Strategy-Proof Allocation Mechanisms at Differentiable Points

Author(s): Mark A. Satterthwaite and Hugo Sonnenschein

Source: *The Review of Economic Studies*, Oct., 1981, Vol. 48, No. 4 (Oct., 1981), pp. 587-597

Published by: Oxford University Press

Stable URL: https://www.jstor.org/stable/2297198

Review of Economic Studies (1981) XLVIII, 587–597 0034-6527/81/00510587$02.00

Strategy-Proof Allocation Mechanisms at Differentiable Points

MARK A. SATTERTHWAITE
Northwestern University
and
HUGO SONNENSCHEIN
Princeton University

1. INTRODUCTION

Consider allocation mechanisms that are single valued and where each agent's strategy space is a set of a priori admissible utility functions. Such an allocation mechanism is strategy-proof if, for each agent, faithfully reporting his true utility function is a dominant strategy. The purpose of this paper is to characterize for the restricted domains associated with economic environments strategy-proof allocation mechanisms at points at which they are differentiable with respect to agents' preferences. Our concern with classical economic environments dictates a framework in which (a) the set of attainable alternatives is a subset of a l-dimensional Euclidean space, (b) the domain of admissible preference n-tuples is restricted (utility functions may be required to satisfy such properties as continuity, monotonicity, and quasiconcavity), and (c) the standard representations of economies are admissible; in particular, the analysis applies to economies with and without production, with and without public goods, and with and without externalities. Indeed, our goal has been to provide a result on strategy-proofness that is as basic for allocation mechanisms within economic environments as the Gibbard–Satterthwaite Theorem (Gibbard 1973 and Satterthwaite 1975) is for voting procedures with unrestricted domain.

The transition from the social choice framework with its minimal mathematical structure on the attainable alternatives and admissible preferences to economic environments with their considerable structure is usefully divided into two steps. For the first step, consider mechanisms that allocate public goods only and are regular on a set of n-tuples of utility functions $\mathscr{R}$. Consideration of mechanisms that also allocate private goods is deferred to the second step. Regularity at $(u_1, \ldots, u_n) \in \mathscr{R}$ means that the allocation changes smoothly as agents change their reported utility functions in a neighbourhood of $(u_1, \ldots, u_n)$. We permit the set U of a priori admissible utility functions to be restricted to any (C^2) open set of utility functions. This requirement that U be open means that the mechanism is broadly applicable, which is to say that it must be defined for more than a "thin slice" of preferences such as those that are representable by additively separable or CES utility functions. With the addition of some technical conditions, we prove that if an allocation mechanism is strategy-proof, regular, and allocates public goods only, then it is dictatorial on $\mathscr{R}$. This corresponds precisely with the Gibbard–Satterthwaite Theorem that strategy-proofness in the social choice framework implies dictatorship.

For the second step, consider broadly applicable mechanisms that allocate private as well as public goods and are regular on a set $\mathscr{R}$. Whereas with "public goods only" each agent's utility evaluation of an allocation depends on every coordinate of the allocation vector (and broad applicability requires that the nature of this dependence be allowed to vary), with "private goods only" (and no externalities) each agent's evaluation of an allocation depends only on what he privately receives. The utcome of this step is fundamentally different than that of the first step where the validity of a Gibbard–Satterthwaite type theorem was confirmed for the public goods only economic environment. The following example demonstrates that this type theorem fails and strategy-proof nondictatorial mechanisms do exist if the economy has private goods only.

There are two pure private goods, x and y, and three consumers. Commodity x is produced from y according to $x = y$. The economy begins with three units of y. Agent two and agent three share the first unit of y in proportions that depend on "the mean curvature" of one's utility function. If it is "very curved" two gets the unit of y and if it is "very flat", three gets the unit of y. The second unit of y is shared according to the rule obtained by replacing one by two, two by three, and three by one in the rule for sharing the first unit of y. Similarly, the third unit of y is shared by replacing one by three, two by one, and three by two in the first unit's rule. Each consumer is then assigned his utility maximal point on his budget line $x + y = \bar{y}$ where $\bar{y}$ is that share of the economy's initial endowment that he receives based on the mean curvatures of the other two agents' utility functions. This mechanism is strategy-proof because each agent's constraint set is exogenous to his own strategy and the mechanism automatically picks his utility maximal point on that set.[1]

A striking feature of the preceding example is that the agent can maintain his bundle unchanged at the same time he causes changes in the bundles that the other agents receive. He does this by changing the mean curvature of his utility function while keeping its gradient constant at his current consumption bundle. We refer to mechanisms for which such action is possible as bossy and note that with public goods only (which means that everybody receives the identical bundle) bossiness is never possible. The above example shows that there exist bossy mechanisms that are strategy-proof and nondictatorial. Thus a Gibbard–Satterthwaite type theorem does not hold for private goods economies if bossy mechanisms are admitted.

If, however, bossy mechanisms are excluded from consideration, then we have been able to establish a Gibbard–Satterthwaite type of theorem for private goods only and mixed public-private goods environments. It states that every non-bossy, regular mechanism that is strategy-proof and broadly applicable is a serial dictatorship. Serial dictatorship means that the mechanism consists of one or more hierarchies of agents where the highest ranking agent in each hierarchy selects his allocation from a feasible set that is exogenously given, the second highest ranking agent selects his allocation from a feasible set that depends on the first agent's choice, the third highest ranking agent selects his allocation from a feasible set that depends on the first and second agents' choices, etc. Consequently, an agent who is high on a hierarchy is a dictator to those agents lower on that hierarchy in the sense that he can affect what is available to them to choose among and they can not affect him reciprocally. He is not, however, necessarily a dictator in the stronger senses of being able to choose any technologically feasible outcome for himself and being able to impose particular outcomes on the other agents. Additional conditions on the nature of the mechanism and the set of admissible utility functions must be added to demonstrate that there is a single hierarchy. Finally, observe that just as in the Gibbard–Satterthwaite Theorem, Pareto optimality is not a condition of our theorem.

