Cowles Foundation Paper 793

Soc Choice Welfare (1991) 8 247–254

Social Choice and Welfare

Inefficiency of strategy-proof allocation mechanisms in pure exchange economies

Lin Zhou*

Cowles Foundation, Box 2125, Yale University, New Haven, CT 06520, USA

Received July 18, 1990 / Accepted February 15, 1991

Abstract. In this paper I prove that in the standard model of $2 \times n\,(n \geqq 2)$ pure exchange economies there is no allocation mechanism that is efficient, non-inversely-dictatorial, and strategy-proof. This strengthens two previous results on this subject by Hurwicz and by Dasgupta, Hammond, and Maskin.

1. Introduction

It has long been recognized that most political and economic institutions are vulnerable to their participants' strategic manipulation. In voting theory Gibbard [3] and Satterthwaite [6] proved that any voting scheme is manipulable unless it is dictatorial. In economic theory many authors have established various results of the same nature. But most results in economic models lack the generality and elegance of the original Gibbard-Satterthwaite theorem. It was only recently that Zhou [8] adopted an approach by Barbera and Peleg [1] to derive in the standard model of pure public good economies a result that is parallel to the Gibbard-Satterthwaite theorem in voting theory. The purpose of this paper is to continue the investigation of the same issue in the standard model of pure exchange economies.

There is a long list of previous contributions on this subject. Two of them are particularly relevant to our discussion here. Hurwicz [4] was the first author to consider formally the manipulability of an allocation mechanism in pure exchange economies. He proved that any efficient and individually rational allocation mechanism is manipulable in two-agent, two-good, pure exchange economies in which both agents' preferences are continuous, strictly convex, and increasing, and both agents' initial endowments are non-zero. A later result of Dasgupta, Hammond, and Maskin [2] replaces the requirement of individual rationality in Hurwicz's result by that of nondictatorship but allows discontinuous individual preferences

* I thank Alvin Klevorick, William Thomson, and an anonymous referee of the journal for many helpful comments.

in an essential way. As pointed out by Hurwicz and Walker [5], these results "thus do not cover situations in which preferences are continuous and there are more than two persons or more than two goods." Although many economists expect that there is a general result free of these seemingly unnecessary restrictions, its exact formulation, not to mention the proof, has remained an open question.

In this paper I prove a result that generalizes substantially the above-mentioned conclusions. More precisely, I show that any efficient and non-inversely-dictatorial allocation mechanism is manipulable for pure exchange economies with two agents and arbitrarily many goods in which both agents' utility functions are continuous, strictly concave, and increasing. Furthermore, this result lends strong support to the conjecture that a similar impossibility can be proved for many-agent economies, even though I still do not have a general proof.

Some necessary basic definitions and the formal statement of the main result are contained in Sect. 2. A proof of the result is given in Sect. 3. Section 4 concludes the paper with a brief discussion on possible extensions of the result to pure exchange economies with arbitrarily many agents and to private economies with production.

2. The model and the main theorem

We begin with the description of $2 \times n\,(n \geqq 2)$ pure exchange economies. There are two *agents* and *n private goods.* Consumption sets for the agents are $X_1 = X_2 = R^n_+$.[1] Each agent has a *utility function* u_i over R^n_+, which is drawn from U, the space of all continuous, strictly quasi-concave, and increasing utility functions.[2] Society has an *aggregate endowment* $\omega \in R^n_{++}$, which is fixed throughout ($\omega = \omega_1 + \omega_2$ if endowments are specified individually). An *allocation* is a vector $(x, \omega - x) \in R^n_+ \times R^n_+$, or equivalently a vector $x \in R^n_+$ that satisfies $x \leqq \omega$, which represents agent 1's share of the endowment.

A *direct allocation mechanism*, or simply a *mechanism*, F is a function from $U \times U$ to R^n_+ such that $F(u_1, u_2) \leqq \omega$.[3]

A mechanism F is *efficient* if for any $(u_1, u_2) \in U \times U$, and any allocation y, we have $u_1(y) > u_1(F(u_1, u_2))$ implies $u_2(\omega - F(u_1, u_2)) > u_2(\omega - y)$.

