STRATEGY PROOFNESS OF VOTING PROCEDURES WITH LOTTERIES AS OUTCOMES AND INFINITE SETS OF STRATEGIES

Aanund Hylland[1]

January, 1980

ABSTRACT

A decision procedure is strategy proof if reporting one's preferences truthfully is always an optimal action. In the procedures discussed here, the final outcome may partly be determined by chance; hence people's voting determines a lottery (on a finite set of basic alternatives). Individual preferences are represented by (von Neumann-Morgenstern) utility functions. If a strategy-proof procedure satisfies the ex post Pareto condition (the final, non-random outcome is Pareto optimal), it is a "random dictatorship." If the ex ante Pareto condition holds (the chosen lottery is Pareto optimal), the procedure is dictatorial.

1. Introduction

Do there exist reasonable voting procedures in which nobody can ever gain by strategic voting? It has long been suspected that the answer is no; the only strategy-proof procedures are the "dictatorial" ones in which one participant unilaterally decides the outcome. Vickrey [24, p. 518] and Dummett and Farquharson [8, p. 34] made precise conjectures to this effect.[2] A formal theorem was proved independently by Gibbard [12] and Satterthwaite [22]: If there are at least three possible outcomes, then every strategy-proof voting procedure is dictatorial.[3]

The Gibbard-Satterthwaite theorem applies only to deterministic and single-valued procedures. That is, it is assumed that when the participants' votes are given, the outcome is unique and can be found by a purely computational process. If this condition is relaxed and chance is allowed to play a role in the decision procedure, there exist strategy-proof procedures that are not dictatorial.[4]

In order to discuss the possibilities of strategic voting in such procedures, some assumptions must be made about people's preferences for risky prospects. Following the tradition of economic theory, it will be assumed that preferences over lotteries can be represented by (von Neumann-Morgenstern) utility functions. That is, a utility value is assigned to each non-random outcome, and one lottery is preferred to another if and only if it has a higher expected utility.[5] Even if the set of non-random outcomes is finite, there exist infinitely many different preference relations of this type.

In most of this paper, only decision procedures that guarantee Pareto-optimal outcomes will be considered. That is, there shall exist no possible decision which, in everybody's opinion, is better than the actual outcome.

This condition can be formulated in two different ways; the requirement can be imposed on the final outcome after all uncertainty has been resolved (the ex post condition), or it can be applied to the lottery resulting from the voting (the ex ante condition). The latter condition is the stronger one. (For precise definitions, see Section 2.) If one looks for "reasonable" strategy-proof procedures, it seems to me that there can be no loss of generality in imposing the Pareto condition; it can hardly be reasonable to choose an outcome which would be unanimously voted down in a pairwise vote with some other possible decision.[6]

Presumably, the principal objection to a dictatorial voting procedure is that it distributes influence in an extremely uneven way. The introduction of chance can mitigate this problem, as shown by the following procedure: Choose one individual by lottery, and let this individual decide the outcome. When preferences are of the type considered here, this procedure is strategy proof. In fact, a whole class of procedures has been defined, since the participants' probabilities of becoming dictator need not be equal; any predetermined probability distribution on the set of individuals defines a strategy-proof procedure.[7] These procedures satisfy the ex post Pareto condition.

When there are at least three possible non-random outcomes, the procedures described in the previous paragraph are the only ones which are strategy proof and satisfy the ex post Pareto condition. This is the main result of this paper (Theorem 1). In fact, the ex post Pareto condition can be weakened somewhat without changing the conclusion; see Theorem 1*. If the ex ante Pareto condition is imposed, only non-random dictatorial procedures are strategy proof (Theorem 2).

In two important papers [13, 14], Gibbard has proved similar results. (He also proves a characterization theorem for all strategy-proof procedures, without imposing the Pareto condition.) The premise of Gibbard's theorems contains the condition that each individual's set of voting strategies be finite. Since there are infinitely many possible preference relations, any procedure which is "unique" in the sense that different preferences lead to different voting behavior, must violate the premise and fall outside the scope of the theorems. Zeckhauser [25, p. 942, Theorem II] demonstrates that uniqueness is important for achieving Pareto optimality. As pointed out by Gibbard [14, p. 598, f.n. 6], procedures in which voting consists in reporting a preferences relation, are not covered by his theorems. Hence the finiteness condition seems to be important and to imply a significant weakening of the results.[8] In this paper, the characterization theorems for strategy-proof procedures satisfying the Pareto condition are strengthened by the removal of this condition.