Our work builds on and complements a long list of previous contributions. These are conveniently classified by whether they originated in the incentive compatibility literature or within the social choice literature. Samuelson, in a classic paper (Samuelson 1954), stated that in an economy with public goods, it would be in an individual's interest to misrepresent his preferences. Hurwicz (1972) showed that even in a standard (finite number of agents) private goods, perfectly competitive economy, an individual can gain by misrepresenting his preferences; thus, gain from misrepresentation does not by itself distinguish private goods economies from economies with public goods. In the same paper, he proved that for two-person, two-good exchange economies, there exists no strategy-proof mechanism that (a) always generates Pareto optimal outcomes, (b) is individually rational, and (c) works for all economies in which agents have convex indifference curves. Green and Laffont (1977) considered incentive compatibility within the context of an economy having one or more public goods and a single, private good. Within this specific context and under the strong restriction on the set U of admissible utility functions that each agent's utility be linear in the private good (i.e. utility is transferrable), they showed that every strategy-proof mechanism that is optimal in the sense of maximizing the sum of the public good components of the agents' utility functions is necessarily a Groves mechanism. See also the work of Clarke (1971), Groves (1973), and Groves and Loeb (1975). Green and Laffont (1979) have summarized and extended this body of work.

Gibbard (1973) and Satterthwaite (1975), for the case of unrestricted domain and in the context of the social choice literature, showed that no strategy-proof voting procedure exists that is nondictatorial and has a range of at least three alternatives. Kalai and Muller (1977) and Maskin (1976*a* and 1976*b*) asked to what degree the set U of admissible utility functions must be reduced in order to obtain a possibility result instead of Gibbard and Satterthwaite's impossibility theorem. They independently derived necessary and sufficient conditions for the set U of admissible utility functions to admit the construction of nondictatorial strategy-proof mechanisms that are derivable from an Arrow social welfare function. Since requiring that economic allocation mechanisms be rationalizable by an Arrow social welfare function is unnaturally restrictive, these results are not satisfactory in the present context. Dasgupta, Hammond, and Maskin (1979) reported the result that for pure exchange economies, there exists no allocation mechanism that is (a) nondictatorial, (b) always achieves Pareto optimality, and (c) works for all economies in which agents have arbitrary strictly convex and strictly monotone preferences.[2]

Three sections follow. Section 2 presents the model and formally states the theorems. Section 3 discusses the key assumptions. Section 4 is devoted to proofs.

2. THE MODEL AND THEOREMS

For simplicity assume that all agents have the same *consumption set* $X \subset R^l$ which is compact, convex, and has a non-empty interior. The class of *admissible utility functions* on X is denoted by U. We assume throughout that U is a convex subset of a linear function space and is endowed with a C^2 topology. The ith agent is defined by his utility function $u_i \in U$. An *allocation mechanism* for an n agent economy is a function $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n): U^n \to X^n$. The allocation mechanism σ is *manipulable* by i at $u \in U^n$ if there exists $\bar{u}_i$ such that $u_i\sigma_i(u\backslash\bar{u}_i) > u_i\sigma_i(u)$, where $(u\backslash\bar{u}_i)$ denotes the vector u with ith coordinate replaced by $\bar{u}_i$, and $u_i\sigma_i(u\backslash\bar{u}_i)$ is shorthand for $u_i[\sigma_i(u\backslash\bar{u}_i)]$. The allocation mechanism is *strategy-proof* on U^n if it is not manipulable by any i at any $u \in U^n$. The allocation mechanism is *broadly applicable* (BA) if U is open.[3]

The allocation mechanism σ is *non-bossy* (NB) if for all $u \in U^n$, for all i, j, and for all $u_j^1, u_j^2, [\sigma_j(u\backslash u_j^1) = \sigma_j(u\backslash u_j^2) \Rightarrow \sigma_i(u\backslash u_j^1) = \sigma_i(u\backslash u_j^2)]$. Let $C^2(X)$ denote all C^2 functions that are defined on X. The strategy-proof mechanism σ is *regular* at $u = (u_1, u_2, \ldots, u_n)$ if:

(a) σ is continuously differentiable in u; in particular, for all $v \in [C^2(X)]^n$, the derivative $D_v\sigma(u) = \lim_{\lambda\to 0} [\sigma(u+\lambda v) - \sigma(u)]/\lambda$ exists with the standard properties that for all $c, d \in R$ and for all $v, w \in [C^2(X)]^n$, $D_{cv+dw} = cD_v\sigma(u) + dD_w\sigma(u)$.

(b) For all i, $B_i(u) = \{x^i \in X | a\ u_i' \text{ exists with } x^i = \sigma_i(u\backslash u_i')\}$ is a m_i-dimensional, $0 \leqq m(i) \leqq l-1$, smooth manifold in a neighbourhood of $\sigma_i(u)$ and is continuously able in u. Formally, there exists $f: X \times U^n \to R^{l-m(i)}$ such that f is C^2 in both variables and $B_i(u) = \{x \in X | f(x, u) = 0\}$.[4]

(c) For all i, $\sigma_i(u)$ is the unique and regular maximizer of u_i on $B_i(u)$.[5]

Let $\mathcal{R} \subset U^n$ denote the set of all regular points. We speak of a regular mechanism to suggest the smoothness and nondegeneracy assumptions on σ that guarantee (a), (b), and (c) above. Whenever we refer to strategy-proof allocation mechanisms, we restrict attention to those for which the regular points $\mathcal{R} \subset U^n$ form an open set.

Let (v_i) denote the vector in $[C^2(X)]^n$ that is zero in all coordinates except the ith and is v_i in that coordinate. Agent i *affects* agent $j (i \neq j)$ at $u \in U^n$ if $v_i \in C^2(X)$ exists such that $D_{(v_i)}\sigma_j(u) \neq 0$. Agent i *affects* j's *utility* $(i \neq j)$ at $u \in U^n$ if a $v_i \in C^2(X)$ exists such that $D_{(v_i)}u_j\sigma_j(u) \neq 0$. If i affects j at u, we write $iA(u)j$, and if i affects j's utility at u we write $i\tilde{A}(u)j$. Finally, define $S^{ij} = \{u \in \mathcal{R}: iA(u)j\}$ and $\tilde{S}^{ij} = \{u \in \mathcal{R}: i\tilde{A}(u)j\}$.