A mechanism F is *inversely dictatorial* if either $F \equiv \mathbf{0}$ (in this case agent 1 is the inverse dictator) or $F \equiv \omega$ (in this case agent 2 is the inverse dictator).

A mechanism F is *strategy-proof* if for any pair (u_1, u_2), and $(\hat{u}_1, \hat{u}_2) \in U \times U$, we have both $u_1(F(u_1, u_2)) \geqq u_1(F(\hat{u}_1, u_2))$ and $u_2(\omega - F(u_1, u_2)) \geqq u_2(\omega - F(u_1, \hat{u}_2))$.

[1] Conventional notation for vector inequalities: Assume x and $y \in R^n$ $x \leqq y$ means $x_i \leqq y_i$ for all i, $x \leq y$ means $x \leqq y$ and $x_i < y_i$ for at least one i, and $x \ll y$ means $x_i < y_i$ for all i.

[2] A utility function $u(\cdot)$ is increasing if $u(x) > u(y)$ for all $x \geq y$ An immediate corollary is that $p \gg 0$ for any subgradient vector p of u at any interior point of R^n_+.

[3] We only consider direct mechanisms in our formal presentation. However, readers should bear in mind that the analysis on direct mechanisms that are strategy-proof extends, via the well-known revelation principle, to general mechanisms that always have dominant strategy equilibria. Another related point. Since we are interested in mechanisms that have dominant strategy equilibria, a direct mechanism is thus defined as a single-valued function instead of a multi-valued correspondence The latter is more appropriate when one considers implementation in Nash equilibrium

Theorem 1. *There is no mechanism F for* $2 \times n$ $(n \geqq 2)$ *pure exchange economies that is efficient, non-inversely-dictatorial, and strategy-proof.*

Theorem 1 is tight. We leave it to readers to construct various examples of mechanisms that satisfy two of the conditions in Theorem 1.

Theorem 1 improves Hurwicz's result by replacing the condition of individual rationality with the much weaker non-inversely-dictatorial condition. First, the former condition requires at least that $\mathbf{0} \leqq F(u_1, u_2) \leqq \omega$ for all (u_1, u_2), while the latter only requires that $\mathbf{0} \leqq F(u_1, u_2)$ for some (u_1, u_2), and $F(v_1, v_2) \leqq \omega$ for some (v_1, v_2). Second, in situations where individual initial endowments are not well specified or are intentionally ignored for a more equitable distribution, the condition of individual rationality is not appropriate. Another improvement of Theorem 1 is that it frees the dimensionality of the commodity space.

Our theorem also improves Dasgupta-Hammond-Maskin's result by restoring continuity of individual preferences. This is significant not only because continuity of preferences is an essential feature of the standard model of pure exchange economies, but also because, as suggested by Satterthwaite and Sonnenschein [7], when discontinuous preferences are allowed "the set of [non-dictatorial] Pareto efficient single-valued social choice rules [whether strategy-proof or not] is empty." Another difference is that I use the condition of no-inverse-dictatorship in Theorem 1 rather than the no-dictatorship condition that Dasgupta-Hammond-Maskin impose. Even though the two conditions are equivalent in two-agent economies, they are different in economies of three or more agents (an agent i in a general many-agent economy is an inverse dictator if he always gets nothing). While the Dasgupta-Hammond-Maskin result is not true in economies of three or more agents (see Sect. 4.1), Theorem 1 constitutes a well-posed conjecture to which no counterexample has been found yet.

3. A proof of Theorem 1

In the proof of Theorem 1, we will use some properties that utility functions in U possess. Although being very intuitive, they do not seem to have appeared before. For completeness of our analysis we present them formally in three lemmas and provide proofs when necessary.