The paper is organized in this way: In Section 2, the model is described in detail, and the necessary concepts are defined. Both formal and informal definitions are given. Section 3 contains the principal theorems and sketches of proof; detailed proofs are given in Appendices A and B. Some generalizations are presented in Sections 4 and 5. In Section 6, an alternative concept of strategy proofness is considered; this concept is generally weaker than the one used in the main theorems. Some concluding remarks are made in Section 7.

2. The Model

So far, it has not been made clear what a "voting procedure" is or what "strategy proofness" means. These and other concepts are defined and discussed in this section. (Additional definitions are introduced, as needed, in later sections.) In many respects, the notation follows Gibbard [13, 14].

There is a fixed set of basic, non-random outcomes, exactly one of which must eventually be chosen. This set is called A; elements of A are called alternatives and denoted x, y, z and w. Subsets of A are denoted B and C. In the main theorems, it will be assumed that A is finite and has at least three elements.

There are finitely many individuals who participate in the decision; they will be denoted $1,2,\ldots,n$, where n is a positive integer. The set of individuals is called N; hence $N = \{1,2,\ldots,n\}$. The letters i, j and k will refer to elements of N, while I, J and K denote subsets of N. For $I \subseteq N$, $\tilde{I}$ is the complement of I relative to N; that is, $\tilde{I} = N \setminus I$.

In the procedures to be studied in this paper, the outcome is determined partly by the participants' behavior (loosely referred to as their votes), and partly by random events. Such procedures can be organized in many ways. Consider, for example, the "random dictatorship" mentioned in the Introduction. As described there, the procedure consisted in first choosing an individual randomly, and then letting this person decide the outcome. Assuming that all individuals have an equal probability of being chosen, an alternative procedure can be described in this way: Ask each person to vote, by secret ballot, for one alternative; then choose one ballot by lottery and let it decide the issue. In the formal model, these procedures are equivalent. If people have preferences of the type assumed here and act accordingly, the probability of any alternative being chosen is the same in the two procedures. In practical applications, there may be reasons for choosing one version rather than the other. For example, the first version requires less transmission of information; only the "dictator" need vote. On the other hand, the secret ballot can be protected in the second version; the identity of the dictator can also be kept secret, which may be desirable. But these considerations fall outside the model.

By the same line of argument, there is no loss of generality in assuming that all procedures are of the following form: First, each individual delivers a vote. They do so simultaneously, and there is only one round of voting. These votes are the input to a computational process, the output of which is an assignment of probabilities to the elements of A. Finally, some random process is employed to choose one element of A, in such a way that the different elements get the right probabilities of being chosen. Suppose that there is given a procedure of a more complicated kind; for example, the voting may be divided into many rounds, and lotteries may be conducted at several stages along the way. Such a procedure can be converted to the form described above. A "vote" in the converted procedure is a complete strategy for voting in the original procedure; it must cover all the rounds, and in the later rounds it must describe the voting behavior as a function of the results of earlier lotteries and votes. When these votes are in, the original procedure can be simulated; if an intermediate lottery is encountered, the probabilities of all possible results of the lottery are computed, and the simulation is continued from each of these results. This simulation defines the computational part of the converted procedure.

Therefore, all procedures are assumed to take the form that the participants' votes determine a lottery on A. Formally, a <u>lottery</u> ρ is a real-valued function defined on A and satisfying $\rho(x) \geq 0$ for all $x \in A$ and $\sum_{x \in A} \rho(x) = 1$. The Greek letters ρ and σ will denote lotteries. For any lottery ρ, $P(\rho)$ denotes the set of alternatives with positive probability in ρ, that is, $P(\rho) = \{x \in A | \rho(x) > 0\}$. Expressions like "$\rho$ is a lottery on B" or "ρ is a lottery in x and y" will sometimes be used. The meaning should be clear: Only elements of B (or only x and y) get positive probability.