The following result establishes that if σ is strategy-proof and satisfies NB and BA, then for each regular point $u \in \mathcal{R}$, the affects relation A is an acyclic relation. Since the sets S^{ij} are open (Lemma 1), it follows that, for each $u \in \mathcal{R}$, there are a collection of serial dictatorships that are fixed in a neighborhood of u. The fact that serial dictatorship rather than dictatorship obtains is analogous to Luce and Raiffa's observation [Luce and Raiffa (1957), p. 344] that a serially dictatorial social welfare function (where agents who are low on the "pecking order" get to determine the rank only of the alternatives those above them are indifferent among) is consistent with all of Arrow's conditions except nondictatorship.

Theorem 1. *If σ be strategy-proof and satisfies NB and BA, then, for all $u \in \mathcal{R}$, $A(u)$ is acyclic.*

Theorem 1 states that, for each $u \in \mathcal{R}$, $A(u)$ is acyclic; it permits the hierarchies of serial dictators to vary as $u \in \mathcal{R}$ varies. Theorem 2 states conditions that are sufficient to guarantee that for all $u \in \mathcal{R}$, a single hierarchy of serial dictators holds.

Theorem 2. *Let σ be strategy-proof and satisfy NB and BA. If in addition $\mathcal{R}$ is connected and A is everywhere total, then there exists a permutation $Q: \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ such that, for all $u \in \mathcal{R}$, $iA(u)j$ if and only if $Q(i) > Q(j)$.*

A mechansim σ is *everywhere total* if $A(u)$ is total for each $u \in \mathcal{R}$.[6] Precisely how restrictive these assumptions of connectedness and totality are, however, has not been worked out.[7]

For economies with pure public goods only, property NB is automatically satisfied and serial dictatorship reduces to dictatorship, so that for this case Theorem 2 reads as follows.

Theorem 3. *Let σ be strategy-proof and satisfy BA. In addition, let $\mathcal{R}$ be connected and A be everywhere total. If σ is defined for public goods only, then σ is dictatorial.*

For σ to be defined for public goods only means that, for all i, j and for all $u \in U^n$, $\sigma^i(u) = \sigma^j(u)$. Dictatorship means formally that an agent i exists whose constraint set B_i is independent of $(u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n)$; furthermore, for each $u \in R$, the allocation that all the agents collectively receive is i's most preferred point on B_i.

3. DISCUSSION OF THE MODEL

Our specification of the mechanism σ is flexible enough to accommodate pure exchange economies, pure public good economies, economies with production, and mixed public-private good economies. For the case of pure public goods economies, the n components of the mechanism $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ are constrained to be identical. For the case of a purely private good economy, no restrictions apply across σ's components except those that constrain the outcome to be feasible. For the case of mixed public and private goods, only the public goods components of the functions $\sigma_i = (\sigma_{i1}, \sigma_{i2}, \ldots, \sigma_{il})$, $i = 1, 2, \ldots, n$, are restricted to be identical across agents.

Specification of each agent's strategy space to be the set U of a priori admissible utility functions is not restrictive because we are concerned in this paper with dominant strategy mechanisms. This assertion follows from Gibbard's observation (Gibbard 1973) that within a given environment a dominant strategy mechanism exists if and only if an equivalent strategy-proof mechanism with strategy space U^n exists.[8] Gibbard's observation means that because we are interested in dominant strategy mechanisms we can, as a matter of convenience, study those mechanisms that limit each agent's strategy space to a set U of admissible utility functions.

The requirement that a mechanism be broadly applicable follows from the observation that while preferences within an economic environment may have considerable a priori structure such as strict convexity, preferences are not naturally limited to any particular parametric form. Here, through our assumption on the openness of U, we assert that if u_i is an admissible utility function describing agent i's preferences, then all those utility functions u_i' that are sufficiently near u_i (in the sense of the C^2 topology) should also be a priori admissible.[9]

The example in the introduction illustrates that mechanisms exist for the allocation of private goods that are strategy-proof, but not dictatorial. Because of this, we introduced the notion of bossiness and showed that a Gibbard–Satterthwaite type theorem obtains when attention is restricted to nonbossy mechanisms. It is therefore appropriate to consider whether non-bossiness is a reasonable or desirable condition to require of mechanisms.[10]

While we have not exhaustively considered this question, we have identified one substantial consideration that bears on nonbossiness's reasonableness and desirability. It relates to simplicity of design. Most allocation mechanisms, including the competitive mechanism, have equilibria that can simply and naturally be defined in terms of an adding up condition and some marginal equalities arising from the several agents' first-order conditions. Such mechanisms might appropriately be called first-order. They necessarily have the property that if several agents change their preferences, but maintain their initial marginal rates of substitution at their initial allocations, then the initial equilibrium is retained unchanged because the changes in preference leave the adding up condition and marginal equalities intact. This, however, means that a bossy mechanism cannot be a first-order mechanism.[11] Thus, the simplicity of first-order mechanisms can only be purchased at the cost of excluding bossy mechanisms from consideration.

One final comment concerning bossiness and non-bossiness is merited. Bossiness only has meaning within the context of pure private goods where each agent cares not at all about what allocations other agents receive. If agents are allowed to care even slightly about other agents' allocations, Theorem 3 applies, and non-bossiness is no longer needed as an assumption to establish that strategy-proofness implies serial dictatorships.

As Hurwicz (1972) observed, the competitive mechanism is not a strategy-proof means for allocating private goods among a finite set of consumers. It is instructive to see how this obtains as an application of our theorems for arbitrary regular mechanisms. Suppose, contrary to the assertion, that the competitive mechanism is strategy-proof for a finite set of consumers. At any regular point $u \in U^n$ the competitive mechanism is first-order; therefore it is non-bossy. Theorem 1 consequently applies and states that the A relation must be acyclic. This, however, is not true for the competitive mechanism. Each agent faces a constraint set B_i (offer curve) that varies with every other agent's preferences. This means that each pair of agents reciprocally affects each other, which is to say that the A relationship is cyclic, not acyclic as the assumption that it is strategy-proof necessarily implies. Therefore the competitive mechanism is not strategy-proof.