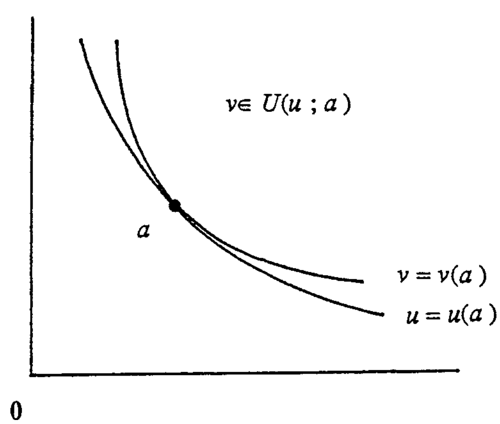

Fig. 1

For a utility function u, and a vector $a \in R^n_+$, let $U(u;a)$ denote the set of all utility functions v in U that satisfy $[v(x) \geqq v(a),\ \mathbf{0} \leqq x \leqq \omega$, and $x \neq a]$ imply $[u(x) > u(a)]$. (See Fig. 1 for an illustration of a utility function $v \in U(u;a)$.)

Lemma 1. *$U(u;a)$ is nonempty for any $u \in U$, and $\mathbf{0} \ll a \ll \omega$.*

Proof. Notice that the lemma is not true if U is the set of quasi-concave and nondecreasing functions. An obvious counterexample is the Leontief utility function. Accordingly, the Leontief function cannot be used in our proof as a desired v. However, we can smooth it to get a desired $v \in U(u;a)$. Let

$$f_i^m(x) = \ln\left(1 + \frac{x_i - a_i}{\|\omega\|}\right) + \frac{1}{\mathrm{m}} \sum_{j \neq i} \ln\left(1 + \frac{x_j - a_j}{\|\omega\|}\right), \quad \text{for each } i .$$

We then define f^m by

$$f^m(x) = \min_{1 \leqq i \leqq n} \{f_i^m(x)\} .$$

By definition, $f^m \in U$. We claim that $f^m \in U(u;\mathrm{a})$ for large enough m. Suppose it is not true. Then for each m there is an x^m such that: (i) $x^m \neq a$; (ii) $\mathbf{0} \leqq x^m \leqq \omega$; (iii) $f^m(x^m) \geqq f^m(a)$; and (iv) $u(x^m) \leqq u(a)$.

Since $f^m(a) = f_i^m(a)$ for each i, (iii) implies $f_i^m(x^m) \geqq f_i^m(a)$ for each i. Because $t \geqq \ln(1+t)$ for all $t > -1$, $f_i^m(x^m) \geqq f_i^m(a)$ further leads to:

$$(x_i^m - a_i) + \frac{1}{m} \sum_{j \neq i} (x_j^m - a_j) \geqq 0 , \quad \text{for each } i . \tag{iii*}$$

Since $\mathbf{0} \ll a \ll \omega$, let $r > 0$ be such that $\mathbf{0} \ll y \leqq \omega$ for all $\|y - a\| \leqq r$. If $\|x^m - a\| < r$, then we replace it by $y^m = a + t^m(x^m - a)$ that satisfies $t^m > 1$ and $\|y^m - a\| = r$. By the definition of r, y^m satisfies (ii). It is also direct to see that y^m satisfies (iii*). Because u is strictly quasi-concave and $u(x^m) \leqq u(a)$, we have $u(y^m) \leqq u(a)$. Hence, we can assume without loss of generality that all x^m satisfy: (i*) $\|x - a\| \geqq r$, (ii), (iii*), and (iv).

Now let m go to infinity. The sequence $\{x^m\}$ (or a subsequence of it) approaches some x by (ii). $x \neq a$ by (i*), and $x \geq a$ by (iii*). Finally, $u(x) \leqq u(a)$ by (iv). But this contradicts the assumption that u is increasing. Q.E.D.