An individual's preferences are represented by a <u>utility scale</u>, which is a real-valued function defined on A. The letters u and v, possibly with subscripts, primes, etc., will denote utility scales. The set of alternatives with maximal value according to u will be denoted T(u). That is, $x \in T(u)$ if and only if $u(x) \geq u(y)$ for all $y \in A$. When A is finite, T(u) is non-empty for every u. The scale u is said to be <u>strict</u> if T(u) has one element, that is, if there exists an $x \in A$ such that $u(x) > u(y)$ for all $y \in A$ with $y \neq x$. Both lotteries and utility scales can be viewed as finite-dimensional real vectors. Hence the inner product of a utility scale u and a lottery ρ can be defined by $u \cdot \rho = \sum_{x \in A} u(x)\rho(x)$. The number $u \cdot \rho$ is referred to as the utility of ρ according to u; the individual prefers ρ to σ if and only if $u \cdot \rho > u \cdot \sigma$. For each $x \in A$, $\hat{x}$ will denote the unit vector in the direction x, that is, $\hat{x}$ is a real-valued function defined on A, $\hat{x}(x) = 1$, and $\hat{x}(y) = 0$ for $y \neq x$. The vector $\hat{x}$ has several interpretations. It is the degenerate lottery which assigns probability 1 to x; alternatively, it is a utility scale in whose underlying preferences x is good while all other elements of A are equally bad. For a lottery ρ, $\hat{x}\cdot\rho = \rho(x)$ is the probability assigned to x by ρ.

A <u>utility profile</u> is an n-tuple $(u_1,\ldots,u_n)$ of utility scales. The symbols $\underline{u}$, $\underline{v}$, $\underline{u}'$, etc. will denote utility profiles, where $\underline{u} = (u_1,\ldots,u_n)$, $\underline{u}' = (u_1',\ldots,u_n')$ and so on. The profile $\underline{u}$ is <u>strict</u> if u_i is strict for every i. The utility profile obtained by substituting v_i for u_i in $\underline{u}$ is denoted $v_i\underline{u}_{-i}$.

When constructing a decision procedure, one must decide which form the individuals' votes shall take. In general, the votes need not take the same form for all the participants. The set from which individual i's vote shall be chosen will be denoted S_i. An element of S_i will be referred to as a <u>strategy</u> for i; earlier, informal expressions like "votes," "voting behavior," etc. have been used. A <u>strategy profile</u> is an element of $S_1 \times \ldots \times S_n$, that is, it is an n-tuple consisting of one strategy for each individual. Strategies for i are denoted s_i, s_i', t_i, etc. and strategy profiles are denoted $\underline{s}$, $\underline{s}'$, $\underline{t}$, etc.; then s_i is the ith coordinate of $\underline{s}$, and so on. The strategy profile obtained by substituting t_i for s_i in $\underline{s}$ is denoted $t_i\underline{s}_{-i}$.

The most general class of voting procedures to be discussed is the class of <u>game forms</u>, defined in this way: A function g is a game form if there exist non-empty sets $S_1,\ldots,S_n$ such that the domain of definition for g is $S_1 \times \ldots \times S_n$, and $g(\underline{s})$ is a lottery for every $\underline{s} \in S_1 \times \ldots \times S_n$. The definition implies that in order to describe a game form, one has to specify both the strategy sets S_i and the function g.[9]

For a given game form, there is no guarantee that all elements of A can actually be chosen. For example, g may be a constant function. The requirement may be imposed that the individuals, at least if they cooperate, shall be able to obtain any alternative with certainty. This is a kind of citizens' sovereignty condition. Formally, the game form g satisfies the attainability condition if, for each $x \in A$, there exists an $\underline{s} \in S_1 \times \ldots \times S_n$ such that $g(\underline{s}) = \hat{x}$.

The concept of "strategy proofness" will now be discussed. Let a game form g with strategy sets $S_1,\ldots,S_n$ be given. For a utility scale u and a strategy $s_i \in S_i$, s_i is said to be u-dominant for i in g if $u \cdot g(s_i \underline{t}_{-i}) \geq u \cdot g(\underline{t})$ for all strategy profiles $\underline{t}$. A strategy profile $\underline{s}$ fits a utility scale $\underline{u}$ if, for every i, s_i is u_i-dominant for i. The game form g is straightforward if, for all i and all u, there exists a strategy $s_i \in S_i$ such that s_i is u-dominant for i. This is the strategy proofness condition which will be used in the main results in Sections 3 and 4.