Serial dictatorship is unattractive not only because it implies a non-symmetric distribution of power, but also because it generates allocations that (generally) are not Pareto optimal with respect to agents' preferences and the underlying, technologically given, production possibility set. To see this, consider an economy in which there are at least two private goods and a strictly convex set of production possibilities. Assume the mechanism σ is strategy-proof, serially dictatorial, and always generates optimal outcomes that are nontrivial in the sense that every consumer receives positive amounts of the private goods. The agent who is at the bottom of a hierarchy of serial dictators faces a strictly convex feasible set because he gets to choose among the residual production possibilities after everyone else has chosen. He cannot affect the choices of the higher ranking agents; therefore, the marginal rates of substitution (for private goods) of the higher ranking agents at their allocations may be taken as fixed and given relative to the strategies of the bottom ranking agent. Moreover, the assumption that σ generates optimal outcomes implies that the marginal rates of substitution of those higher ranking agents are all equal. Consequently, for optimality of the outcome to be preserved, the bottom ranking agent must choose that unique point on his strictly convex feasible set that results in him having the same marginal rate of substitution for private goods that the higher ranking agents have. But his preferences, or a perturbation of his preferences, will be such that he chooses a different point. This destroys optimality and contradicts the assumption that σ always generates optimal outcomes. Therefore, serial dictatorship violates optimality.

A restatement of this result is that if Pareto efficiency is to be achieved, then agents generally must be able to affect each other reciprocally. This implies that attempting to construct allocation mechanisms that are efficient, non-bossy, and strategy-proof for general cases, where agents' consumption sets have dimension of at least two ($l \geqq 2$) and preferences are broadly applicable, are certain to fail. In particular, trying to piece together on the domain of admissible preference n-tuples U^n a complicated pattern of changing and distinct serial dictatorships can only achieve strategy-proofness and non-bossiness, but not efficiency. Thus the properties of first-orderness, efficiency, and strategy-proofness appear to be in conflict. Consequently one might interpret the present analysis as supporting the importance of bossiness in the construction of efficient and strategy-proof mechanisms.[12]

4. PROOFS

Lemma 1. *Let σ be strategy-proof. Then, for all $i \neq j$, (i) $\tilde{S}^{ij} \subset S^{ij}$, (ii) S^{ij} is open, and (iii) $\tilde{S}^{ij}$ is open.*

Proof. The first claim follows from the fact $i\tilde{A}(u)j$ implies $iA(u)j$. The latter two are a consequence of the continuity of the derivatives assumed in the definition of regular. ‖

Lemma 2. *Let σ be strategy-proof and satisfy NB and BA. Then, for all $i \neq j$, $\tilde{S}^{ij} \cap \tilde{S}^{ji} = \varnothing$.*

Proof. We assume $1\tilde{A}(\bar{u})2$ at the regular utility n-tuple $\bar{u}$ and show that $2\tilde{A}(\bar{u})1$ is impossible. With the understanding that $\bar{u}_3, \bar{u}_4, \ldots, \bar{u}_n$ are fixed, we imagine that σ depends only on u_1 and u_2. The assumption $1\tilde{A}(\bar{u})2$ means there exists $v_1 \in C^2(X)$ such that $D_{(v_1)}\bar{u}_2\sigma_2(\bar{u}) > 0$. Let $v_2 \in C^2(X)$ be arbitrary. We will show $D_{(v_2)}\bar{u}_1\sigma_1(\bar{u}) \neq 0$ leads to a contradiction.

Define $B(\lambda) = B_2(\bar{u}_1 + \lambda v_1, \bar{u}_2)$ where $\bar{u}_3, \ldots, \bar{u}_n$ have been suppressed. By regularity, $B(\lambda)$ is a smooth manifold in a neighbourhood of $\sigma_2(\bar{u})$, provided that λ is sufficiently small. By strategy-proofness and regularity, $\sigma_2(\bar{u}_1 + \lambda v_1, \bar{u}_2)$ is the unique maximizer of $\bar{u}_2$ on $B(\lambda)$ (again provided that λ is small) and by hypothesis $D_{(v_1)}\bar{u}_2\sigma_2(\bar{u})>0$. Similarly, for λ small enough, $\sigma_2(\bar{u}_1, \bar{u}_2+\lambda v_2)$ is the unique maximizer of $(\bar{u}_2+\lambda v_2)\in U$ on $B(0)$. Since σ satisfies BA, for any λ sufficiently small, there exists $w\in U$ such that $\sigma_2(\bar{u}_1, \bar{u}_2+\lambda v_2)$ is the unique maximizer of w on $B(0)$ and $\sigma_2(\bar{u}_1+\lambda v_1, \bar{u}_2)$ is the unique maximizer of w on $B(\lambda)$.[13]

Since σ is strategy-proof $\sigma_2(\bar{u}_1, w)=\sigma_2(\bar{u}_1, \bar{u}_2+\lambda v_2)$ and $\sigma_2(\bar{u}_1+\lambda v_1, w)=\sigma_2(\bar{u}_1+\lambda v_1, \bar{u}_2)$. By NB $\bar{u}_1\sigma_1(\bar{u}_1, \bar{u}_2+\lambda v_2)=\bar{u}_1\sigma_1(\bar{u}_1, w)$ and $\bar{u}_1\sigma_1(\bar{u}_1+\lambda v_1, w)=\bar{u}_1\sigma_1(\bar{u}_1+\lambda v_1, \bar{u}_2)$. But by strategy-proofness $\bar{u}_1\sigma_1(\bar{u}_1, w)\geqq\bar{u}_1\sigma_1(\bar{u}_1+\lambda v_1, w)$. Therefore

$$\bar{u}_1\sigma_1(\bar{u}_1, \bar{u}_2+\lambda v_2)\geqq\bar{u}_1\sigma_1(\bar{u}_1+\lambda v_1, \bar{u}_2) \tag{1}$$

for all λ sufficiently small. We may assume (if necessary replace v_2 with $-v_2$) that there exists a sequence of positive $\lambda_n\to 0$ such that for all n

$$\bar{u}_1\sigma_1(\bar{u}_1, \bar{u}_2)\geqq\bar{u}_1\sigma_1(\bar{u}_1, \bar{u}_2+\lambda_n v_2) \tag{2}$$

