount available in the economy. (It was introduced just before Lemma 2.) CW is obtained from the list of constrained demand correspondences as W is obtained from the list of demand correspondences.

Theorem 2: For all e in E , the set of equilibrium allocations of the quasi-game (M^S, V) is equal to $CW(e)$.

Proof: Given e in E , let $N^S(e) \subset M^S \times X$ be the set of equilibrium pairs of (M^S, V) . First we observe that lemmas 1 and 2 apply as well to (M^S, V) as to (M, V) . Let now $(m, z) \in N^S(e)$ be given. By lemma 1, there exists $p \in S^m$ such that, for all i , m_i has a unique hyperplane of support with normal p at z_i . This hyperplane of support contains w_i . $z_i = w_i - \sum_{j \neq i} (z_j - w_j)$ belongs to $h(m_{-i})$. By the smoothness requirement, $h(m_{-i})$ admits of a unique supporting hyperplane at z_i , with normal p .

We now claim that (p, z) is a constrained competitive equilibrium for e . If not, the set $\{z'_i \in X_i \mid z'_i \succ_{m_i} z_i\} \cap \{z'_i \in X_i \mid z'_i \preceq w\}$ would not be supported at z_i by a hyperplane with normal p . Then $h(m_{-i})$ would be transversal to it, and the condition of lemma 2 would be violated.

Conversely, given $z \in CW(e)$, let p be a supporting price for z , and let agent $i \in I$ announce m_i such that (a): $z_i \in m_i$, (b): $p \in H(m_i, z_i)$. Then $(m, z) \in N^S(e)$.

This result can be generalized to economies without transferable utility.

Theorem 3: Let E' be the class of pure-exchange economies with n agents and m commodities. Each agent is characterized by a list (X_i, w_i, \succsim_i) , where $X_i = R_+^m$, $w_i \in R_+^n$, and \succsim_i is a convex, continuous preference relation, strictly monotonic in the first commodity. Let (M^S, Φ) be the abstract quasi-game defined in the preceding theorem. For every e in E' , the set of equilibrium allocations of this quasi-game coincides with $CW(e)$.

Proof: It is identical to that of Theorem 2.

Remark: Since $M_i^S \subset M_i$ for all i , and any pair (m, z) which is an equilibrium of the quasi-game (M^S, V) is such that ω belongs in $P(m)$, it follows that, for all i , there is no m_i' in M_i which is better than m_i . Consequently, $CW(e) \subset N(e)$ for all e .

Instead of the above smoothness requirement, another redefinition of the quasi-game is now proposed that has the same consequences. For simplicity, we will take ω_i to be in $\text{int } X_i$ for all i .

Let now \bar{M}_i be the set of elements m_i of M_i such that (a): $H(m_i, \omega_i)$ be a singleton, p_i , and (b): $G_i(m_i) \equiv m_i \cap \{z_i' \in X_i \mid p_i z_i' = p_i \omega_i\}$ spans a subspace of dimension $m - 1$. Define $\bar{M} = \bar{M}_1 \times \dots \times \bar{M}_n$.

Denoting the relative interior of a set A as $\text{ri } A$, the value performance correspondence is modified as follows:

$$\bar{V}(m) = \begin{cases} V(m) & \text{if } \omega \notin P(m) \\ \{z \in V(m) \mid \forall i, z_i \in \text{ri } G_i(m_i) \text{ or } z \in CW(m) \setminus W(m)\} & \text{if } \omega \in P(m) \end{cases} .$$

Theorem 4: For all e in E , the set of equilibria of the quasi-game (\bar{M}, \bar{V}) coincides with $CW(e)$.

Proof: One could prove, as in lemma 1, that $(m, z) \in \bar{M} \times X$ is an equilibrium pair of (\bar{M}, \bar{V}) only if $\omega \in P(m)$. Since $z \in \bar{V}(m) \Rightarrow z \in P(m)$, there exists $p \in S^m$ such that $p \in H(m_i, z_i)$ for all i . By the definition of \bar{M} , $p = p_i = p_j$ for all i and j in I . Since $\omega \in P(m)$, $z_i \in \text{ri } G_i(m_i)$ or $z \in CW(m) \setminus W(m)$. In addition, $z_i \in \{z_i' \in X_i \mid p z_i' = p z_i\}$. Assume now that $z \notin CW(e)$. Then, the set $\{z_i' \in X_i \mid z_i' \succ_{m_i^0} z_i\} \cap \{z_i' \in X_i \mid z_i' \preceq \omega\}$ would be supported at z_i only by hyperplanes with normal $p' \neq p$. And $h(m_{-i}) \cap \{z_i' \in X_i \mid z_i' \succ_{m_i^0} z_i\} \cap \{z_i' \in X_i \mid z_i' \preceq \omega\}$ would be non-

empty, in contradiction with lemma 2 (the proof of which is omitted).

Theorem 5: Same statement as Theorem 3, with (M^S, V) replaced by (\bar{M}, \bar{V}) .

Proof: Same as above theorem.

5. Concluding Comment

A performance correspondence Φ defined on E is implementable by a game (N, h) with strategy space $N = N_1 \times \dots \times N_n$, and outcome function $h: N \rightarrow X$, iff, for all e in E , the set of Nash equilibrium allocations of (N, h) coincides with $\Phi(e)$. Maskin [5] has shown that a performance correspondence Φ is implementable by a game only if it is monotonic:

Definition: $\Phi: E \rightarrow X$ is monotonic iff:

$$\forall e, e' \in E, \forall z \in \Phi(e), [\forall z' \in X, \forall i \in I, z \succ_{m_i} z' \Rightarrow z \succ_{m_i} z'] \Rightarrow z \in \Phi(e').$$

Similarly, a performance correspondence Φ defined on E is implementable by a quasi-game (N, h) with strategy space $N = N_1 \times \dots \times N_n$ and outcome correspondence $h: N \rightarrow X$, iff, for all e in E , the set of equilibrium allocations of (N, h) , as defined on pg. 6, coincides with $\Phi(e)$.