If individual i's preferences are represented by u and s_i is u-dominant for i, then s_i is an optimal action for i no matter what the other individuals do. This means that straightforwardness is a relatively strong condition.[10] Essentially, it is required that person i's choice of strategy be based only on i's own preferences and neither on knowledge nor on probabilistic expectations about other people's preferences. The informal concept of strategy proofness can be made precise in other ways, in general leading to weaker conditions. The issue is discussed in Sections 6 and 7.

When straightforwardness is required, the definition of a game form is unnecessarily general. There is essentially no loss of generality in assuming that an individual casts a vote by reporting a utility scale, and that revealing one's preferences correctly is always a dominant strategy. This is shown in the corollaries in Section 4, and it is also a consequence of more general results.[11] To formalize this idea, let a *decision scheme* be a game form each of whose strategy sets is equal to the set of utility scales. That is, a decision scheme is a function f defined on the set of all utility profiles such that, for every utility profile $\underline{u}$, $f(\underline{u})$ is a lottery.[12] For a given utility profile $\underline{u}$ and an individual i, i *can manipulate* f at $\underline{u}$ if there exists a utility scale v_i such that $u_i \cdot f(v_i \underline{u}_{-i}) > u_i \cdot f(\underline{u})$. The decision scheme f is *strategy proof* if there exist no $\underline{u}$ and i such that i can manipulate f at $\underline{u}$.[13] The definitions imply that f is strategy proof if and only if, for all utility scales u and all individuals i, u is u-dominant for i. If f is strategy proof then f is straightforward, but the converse need not hold, since there may exist u, v and i with $u \neq v$ such that v is u-dominant for i.

For a given utility profile $\underline{u}$ and alternatives x and y, x is said to *dominate* y in $\underline{u}$ (or y is dominated by x in $\underline{u}$) if $u_i(x) > u_i(y)$ for all i. The alternative x is *ex post Pareto optimal* for $\underline{u}$ if there exists no alternative that dominates x in u. The set of ex post Pareto-optimal alternatives for $\underline{u}$ is denoted $PO(\underline{u})$; formally, $x \in PO(\underline{u})$ if and only if for every $y \in A$ there exists an $i \in N$ such that $u_i(x) \geq u_i(y)$. A lottery ρ is *ex post Pareto optimal* for $\underline{u}$ if $P(\rho) \subset PO(\underline{u})$, that is, if every alternative that can be chosen by the lottery ρ is ex post Pareto optimal.

Similarly, for a profile $\underline{u}$ and lotteries ρ and σ, ρ dominates σ in $\underline{u}$ if $u_i \cdot \rho > u_i \cdot \sigma$ for all i. The lottery ρ is *ex ante Pareto optimal* for $\underline{u}$ if there exists no lottery that dominates ρ in $\underline{u}$. As will be proved formally in Section 3, this is a stronger condition than ex post optimality.

A decision scheme f satisfies the *ex post* (*ex ante*) *Pareto condition* if, for all utility profiles $\underline{u}$, $f(\underline{u})$ is ex post (ex ante) Pareto optimal for $\underline{u}$.

The dominance relations defined above are the strong relations; strict inequality $u_i(x) > u_i(y)$ or $u_i \cdot \rho > u_i \cdot \sigma$ is required for every i. This leads to weak versions of the Pareto condition. The alternative is to say that x dominates y if $u_i(x) \geq u_i(y)$ for all i, with strict inequality for at least one i, and similarly for dominance among lotteries. From these definitions, stronger Pareto conditions would result. The weak conditions are sufficient to prove the main theorems; therefore, there is no reason to weaken the theorems by using the strong versions.

What should be meant by the statement "individual i is a dictator for the decision scheme f"? If person i has a unique top-ranked alternative (that is, if i's preferences are represented by a strict utility scale), dictatorship must obviously imply that this top-ranked alternative is chosen with certainty. On the other hand, if two or more alternatives are tied at the top of i's ranking, several interpretations are possible. The definition chosen here only requires that the outcome be a lottery over i's top-ranked alternatives. Any such lottery will satisfy the dictatorship condition, and the particular outcome may depend on the utility scales of individuals other than i. In other words, if i is a dictator, then i is guaranteed an optimal outcome but does not necessarily control the choice of one lottery from among those which are optimal for i. Of course, if i does have a unique top-ranked alternative, there is no such choice to be made, and i controls the outcome.

The formal definition is the following: A decision scheme f is <u>dictatorial</u> for i if $P(f(\underline{u})) \subseteq T(u_i)$ for every utility profile $\underline{u}$. A scheme is dictatorial if it is dictatorial for some i.