From (1) and (2),

$$0\geqq\frac{\bar{u}_1\sigma_1(\bar{u}_1, \bar{u}_2+\lambda_n v_2)-\bar{u}_1\sigma_1(\bar{u}_1, \bar{u}_2)}{\lambda_n}\geqq\frac{\bar{u}_1\sigma_1(\bar{u}_1+\lambda_n v_1, \bar{u}_2)-\bar{u}_1\sigma_1(\bar{u}_1, \bar{u}_2)}{\lambda_n} \tag{3}$$

As λ_n approaches zero (3) becomes

$$0\geqq D_{(v_2)}\bar{u}_1\sigma_1(\bar{u})\geqq D_{(v_1)}\bar{u}_1\sigma_1(\bar{u}). \tag{4}$$

The right hand side of (4) must be zero because σ is strategy-proof; if it were not zero, then the first order condition for $\bar{u}_1$ being agent one's dominant strategy would not be met. Therefore $0\geqq D_{(v_2)}\bar{u}_1\sigma_1(\bar{u})\geqq 0$, which is to say that $D_{(v_2)}\bar{u}_1\sigma_1(\bar{u})=0$. Thus agent two cannot affect agent one's utility at $\bar{u}$. ||

Lemma 3. *If σ is strategy-proof and satisfies BA, then, for all $i\neq j$, $\bar{\tilde{S}}^{ij}\supset S^{ij}$ where $\bar{\tilde{S}}^{ij}$ denotes the closure of $\tilde{S}^{ij}$.*

Proof. Suppose the lemma is false. A $u\in\mathcal{R}$ therefore exists such that: (a) $iA(u)j$ and not $i\tilde{A}(u)j$ and (b) a neighbourhood $N(u)=N(u_1)\times N(u_2)\times\cdots\times N(u_n)\subset\mathcal{R}$ exists for which $u'\in N(u)$ implies $iA(u')j$ and not $i\tilde{A}(u')j$. Regularity and BA imply that we may select a neighbourhood $\bar{N}=N(\sigma_j(u))\subset X$ so that (a) corresponding to each x in $B_j(u)\cap\bar{N}$ is an admissible utility function $u_j^x\in N(u_j)$ that has its maximum on $B_j(u)$ at x and (b), for all $u'\in N(u)$, the manifold $B_j(u')\cap\bar{N}$ is smooth, continuously differentiable in u, and m-dimensional where $0\leqq m\leqq l-1$. Note that because σ is strategy-proof $\sigma_j(u\backslash u_j^x)=x$. Because the proof for the general case where the dimensionality m of $B_j(u)$ may have any value between 0 and $l-1$ is lengthy, we present here a proof for the special case where $m=l-1$. Proof for the general case is presented in Satterthwaite and Sonnenschein's technical momorandum (1979).

Let $\perp(z)$ denote for any $z\in\bar{N}$ the normal to $B_j(u)$ that passes through z. Establish a new coordinate system for the neighbourhood $\bar{N}$. Let a point that is z in the original system become the point (x, y) in the new system where (i) $x=\perp(z)\cap B_j(u)$ and (ii) y is the Euclidean distance (up) from x to z. Figure 1 illustrates this transformation.

Let $v_i\in C^2(X)$ be arbitrary and for each $(x, 0)\in\bar{N}$, define $f^x(\lambda)$ to be the second component of the unique point $\perp(x, 0)\cap B_j(\lambda)$ where $B_j(\lambda)=B_j(u\backslash u_i+\lambda v_i)$. Thus $\perp(x, 0)\cap B_j(\lambda)\equiv(x, f^x(\lambda))$. By assumption, $D_{(v_1)}u_j^x\sigma_j(u\backslash u_j^x)=0$ for all $(x, 0)\in\bar{N}$. Therefore $df^x(0)/d\lambda=0$ for all $(x, 0)\in\bar{N}$ because if otherwise, then changing λ would cause j's constraint set to move in the direction $\perp(x, 0)$, thus making feasible points that j prefers to $\sigma_j(u\backslash u_j^x)$. Figure 2 illustrates this.