Let $C(u_1, u_2)$ denote the contract curve of two utility functions u_1 and u_2. A continuous curve C from the origin to ω is a continuous mapping from $[0, 1]$ to R^n_+ such that $c(0) = \mathbf{0}$ and $c(1) = \omega$. It is increasing if $c(t) \gg c(t')$ for all t, t' with $t > t'$. The next lemma is obvious.

Lemma 2. *Any increasing and continuous curve C from the origin to ω is the contract curve of some utility functions u_1 and u_2.*

We now consider a strengthening of Lemma 2 in which we let one of the utility function, say u_2, be fixed at the beginning. In this case there might not be a $u_1 \in U$ such that $C = C(u_1, u_2)$ because indifference surfaces of u_2 can flatten out near the origin.

For example, let $X_1 = X_2 = R^2_+$, $\omega = (1, 1)$, $u_2(x_1, x_2) = x_1 x_2$, and the curve $C: c(t) = (t, 1 - (1-t)^2)$, $t \in [0, 1]$. The gradient vector of u_2 at $c(t)$ is thus $(1 - t, 1)$ (up to a positive scalar). If C is the contract curve $C(u_1, c_2)$ of u_1 and u_2 for some $u_1 \in U$, then $(1 - t, 1)$ should also be a subgradient vector of u_1 at

$c(t)$. Letting t approach 1, we would have $(0, 1)$ as a subgradient vector of u_1 at $(1, 1)$. This contradicts the assumption that u_1 is increasing.

Nevertheless, we can show that C is *almost* the contract curve of some $u_1 \in U$ and u_2.

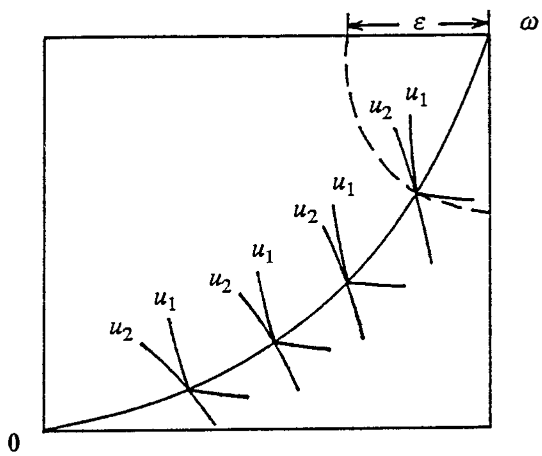

Fig. 2

Lemma 3. *Assume that C is an increasing and continuous curve from the origin to ω. Then for any $u_2 \in U$ and any $\varepsilon > 0$, there is a $u_1 \in U$ such that $[a \in C(u_1, u_2)$ and $\|\omega - a\| \geqq \varepsilon]$ imply $[a \in C]$. (See Fig. 2 for an illustration of Lemma 3.)*

Proof. First consider an arbitrary point $a \in C$, and $a \neq \omega$. Consider all subgradient vectors of u_2 at $(\omega - a)$: $P(u_2;\ \omega - a) = \{p \in R^n \mid \|p\| = 1$, and $u_2(x) \leqq u_2(\omega - a)$ for all x satisfying $p \cdot x \leqq p \cdot (\omega - a)\}$. Since u_2 is strictly quasi-concave, $P(u_2;\ \omega - a)$ in nonempty. And since u_2 is increasing and each p in $P(u_2;\ \omega - a)$ is normalized, $\alpha(a) = \inf\{p_i \mid p \in P(u_2;\ \omega - a)\}$ is positive. We then consider $\alpha = \inf\{\alpha(a) \mid a \in C$ and $\|\omega - a\| \geqq \varepsilon\}$. Because $P(u_2;\ a)$ is upper-hemi-continuous, it is easy to show that α is positive as well.

Now we take a function f^m as defined in the proof of Lemma 1 with $a = \mathbf{0}$ and move the indifference surface S of f^m passing $\mathbf{0}$ along C to generate a family of surfaces $\{S(t)\}$ with $S(t) = \{c(t)\} + S$. We can define a utility function u_1: $u_1(x) = t$ if and only if $x \in S(t)$. (It is easy to extend u_1 to R^n_+ by extending $c(t)$.) Obviously, $u_1 \in U$.