For some positive integer m, let $f_1, \ldots, f_m$ be decision schemes, and let $a_1, \ldots, a_m$ be nonnegative numbers with sum 1. If f is defined by

$f(\underline{u}) = a_1 f_1(\underline{u}) + \ldots + a_m f_m(\underline{u})$ for all $\underline{u}$, then f is a decision scheme. When this equation holds, f is said to be a <u>probability mixture</u> of $f_1, \ldots, f_m$, and f is written $f = a_1 f_1 + \ldots + a_m f_m$. A similar definition can be made for game forms.

3. Main Theorems

In this section, it is assumed that the set A of alternatives is finite and has at least three elements, and that the set N of individuals is non-empty and finite.

Theorem 1: If the decision scheme f is strategy proof and satisfies the ex post Pareto condition, then f is a probability mixture of dictatorial decision schemes.

The theorem is proved in Appendix A, Lemmas A1 - A9. Only a brief outline of the proof is given here. The desired conclusion is that there exist nonnegative numbers $a_1,\ldots,a_n$ with sum 1 and decision schemes $f_1,\ldots,f_n$, such that each f_i is dictatorial for i, and

$$f = a_1 f_1 + \ldots + a_n f_n \quad . \tag{1}$$

Of course, some number a_i may be 0, in which case f_i is irrelevant. The proof starts by considering utility profiles of the following special type: Let I be a subset of N, and let x and y be distinct alternatives. Suppose that $\underline{u}$ satisfies $T(u_i) = \{x\}$ for $i \in I$, $T(u_j) = \{y\}$ for $j \in \tilde{I}$, and x dominates any $z \notin \{x,y\}$ in $\underline{u}$. The Pareto condition shows that $f(\underline{u}) = a\hat{x} + (1-a)\hat{y}$ for some number a. The number a turns out to be the same for all profiles $\underline{u}$ of the type just described. Moreover, a does not depend on the particular alternatives x and y; it is a function solely of the set I and can be denoted a(I). This function satisfies $a(\tilde{I}) = 1 - a(I)$, $a(\phi) = 0$, and $a(N) = 1$. The next step is to consider profiles $\underline{u}$ with $T(u_i) = \{x\}$ for $i \in I$; nothing is now assumed about u_j for $j \in \tilde{I}$. The inequality $\hat{x} \cdot f(\underline{u}) \geq a(I)$ can be established. Then it is possible to prove that the function a is a probability measure on N. That is, the numbers $a(\{i\})$ are nonnegative, they sum to 1, and $a(I) = \sum_{i \in I} a(\{i\})$. For every i, the number a_i of (1) will be equal to $a(\{i\})$. From the facts now stated, it is easy to prove that (1) holds on all strict utility profiles. (When $\underline{u}$ is strict, the condition that f_i is dictatorial for i determines $f_i(\underline{u})$ uniquely.) For profiles $\underline{u}$ which are not strict, the duality theorem for linear programming can finally be used to find lotteries $f_i(\underline{u})$ such that (1) holds.

The theorem does not say that the number a_i and the scheme f_i of (1) are uniquely determined by f. The number a_i is, however, uniquely determined by f. For a given individual i and distinct alternatives x and y, let $\underline{u}$ satisfy $T(u_i) = \{x\}$ and $T(u_j) = \{y\}$ for $j \neq i$. Then (1) requires $f(\underline{u}) = a_i\hat{x} + (1-a_i)\hat{y}$, which shows that f determines a_i. Whenever u_i is a strict utility scale, the condition that f_i is dictatorial for i leaves only one possible value for $f_i(\underline{u})$. If $a_i > 0$ and u_j is strict for all $j \neq i$, only one value of $f_i(\underline{u})$ can satisfy (1). Therefore, only when $T(u_i)$ has two or more elements for two or more values of i can equation (1) fail to determine the lotteries $f_i(\underline{u})$ uniquely. (Here the possibility $a_i = 0$ is ignored.)

Strictly speaking, Theorem 1 does not characterize the class of strategy-proof decision schemes satisfying the ex post Pareto condition. The reason is that a scheme f of the form (1) need not be strategy proof. (Such an f will, however, satisfy the Pareto condition.) If $i \neq j$ and u_i is not strict, then $f_i(\underline{u})$ may depend on u_j in a way which contradicts strategy proofness.