Lemma 4: A performance correspondence $\Phi: E \rightarrow X$ is implementable by a quasi-game only if it is monotonic.

Proof: It is virtually identical to that of Maskin's Theorem, and we will omit it.

The constrained Walrasian correspondence CW is monotonic, and can indeed be implemented by the quasi-game (M^S, V) as we saw above. On the other hand, we know that neither CW nor V can be implemented by their own quasi-games. The question arises as to whether V can be implemented by some other quasi-game. A negative answer follows from Lemma 4 and the following:

Lemma 5: The value performance correspondence is not monotonic.

Proof: Given a preference relation \succsim_i for agent $i \in I$, and z_i in X_i , we will say that \succsim'_i is obtained by a monotonic transformation of \succsim_i at z_i iff

$$\forall z'_i \in X_i, z_i \succsim_i z'_i \Rightarrow z_i \succsim'_i z'_i.$$

In Figure 5, the solid lines represent the preferences \succsim_1 and \succsim_2 of the two agents of some economy e , and $\{z\} = V(m)$. z is the middle of $[z'_1, z''_2]$. The dotted lines represent a preference relation \succsim'_1 for agent 1. It is easy to check that \succsim'_1 is obtained by a monotonic transformation of \succsim_1 at z_1 . It is also clear that $\{\tilde{z}\} = V(m'_1, m_2)$ is different from z . Therefore, $V: E \rightarrow X$ is not monotonic.

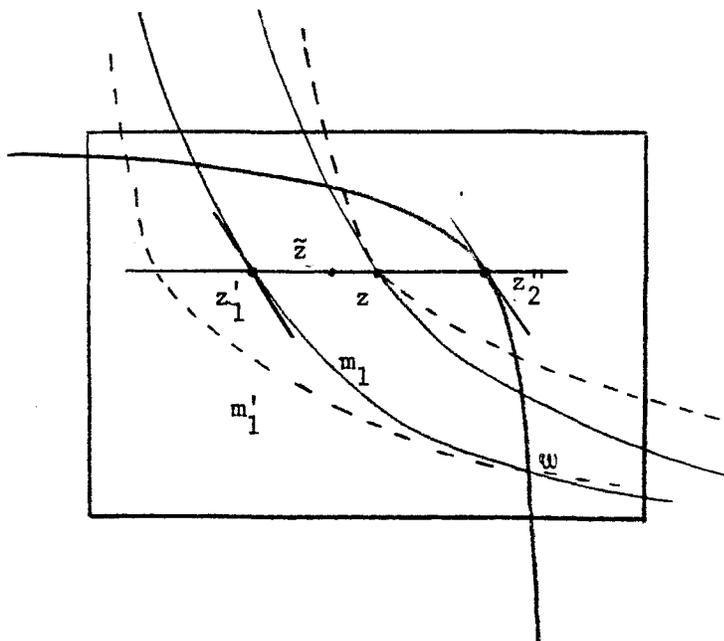


Figure 5

Theorem 6: The Shapley-value performance correspondence is not implementable by a quasi-game.

Proof: It is a straightforward consequence of Lemmas 4 and 5.

REFERENCES

- [1] Hurwicz, L., "On Informationally Decentralized Systems," Chapter 14, Decision and Organization, C.B. McGuire and R. Radner, (eds.), North-Holland, Amsterdam, 1972.
- [2] Hurwicz, L., "Outcome Functions Yielding Walrasian and Lindhal Allocations at Nash Equilibrium Points for Two or More Agents," Review of Economic Studies, 46 (1979), pp. 217-225.
- [3] Hurwicz, L., "On the Interaction Between Information and Incentives in Organizations," in K. Krippendorff, ed., Communication and Control in Society, New York, NY, Scientific Publishers, Inc., 1978.
- [4] Hurwicz, L., E. Maskin, and A. Postlewaite, "Implementation with Unknown Endowments," mimeograph, June 1979.
- [5] Maskin, E., "Nash-Equilibrium and Welfare-Optimality," Mathematics of Operations Research, forthcoming.
- [6] Otani, Y. and J. Sicilian, "Equilibrium of Walras Preferences Games," University of Kansas mimeo, October 1978.
- [7] Shapley, L., "A Value for n-person Games," in Contributions to the Theory of Games, Kuhn and Tucker, eds. 1953.
- [8] Shapley, L. and M. Shubik, "Pure Competition, Coalitional Power, and Fair Division," International Economic Review, 10 (1969), pp. 337-362.
- [9] Sobel, J., "Distortion of Utilities and the Bargaining Problem," University of San Diego, Discussion Paper 79-29.
- [10] Thomson, W., "The Equilibrium Allocations of Walras and Lindhal Manipulation Games," mimeograph, June 1979.
- [11] Thomson, W., "On the Manipulability of Resource Allocation Mechanisms Designed to Achieve Individually-Rational and Pareto-Optimal Outcomes," mimeograph, September 1979.
- [12] Vickrey, W., "Counterspeculation, Auctions, and Competitive Sealed Tenders," The Journal of Finance, May 1961.