When m is large enough, any vector p satisfying $[\|p\| = 1]$ and $[p_i \geqq \alpha$ for all $i]$ is a subgradient vector of f^m at $\mathbf{0}$, thus u_1 and u_2 are tangent to each other at any point a satisfying $[a \in C$ and $\|\omega - a\| \geqq \varepsilon]$. Also when m is large enough, the upper contour set of u_1 at the point $b \in C$ with $\|\omega - b\| = \varepsilon$ is contained in the set $\{a \mid \|\omega - a\| \leqq \varepsilon\}$. Now choose a large m so that both conditions above are satisfied. The corresponding u_1 is then a desired utility function. Q.E.D.

Proof of Theorem 1. Suppose there exists a mechanism F that is efficient, non-inversely-dictatorial, and strategy-proof. We show that it leads to a contradiction. Let $R(F)$ denote the range of F.

Step 1. (i) $\mathbf{0} \notin R(F)$; and (ii) $\omega \notin R(F)$.
Suppose (i) is not true. Take any (u_1, u_2) such that $F(u_1, u_2) = \mathbf{0}$. Then $F(\hat{u}_1, u_2) = \mathbf{0}$ for any $\hat{u}_1 \in U$ since otherwise 1 can manipulate at (u_1, u_2). Furthermore,

$F(\hat{u}_1, \hat{u}_2) = \mathbf{0}$ for any $\hat{u}_2 \in U$ since otherwise 2 can manipulate at $(\hat{u}_1, \hat{u}_2)$. Thus $F \equiv \mathbf{0}$. But this says that 1 is an inverse dictator which contradicts our assumption. Hence (i) must be true. So is (ii).

Step 2. $[a \in R(F),\ \mathbf{0} \ll a \leq b \ll \omega]$ imply $[b \notin R(F)]$.
If this is not true, then there are (u_1, u_2) and $(\hat{u}_1, \hat{u}_2)$ such that $a = F(u_1, u_2)$ and $b = F(\hat{u}_1, \hat{u}_2)$. Consider $F(u, u_2)$ for any $u \in U(u_1; a)$. Since F is strategy-proof, we have

$$u_1(a) = u_1(F(u_1, u_2)) \geqq u_1(F(u, u_2)), \quad \text{and}$$

$$u(F(u, u_2) \geqq u(F(u_1, u_2)) = u(a) .$$

Hence, by the definition of $U(u_1; a)$, we must have $F(u, u_2) = a$. Similarly, we must have $F(v, \hat{u}_2) = b$ for any $v \in U(\hat{u}_1; b)$. By Lemma 1, both $U(u_1; a)$ and $U(\hat{u}_1; b)$ are nonempty. Take two functions $u \in U(u_1; a)$ and $v \in U(\hat{u}_1; \mathrm{b})$ that are normalized so that $u(a) = v(a)$ and $u(b) = v(b)$. Now let $\tilde{u}_1 = \min\{u, v\}$. Obviously, $\tilde{u}_1 \in U(u_1; a) \cap U(\hat{u}_1; b)$. Hence, we have

$$F(\tilde{u}_1, u_2) = a , \quad \text{and} \quad F(\tilde{u}_1, \hat{u}_2) = b .$$

They lead to $\hat{u}_2(\omega - F(\tilde{u}_1, \hat{u}_2)) = \hat{u}_2(\omega - b) < \hat{u}_2(\omega - a) = \hat{u}_2(\omega - F(\tilde{u}_1, u_2))$ because $a \leq b$ and $\hat{u}_2$ is increasing. But this contradicts the assumption that F is strategy-proof.