A complete characterization must probably be quite complicated. One difficulty lies in the fact that f can be strategy proof without each f_i having this property; even if some individual i can manipulate f_j at a certain profile, i's gain in f_j from incorrect reporting of preferences may be more than offset by the effect on f_k for $k \neq j$. The converse implication does hold; it is easy to prove that a probability mixture of strategy-proof schemes is a strategy-proof scheme. Therefore, for any nonnegative numbers $a_1,\ldots,a_n$ with sum 1, there exist decision schemes of the form (1) that satisfy the premise of Theorem 1. In particular, such schemes can be constructed by letting each f_i be strategy proof. (It is always assumed that f_i is dictatorial for i.) One possibility is to let f_i depend only on i's utility scale; for example, $f_i(\underline{u})$ can be the even-chance lottery over $T(u_i)$, or it can be the degenerate lottery corresponding to the alternative in $T(u_i)$ which is first in some fixed ordering of A. Another possible f_i is the "serial dictatorship," defined as follows: Let $i_1 = i, i_2, \ldots, i_n$ be a fixed ordering of the individuals. For a given $\underline{u}$, define $T_1 = T(u_i)$, let T_2 be the elements of T_1 which have maximal utility in u_{i_2}, etc. Formally,

$$x \in T_k \text{ if and only if } x \in T_{k-1} \text{ and } u_{i_k}(x) \geq u_{i_k}(y) \text{ for all } y \in T_{k-1}.$$

Finally, $f_i(\underline{u})$ shall be some lottery over T_n. This scheme, and any probability mixture of such schemes, will also satisfy the strong ex post Pareto condition.

If decision schemes are defined only on the set of strict profiles, the remarks in the last paragraph do not apply. On this domain, any probability mixture of dictatorial schemes is strategy proof, and Theorem 1 provides a complete characterization of the schemes considered there. Moreover, the theorem shows that when f satisfies the premise and $\underline{u}$ is strict, then $f(\underline{u})$ depends only on the individuals' top-ranked alternatives. In particular, $f(\underline{u})$ depends only on the individual preferences and not on the particular utility scales chosen to represent them.[14] Any decision scheme covered by the theorem is also neutral on this domain, in the sense that alternatives are treated equally.[15]

The assumption that A has at least three elements is obviously necessary. If there are only two alternatives, simple majority vote, with ties broken by lottery, is a counterexample to Theorem 1.

A decision scheme which satisfies the ex post Pareto condition must also satisfy the attainability condition: For any $x \in A$, let $\underline{u}$ be a profile with $T(u_i) = \{x\}$ for all i. Then $PO(\underline{u}) = \{x\}$, the Pareto condition requires $f(\underline{u}) = \hat{x}$, and attainability is proved. The converse implication does not hold; for arbitrary decision schemes attainability is a weaker condition than ex post Pareto optimality. Theorem 1 remains true, however, even if this weaker condition is used.

Theorem 1*: If the decision scheme f is strategy proof and satisfies the attainability condition, then f is a probability mixture of dictatorial decision schemes.

The proof is given in Appendix A; Lemmas A10 - A13 reduce the proof of Theorem 1* to that of Theorem 1. Comments made above to Theorem 1 also apply to Theorem 1*.

The ex ante Pareto condition is really stronger than the ex post condition, as shown by the following theorem.

Theorem 2: If the decision scheme f is strategy proof and satisfies the ex ante Pareto condition, then f is dictatorial.

Below, Theorem 2 is proved as a corollary to Theorem 1. An alternative, direct proof is given in Appendix B. The latter is considerably shorter and simpler than the proof of Theorem 1.

Proof that Theorem 2 follows from Theorem 1: Suppose that f satisfies the premise of Theorem 2. The first step is to prove the ex post Pareto condition. If this condition does not hold, there exist $\underline{u}$, ρ, x and y such that $f(\underline{u}) = \rho$, $\rho(x) > 0$, and $u_i(y) > u_i(x)$ for all i. Let $\sigma = \rho - \rho(x)\cdot\hat{x} + \rho(x)\cdot\hat{y}$; then σ is a lottery and $u_i\cdot\sigma > u_i\cdot\rho$ for all i. This contradicts the ex ante Pareto condition and proves the assertion.