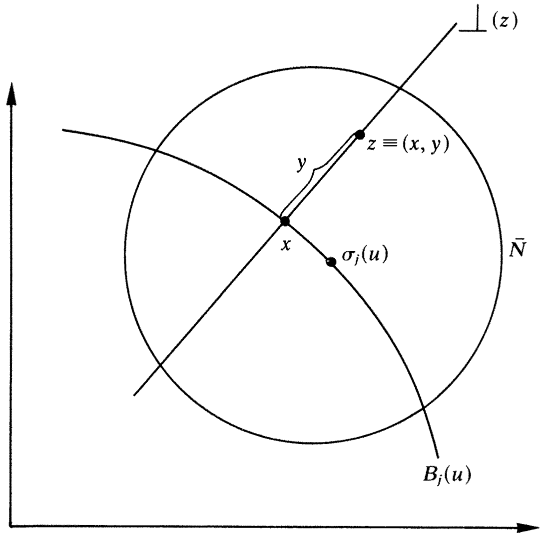

FIGURE 1

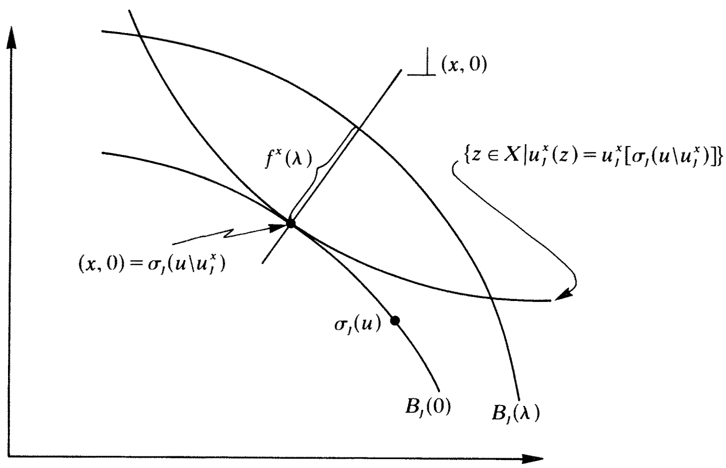

FIGURE 2

Define for all $(x, y) \in \bar{N}$, $F(x, y; \lambda) = y - f^x(\lambda)$. Note that within $\bar{N}$ $B_j(\lambda)$ is represented by the solution to $F(x, y; \lambda) = 0$ and that, by (b) of the definition of regularity, F is C^2. Therefore, since σ is strategy-proof, the point $\sigma_j(u \backslash u_i + \lambda v_i)$ maximizes u_j subject to $F(x, y; \lambda) = 0$. Regularity guarantees that the derivative $D_{(v_i)}\sigma_j(u)$ exists. Finally, the definition of F and the result $df^x(0)/d\lambda = 0$ together imply that for all points $(x, 0) \in B_j(0) \cap \bar{N}$

$$\frac{dF(x, 0; 0)}{d\lambda} = -\frac{df^x(0)}{d\lambda} = 0, \tag{5}$$

which is to say that j's constraint set $B_j(\lambda)$ is fixed and does not move as λ varies about 0. Therefore $D_{(v_i)}\sigma_j(u)=0$ because the only way i can affect j is by moving j's constraint set $B_j(\lambda)$. This contradicts the hypothesis that $iA(u)j$, which completes the proof. ‖

Lemma 4. *If σ is strategy-proof and satisfies NB and BA, then for all $i \neq j$, $S^{ij} \cap S^{ji} = \varnothing$.*

Proof. By Lemma 1, $\tilde{S}^{ij}$ and $\tilde{S}^{ji}$ are open and by Lemma 2 $\tilde{S}^{ij} \cap \tilde{S}^{ji} = \varnothing$. Therefore $\bar{\tilde{S}}^{ij} \cap \tilde{S}^{ji} = \varnothing$ and so $S^{ij} \cap \tilde{S}^{ji} = \varnothing$ follows from Lemma 3. Since S^{ij} and $\tilde{S}^{ji}$ are open and disjoint, $S^{ij} \cap \bar{\tilde{S}}^{ji} = \varnothing$. Applying Lemma 3 once again gives $S^{ij} \cap S^{ji} = \varnothing$, which is the required result.

Lemma 5. *If σ is strategy-proof and satisfies NB and BA, then, for all distinct i, j, and k, $\tilde{S}^{ij} \cap \tilde{S}^{jk} \subset \bar{\tilde{S}}^{ik}$.*

Proof. Let $\bar{u} \in \tilde{S}^{ij} \cap \tilde{S}^{jk}$ and $N(\bar{u}) \subset \tilde{S}^{ij} \cap \tilde{S}^{jk}$ be arbitrary. We will show that a $u \in N(\bar{u})$ exists such that $u \in \tilde{S}^{ik}$. Assume without loss of generality that $i=1$, $j=2$, and $k=3=n$. Since $\bar{u} \in \tilde{S}^{12}$ there exists v_1 such that $D_{(v_1)}\bar{u}_2\sigma_2(\bar{u})>0$, and since $\bar{u} \in \tilde{S}^{23}$ there exists v_2 such that $D_{(\alpha v_2)}\bar{u}_3\sigma_3(\bar{u})>0$ for every scalar $\alpha>0$. *BA* and regularity imply a $\tilde{u}_2$, a $\bar{\lambda} \in (0,1]$, and a $\alpha>0$ exist such that $(\bar{u}_1, \tilde{u}_2, \bar{u}_3) \in N(\bar{u})$ and $\tilde{u}_2$ attains a unique maximum on $B_2(\bar{u}_1+\lambda v_1, \bar{u}_2, \bar{u}_3)$ at $\sigma_2(\bar{u}_1+\lambda v_1, \bar{u}_2+\alpha\lambda v_2, \bar{u}_3)$ for all $\lambda \in [0, \bar{\lambda}]$. Therefore, for all $\lambda \in [0, \bar{\lambda}]$, $\sigma_2(\bar{u}_1+\lambda v_1, \tilde{u}_2, \bar{u}_3) = \sigma_2(\bar{u}_1+\lambda v_1, \bar{u}_2+\alpha\lambda v_2, \bar{u}_3)$ because σ is strategy-proof and $\sigma_3(\bar{u}_1+\lambda v_1, \tilde{u}_2, \bar{u}_3) = \sigma_3(\bar{u}_1+\lambda v_1, \bar{u}_2+\alpha\lambda v_2, \bar{u}_3)$ because σ satisfies NB. Therefore $D_{(v_1)}\bar{u}_3\sigma_3(\bar{u}_1, \tilde{u}_2, \bar{u}_3) = D_{(v_1,\alpha v_2,0)}\bar{u}_3\sigma_3(\bar{u}_1, \bar{u}_2, \bar{u}_3) = D_{(v_1)}\bar{u}_3\sigma_3(\bar{u}) + D_{(\alpha v_2)}\bar{u}_3\sigma_3(\bar{u})$ where the second equality follows from the linearity of the D operator. If this expression is nonzero, then $(\bar{u}_1, \tilde{u}_2, \bar{u}_3) \in \bar{\tilde{S}}_{13}$. Suppose, on the other hand, it is zero. $D_{(\alpha v_2)}\bar{u}_3\sigma_3(\bar{u}) \neq 0$ by hypothesis; therefore $D_{(v_1)}\bar{u}_3\sigma_3(\bar{u}) \neq 0$ and, thus, $\bar{u} \in \tilde{S}_{13}$. Consequently, in either of the possible cases, there exists a $u \in N(\bar{u})$ such that $u \in \tilde{S}^{13}$, which completes the proof. ‖

Proof of Theorem 1. Without loss of generality, suppose that $S^{12} \cap S^{23} \cap \cdots \cap S^{(J-1)J} \neq \varnothing$. We will show that necessarily $S^{12} \cap S^{23} \cap \cdots \cap S^{(J-1)J} \cap S^{J1} = \varnothing$. Because S is open, a u and $N(u)$ exist such that $u \in N(u) \subset S^{12} \cap S^{23} \cap \cdots \cap S^{(J-1)J}$. Lemma 3 states that $S^{ij} \subset \bar{\tilde{S}}^{ij}$; therefore $u \in N(u) \subset \bar{\tilde{S}}^{12} \cap \bar{\tilde{S}}^{23} \cdots \cap \bar{\tilde{S}}^{(J-1)J}$. This implies that a $\bar{u} \in N(u)$ exists such that $\bar{u} \in \tilde{S}^{12} \cap \tilde{S}^{23} \cap \cdots \cap \tilde{S}^{(J-1)J}$ because every neighbourhood of a point contained in the closure of a set must meet the set. Consequently $\tilde{S}^{12} \cap \tilde{S}^{23} \cap \cdots \cap \tilde{S}^{(J-1)J} \neq \varnothing$.