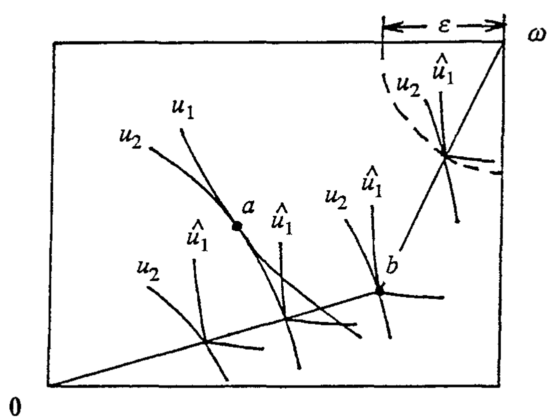

Fig. 3

Step 3. $[a = F(u_1, u_2),\ \mathbf{0} \ll b \ll \omega$, and $u_1(b) > u_1(a)]$ imply $[b \notin R(F)]$.
Suppose $b \in R(F)$. We consider the curve $\overline{Ob\omega}$, u_2, and $\varepsilon = \min(\omega_i - b_i) > 0$. By Lemma 3, there is a $\hat{u}_1 \in U$ such that $[x \in C(\hat{u}_1, u_2)$ and $\|\omega - x\| \geqq \varepsilon]$ imply $[x \in \overline{Ob\omega}]$ (see Fig. 3.) Hence, for any $y \in C(\hat{u}_1, u_2) \backslash \{\mathbf{0}\}$, either $y \leq b$, or $b \leq y$. So $\{b\} = C(\hat{u}_1, u_2) \cap R(F)$ by Step 2. It leads to $F(\hat{u}_1, u_2) = b$ because F is efficient. But then $u_1(F(\hat{u}_1, u_2)) = u_1(b) > u_1(a) = u_1(F(u_1, u_2))$, which contradicts that F is strategy-proof.

Step 4. There is a unique allocation $d \in R(F) \cap \overline{O\omega}$ such that $\mathbf{0} \ll d \ll \omega$.
Consider the straight line $\overline{O\omega}$. Let $u_1 = u_2 = u$ be such that

$$u(x) = \sum_{i=1}^{n} p_i \ln(x_i + \omega_i) , \quad \text{in which } p_i > 0 \text{ for all } i .$$

It is straightforward that $C(u_1, u_2) = \overline{O\omega}$. Because F is efficient, $F(u_1, u_2) = d \in R(F) \cap \overline{O\omega}$. While $\mathbf{0} \ll d \ll \omega$ follows from step 1, the uniqueness of d follows from Step 2.

Step 5 $b \notin R(F)$ for each $b \neq d$ that satisfies $\mathbf{0} \ll b \ll \omega$.
When $b \leq d$, $b \notin R(F)$ by Step 2. When $b \nleq d$, without loss of generality we may assume that $b_1 > d_1$. Now let $u_1 = u_2 = v$ be

$$v(x) = \ln(x_1 + \omega_1) + \beta \cdot \sum_{i=2}^{n} \ln(x_i + \omega_i) \ ,$$

where β is a positive number smaller than $\dfrac{\ln(b_i + \omega_i) - \ln(d_i + \omega_i)}{2(n-1)\ln(2\|\omega\|)}$. $F(u_1, u_2) = d$ by Step 4, and $u_1(b) > u_1(d)$ by the choice of β. Hence, $b \notin R(F)$ by Step 3.

Now we can complete the proof. Take a point e not on $\overline{O\omega}$ satisfying $\mathbf{0} \ll e \ll \omega$ and consider $\overline{Oe\omega}$. By Lemma 2, we can find a pair $(u_1, u_2) \in U \times U$ such that $C(u_1, u_2) = \overline{Oe\omega}$. But Steps 1 and 5 then imply $R(F) \cap C(u_1, u_2) = \phi$. This contradicts the assumption that F is efficient. Q.E.D.