Theorem 1 now implies that f is of the form (1). Obviously, $0 \leq a_i \leq 1$ for all i. Suppose some a_i satisfies $0 < a_i < 1$. Let x, y and z be three distinct alternatives, and find a utility scale $\underline{u}$ which satisfies the following: For the given i, $u_i(x) = 1$, $a_i < u_i(z) < 1$, $u_i(y) = 0$, and $u_i(w) < 0$ for $w \notin \{x,y,z\}$. For $j \neq i$, $u_j(y) = 1$, $1 - a_i < u_j(z) < 1$, $u_j(x) = 0$, and $u_j(w) < 0$ for $w \notin \{x,y,z\}$. Equation (1) gives $f(\underline{u}) = a_i\cdot\hat{x} + (1-a_i)\cdot\hat{y}$, which implies $u_i(z) > a_i = u_i\cdot f(\underline{u})$, and $u_j(z) > 1 - a_i = u_j\cdot f(\underline{u})$ for all $j \neq i$. Hence $f(\underline{u})$ is dominated by the lottery $\hat{z}$, contradicting the ex ante Pareto condition.

Therefore, each a_i is either 0 or 1. Since $a_1 + \ldots + a_n = 1$, this implies the existence of an i such that $a_i = 1$ and $a_j = 0$ for $j \neq i$. Hence $f = f_i$, f is dictatorial for i, and the proof is complete.[16] ||

As was the case for Theorem 1, Theorem 2 does not characterize the class of decision schemes satisfying the premise. When f is dictatorial for i, there may exist $j \neq i$ and $\underline{u}$ such that j can manipulate f at $\underline{u}$. If only strict profiles are allowed, this is not possible, and the theorem provides a complete characterization of the schemes under consideration.

On the domain of strict utility profiles, a dictatorial scheme is also, in a trivial way, coalitionally strategy proof. By definition, f is coalitionally strategy proof if there do not exist strategy profiles $\underline{u}$ and $\underline{v}$ and a set I of individuals such that $v_j = u_j$ for all $j \in \tilde{I}$ and $u_i \cdot f(\underline{v}) > u_i \cdot f(\underline{u})$ for all $i \in I$. That is, there exists no situation in which a group of individuals, by a coordinated change in their strategies, can achieve a change in the outcome from which they all benefit. If coalitional strategy proofness is added to the premise of Theorem 1, then f must be dictatorial. This follows from the proof above of Theorem 2: Instead of using ex ante Pareto optimality to rule out the possibility $0 < a_i < 1$, coalitional strategy proofness could have been used, with $I = N$. The same conclusion follows even if coalitional strategy proofness is assumed to hold only when the coalition I has two members. The proof is easy: Suppose that (1) holds, and assume $a_i > 0$ and $a_j > 0$ for distinct individuals i and j. Let x, y and z be different alternatives, and let $\underline{u}$ be a strict profile such that $u_i(x) = u_j(y) = 1$, $a_i/(a_i+a_j) < u_i(z) < 1$, $a_j/(a_i+a_j) < u_j(z) < 1$, $u_i(y) = u_j(x) = 0$, and $u_i(w) < 0$ and $u_j(w) < 0$ for $w \notin \{x,y,z\}$. Individuals i and j will gain if they both report preferences in which z is the unique top-ranked alternative.

The conclusion of this section is that a strategy-proof decision scheme that satisfies the Pareto condition must be of a very special form. Decision schemes of this form will probably not be considered reasonable or desirable; see further discussion in Section 7.

4. Arbitrary Sets of Strategies

In Section 3, only decision schemes were considered. That is, the theorems apply only to procedures in which voting consists in reporting a utility scale. Moreover, the strategy proofness condition requires that reporting one's preferences correctly always be an optimal action. The principal result of this section is that there is nothing to be gained by relaxing these conditions and considering arbitrary straightforward game forms. The conditions on the cardinality of A and N, introduced in Section 3, are still in effect.