Pick any $\bar{u}$ and neighbourhood $N(\bar{u})$ such that $\bar{u} \in N(\bar{u}) \subset \tilde{S}^{12} \cap \tilde{S}^{23} \cap \cdots \cap \tilde{S}^{(J-1)J}$. By Lemma 5, a $\bar{u}'$ exists such that $\bar{u}' \in N(\bar{u}) \cap \tilde{S}^{13}$, which is equivalent to saying that $\bar{u} \in \bar{\tilde{S}}^{13}$. We now show that $\bar{u} \in \bar{\tilde{S}}^{14}$. Pick a neighbourhood $N(\bar{u}')$ such that $N(\bar{u}') \subset N(\bar{u}) \cap \tilde{S}^{13}$. Since $N(\bar{u}) \subset \tilde{S}^{12} \cap \tilde{S}^{23} \cap \tilde{S}^{34} \cap \cdots \cap \tilde{S}^{(J-1)J}$, $N(\bar{u}') \subset \tilde{S}^{34}$ also. Thus $N(\bar{u}') \subset \tilde{S}^{13} \cap \tilde{S}^{34}$ and, by Lemma 5, a $\bar{u}'' \in N(\bar{u}') \cap \tilde{S}^{14}$ exists. Since $\bar{u}'' \in N(\bar{u}') \subset N(\bar{u})$, $\bar{u}'' \in N(\bar{u})$, which implies that $\bar{u} \in \bar{\tilde{S}}^{14}$. Repeated application of this argument leads to the conclusion $\bar{u} \in \bar{\tilde{S}}^{1J}$. Since $\bar{u}$ was picked arbitrarily from $\tilde{S}^{12} \cap \tilde{S}^{23} \cap \cdots \cap \tilde{S}^{(J-1)J}$, it follows that $\tilde{S}^{12} \cap \tilde{S}^{23} \cap \cdots \cap \tilde{S}^{(J-1)J} \subset \bar{\tilde{S}}^{1J}$.

$\bar{\tilde{S}}^{1J} \cap S^{J1} = \varnothing$ because $\tilde{S}^{1J} \subset S^{J1}$, $\tilde{S}^{1J} \cap S^{J1} = \varnothing$ implies $\bar{\tilde{S}}^{1J} \cap S^{J1} = \varnothing$, and Lemma 4 states that $S^{1J} \cap S^{J1} = \varnothing$. Therefore $\bar{\tilde{S}}^{1J} \subset \mathscr{R} \backslash S^{J1}$ where $\mathscr{R} \backslash S^{J1}$ is the complement of S^{J1} relative to $\mathscr{R}$. This means that $\tilde{S}^{12} \cap \tilde{S}^{23} \cap \cdots \cap S^{(J-1)J} \subset \bar{\tilde{S}}^{1J} \subset \mathscr{R} \backslash S^{J1}$ or, equivalently,

$$\tilde{S}^{12} \cap \tilde{S}^{23} \cap \cdots \cap \tilde{S}^{(J-1)J} \cap S^{J1} = \varnothing. \tag{6}$$

Since each of the J sets on the left hand side of (6) are open in $\mathcal{R}$; alternating applications of (a) the rule that if $A \cap B = \varnothing$, then $\bar{A} \cap B = \varnothing$ and (b) Lemma 3 gives:

$$\begin{aligned} \bar{\tilde{S}}^{12} \cap (\tilde{S}^{23} \cap \cdots \cap \tilde{S}^{(J-1)J} \cap S^{J1}) &= \varnothing \\ S^{12} \cap (\tilde{S}^{23} \cap \cdots \cap \tilde{S}^{(J-1)J} \cap S^{J1}) &= \varnothing \\ \bar{\tilde{S}}^{23} \cap (S^{12} \cap \tilde{S}^{34} \cap \cdots \cap \tilde{S}^{(J-1)J} \cap S^{J1} &= \varnothing \\ S^{23} \cap (S^{12} \cap \tilde{S}^{34} \cap \cdots \cap \tilde{S}^{(J-1)J} \cap S^{J1}) &= \varnothing \\ \bar{\tilde{S}}^{34} \cap (S^{12} \cap S^{23} \cap \tilde{S}^{45} \cap \cdots \cap \tilde{S}^{(J-1)J} \cap S^{J1}) &= \varnothing \\ &\vdots \\ S^{12} \cap S^{23} \cap \cdots \cap S^{J1} &= \varnothing, \end{aligned} \tag{7}$$

which is the required result. ||

Theorems 2 and 3 follow immediately from Theorem 1.

First version received October 1979; *final version accepted May* 1981 (*Eds.*).

Lemma 2 was reported by Satterthwaite (1976); it provided the starting point for the present analysis. We are greatly indebted to Donald Brown, who suggested that the serial dictatorship of binary and Pareto social welfare functions (see Luce and Raiffa (1957, p. 344)) might have a parallel in the present framework. This led to our formulation of our main result: Theorem 1. The proof of Lemma 3 was suggested by William Novshek, who pointed out that an early formal argument was inadequate. Carl Simon helped us to make the notion of a regular mechanism precise. Ehud Kalai pointed out that in the definition of regularity, we could not assume the dimensionality of $B_j(u)$ to be $l-1$. This committee was rounded out by Salvador Barbera, whose criticism and deep understanding helped to improve the exposition of the result. Finally, we gratefully acknowledge the research support of the NSF and, in Satterthwaite's case, the J. L. Kellogg Graduate School of Management's Center for Advanced Study in Managerial Economics and Decision Sciences.