4. Concluding remarks

4.1. Although we still cannot prove Theorem 1 for economies with arbitrarily many agents, we believe that it is a very promising conjecture. The technical difficulty we encounter when we try to prove it for many-agent economies is that the geometry of the set of efficient allocations becomes too complicated. Although we know that it is generally an $(m-1)$-dimensional manifold (here m is the number of agents), it is difficult to visualize in the commodity space with the only exception $m = 2$ where it can be represented by a contract curve. Since the use of contract curves is essential in our current proof of Theorem 1 (as well as in the cited works by Hurwicz, and by D-H-M), it seems that a more thorough investigation of the geometry of the set of efficient allocations is called for to extend Theorem 1 to arbitrarily-many-agent economies.

At this point we need to clear up a misunderstanding some readers might have with regard to results for two-agent economies and many-agent ones. In various occasions many authors have claimed that results similar to those by Hurwicz, and by D-H-M exist for many-agent economies. They usually use the argument that there is a non-dictatorial mechanism for two-agent economies if and only if there is a non-dictatorial mechanism for arbitrarily-many-agent ones. But this is false for private goods economies as shown by the following dictator-making example in Satterthwaite and Sonnenschein [7]. There are three agents. Either agent 1 or 2 gets the whole endowment depending on whether the third agent's marginal rate of substitution at (1, 1) is greater or less than one. This mechanism is efficient, non-dictatorial, and strategy-proof even though no such mechanism exists in two-agent economies by Theorem 1. (Notice, however, the example does not contradicts our conjecture since agent 3 is an inverse dictator).

4.2. When production is introduced into the model, Theorem 1 is no longer always valid. For example, if the technology is linear, i.e. $Y = \{y \in R^n \mid p \cdot y \leqq 0\}$, then the mechanism F that assigns each agent his most desired bundle within the budget line $\{x \in R^n_+ \mid p \cdot x \leqq \frac{1}{2} p \cdot \omega\}$ satisfies all conditions in Theorem 1. This is, however, the only situation we can think of in which Theorem 1 fails. It is very likely that Theorem 1 remains true for all private goods economies with non-linear production technologies. This is another important direction along which extension of Theorem 1 is worth considering.

4.3. In [5] Hurwicz and Walker consider a model that includes pure exchange economies when there is a commodity in which all agents' preferences are linear. To get a result like Theorem 1 which deals with general preferences, they have to assume a priori that a mechanism under consideration is continuous on the set of quasi-linear preferences for a particular topology and that it has to choose some interior allocation. But given that the set of quasi-linear preferences is nowhere dense (in almost all sensible topologies) in the space of all preferences, these assumptions are too strong. It is logically conceivable that in order to achieve both efficiency and strategy-proofness a mechanism must be discontinuous on a small set or has to choose boundary allocations on this small set. Finally, the technique employed there is very specific to quasi-linear preferences and does not generalize.

References

1. Barbera S, Peleg B (1990) Strategy-proof voting schemes with continuous preferences. Soc Choice Welfare 7. 31–38
2. Dasgupta P, Hammond P, Maskin E (1979) The implementation of social choice rules. Rev Econ Stud 46: 153–170
3. Gibbard A (1973) Manipulation of voting schemes: a general result. Econometrica 41 587–602
4. Hurwicz L (1972) On informationally decentralized systems In: McGuire CB, Radner R (eds) Decision and Organization North-Holland, Amsterdam, pp 297–336
5. Hurwicz L, Walker M (1990) On the generic non-optimality of dominant-strategy allocation mechanisms: a general theorem that includes pure exchange economies Econometrica 58: 683–704
6. Satterthwaite MA (1975) Strategy-proofness and Arrow's conditions existence and correspondence theorems for voting procedures and social welfare functions. J Econ Theory 10: 187–217
7. Satterthwaite MA, Sonnenschein H (1981) Strategy-proof allocation mechanisms at differentiable points. Rev Econ Stud 48. 587–597
8. Zhou L (1991) Impossibility of strategy-proof mechanisms in economies with pure public goods. Rev Econ Stud 58: 107–119