NOTES

1. This mechanism is also Pareto efficient.
2. We note two objections to their findings. First, the claimed result is incorrect when there are more than two agents. To see this let the first agent get all of the endowment when the third agent's marginal rate of substitution at (1, 1) is less than unity, and let the second agent get all of the endowment if the third agent's marginal rate of substitution at (1, 1) is greater than or equal to unity. Clearly truth is a dominant strategy and there is no dictator. The third agent is a dictator-maker, and the final two agents alternate as dictators. Second, the assumption that the mechanism must choose Pareto efficient allocations even when preferences are not continuous is very strong. For some standard convex and compact allocation possibility sets A, the set of Pareto efficient single-valued social choice rules is empty, and thus all such mechanisms are dictatorial whether or not they are strategy-proof! To verify this assertion, observe that there exist strictly convex and strictly monotone preference relations over R^2_+ for which there is no best element in the set $A = \{(x_1, x_2) | x + x_2 \leqq 1\}$.
3. In defining broad applicability, we do not require strict concavity, strict monotonicity, or other economically relevant properties in addition to the required openness of U because the theorems that follow are proved using the openness property only.
4. Note that $B_i(u)$ is independent of u_i.
5. Since σ is strategy-proof, $\sigma_i(u)$ maximizes u_i on $B_i(u)$. For $\sigma_i(u)$ to be a regular maximizer or u_i on $B_i(u)$, we must have (i) for all i, the gradient of u_i evaluated at $\sigma_i(u)$ does not vanish and (ii) its relevant bordered Hessian does not vanish.
6. A relation Q on S is total if for all $s, t \in S$ either sQt, tQs, or $s = t$.
7. For example, the mechanism defined on n-tuples of single-peaked preferences (n odd) over an interval that picks the median individual's peak does not give rise to an $\mathcal{R}$ that is connected. But note that for most u, the conditions of Theorem 1 are satisfied and $A(u)$ is acyclic since only one agent (the median agent) affects the choice.
8. This can be restated in Gibbard's terminology: a straightforward game form exists within a given environment if and only if a nonmanipulable voting scheme exists within the environment.

9. The set of all concave C^2 utility functions is not broadly applicable because that set is not open; linear utility functions, which are concave, may be C^2 perturbed an arbitrarily small amount with the consequence that they are no longer concave. The set of strictly concave C^2 utility functions is, however, open and, thus, satisfies broad applicability.

10. As we have already noted, non-bossiness is trivially satisfied for the case of a pure public goods economy since a change in any one agent's allocation means, by definition, an identical change in every other agent's allocation. Thus, for example, non-bossiness is present in practically all of the literature on social choice.

11. The reason is simple. If a mechanism is first order, then the only way agent i can change agent j's allocation is to change his marginal rate of substitution, which for first order mechanisms means his own allocation also changes. But bossiness requires that i be able to change j's allocation without changing his own allocation. Therefore, bossiness is incompatible with being a first order mechanism.

12. Non-differentiability may also be important.

13. The utility function $w \in U$ may be constructed by picking a function $v_w \in C^2(X)$ with the properties

$$\nabla(\bar{u}_2 + v_w) = \nabla(\bar{u}_2 + \lambda v_2)$$

when evaluated at $\sigma_2(\bar{u}_1, \bar{u}_2 + \lambda v_2)$ and

$$\nabla(\bar{u}_2 + v_w) = \nabla \bar{u}_2$$

when evaluated at $\sigma_2(\bar{u}_1 + \lambda v_1, \bar{u}_2)$ and then defining $w \equiv \bar{u}_2 + v_w$. For λ small enough and judicious choice of v_w, the resulting w will be an element of U, since U is open.

REFERENCES

CLARKE, E. H. (1971), "Multipart Pricing of Public Goods", *Public Choice*, **8**, 19–33.

DASGUPTA, P., HAMMOND, P. and MASKIN, E. (1979), "The Implementation of Social Choice Rules", *Review of Economic Studies*, **44**, 185–216.

GIBBARD, A. (1973), "Manipulation of Voting Schemes: A General Result", *Econometrica*, **41**, 587–602.

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

GREEN, J. and LAFFONT, J.-J. (1979) *Incentives in Public Decision-Making* (Amsterdem: North-Holland).

GROVES, T. (1973), "Incentives in Teams", *Econometrica*, **41**, 587–602.

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

HURWICZ, L. (1972), "On Informationally Decentralized Systems", McGuire, B. and Radner, R. (eds.) *Decision and Organization* (Amsterdam: North-Holland Press) 297–336.

KALAI, E. and MULLER, E. (1977), "Characterization of Domains Admitting Nondictatorial Social Welfare Functions and Nonmanipulable Voting Procedures", *Journal of Economic Theory*, **16**, 457–469.

LUCE, R. D. and RAIFFA, H. (1957) *Games and Decisions* (New York: Wiley).

MASKIN, E. (1976*a*), "On Strategy Proofness and Social Welfare Functions on Restricted Domains" (mimeo, Cambridge University).

MASKIN, E. (1976*b*), "Social Welfare Functions on Restricted Domains" (mimeo, Cambridge University).

SAMUELSON, P. A. (1954), "The Pure Theory of Public Expenditure", *Review of Economics and Statistics*, **36**, 387–389.

SATTERTHWAITE, M. A. (1975), "Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions", *Journal of Economic Theory*, **10**, 187–217.

SATTERTHWAITE, M. A. (1976), "Straightforward Allocation Mechanisms" (Discussion Paper No. 253, Center for Mathematical Studies in Economics and Management Science, Northwestern University).

SATTERTHWAITE, M. A. and SONNENSCHEIN, H. (1979), "Technical Note to 'Strategy-Proof Allocation Mechanisms'" (Discussion Paper No. 396, Center for Mathematical Studies in Economics and Management Science, Northwestern University).