Let g be a game form with strategy sets $S_1,\ldots,S_n$. For a given individual i, a strategy selection for i is a function from the set of utility scales into S_i; that is, it is a rule which chooses a strategy from S_i for each possible preference representation.[17] Strategy selections for i are denoted $\tilde{s}_i$ and $\tilde{t}_i$. The function $\tilde{s}_i$ is a dominant strategy selection for i if, for every utility scale u, $\tilde{s}_i(u)$ is u-dominant for i. Symbols like $\underline{\tilde{s}}$ and $\underline{\tilde{t}}$ will denote functions defined on utility profiles with strategy profiles as values, such that $\underline{\tilde{s}}(\underline{u}) = (\tilde{s}_1(u_1),\ldots,\tilde{s}_n(u_n))$, etc. Such a function is called a selection profile. If each $\tilde{s}_i$ is dominant for i, $\underline{\tilde{s}}$ is also said to be dominant. Note that $\underline{\tilde{s}}$ is not a general function of $\underline{u}$, since the ith coordinate of $\underline{\tilde{s}}(\underline{u})$ depends only on u_i. When $\underline{\tilde{s}}$ is given, a function f can be defined by $f(\underline{u}) = g(\underline{\tilde{s}}(\underline{u}))$ for all utility profiles $\underline{u}$. The function f is a decision scheme, and it will be denoted $g \circ \underline{\tilde{s}}$.

Suppose that g is straightforward. Then there exists a dominant strategy selection for every individual. Let $\underline{\tilde{s}}$ be a dominant selection profile. The decision scheme $g \circ \underline{\tilde{s}}$ is strategy proof. This follows directly from the definitions. (It is essential that the ith coordinate of $\underline{\tilde{s}}(\underline{u})$ depends only on u_i.) If $g \circ \underline{\tilde{s}}$ satisfies the ex post Pareto condition, Theorem 1 can be used to conclude that $g \circ \underline{\tilde{s}}$ is a probability mixture of dictatorial schemes. Dominant strategies need not be unique. Hence there can exist a dominant selection profile $\underline{\tilde{t}}$ different from $\underline{\tilde{s}}$. As it turns out, $g \circ \underline{\tilde{s}}$ and $g \circ \underline{\tilde{t}}$ are almost equal, in the sense that the numbers $a_1,\ldots,a_n$ of equation (1) are the same for the two schemes. To derive these results, it is not necessary to assume that $g \circ \underline{\tilde{s}}$ and $g \circ \underline{\tilde{t}}$ satisfy the Pareto condition; a weaker version of Pareto optimality will suffice.

Corollary 1: Let g be a straightforward game form. Assume that for each utility profile $\underline{u}$ there exists a strategy profile $\underline{s}$ such that $\underline{s}$ fits $\underline{u}$ and $g(\underline{s})$ is ex post Pareto optimal for $\underline{u}$. Then there exist nonnegative numbers $a_1,\ldots,a_n$ with sum 1 such that the following holds: For any $\underline{u}$ and any $\underline{s}$ that fits $\underline{u}$, there exist lotteries $\rho_1,\ldots,\rho_n$ such that $g(\underline{s}) = a_1\rho_1 + \ldots + a_n\rho_n$, and $P(\rho_i) \subseteq T(u_i)$ for all i.

The $\underline{s}$ of the premise may depend on $\underline{u}$ in an arbitrary way; hence the condition is weaker than a requirement that some decision scheme $g \circ \underline{\tilde{s}}$ satisfy the Pareto condition. This also has the following consequence: Define a decision scheme f by $f(\underline{u}) = g(\underline{s})$, where $\underline{s}$ is some strategy profile such that $\underline{s}$ fits $\underline{u}$ and $P(g(\underline{s})) \subseteq PO(\underline{u})$. Then f need not be strategy proof. Therefore, the proof of the corollary cannot rely on this f.

Proof: The first step is to prove that $g(\underline{s})$ is ex post Pareto optimal for $\underline{u}$ whenever $\underline{s}$ fits $\underline{u}$. Let $\underline{u}$, $\underline{s}$ and x be such that $\underline{s}$ fits $\underline{u}$ and $x \notin PO(\underline{u})$. Then there exists a y such that $u_i(x) < u_i(y)$ for all i. A utility profile $\underline{v}$ is constructed as follows: For all i, $v_i(x)$ satisfies $u_i(x) < v_i(x) < u_i(y)$, and $v_i(z) = u_i(z)$ for all $z \neq x$ (including $z = y$). That is, v_i is obtained from u_i by increasing the utility of x and leaving all other alternatives unchanged, the increase for x being so small that it is ranked below y in v_i. Choose $\underline{t}$ such that $\underline{t}$ fits $\underline{v}$ and $P(g(\underline{t})) \subset PO(\underline{v})$; such a $\underline{t}$ exists by assumption. Since y dominates x in $\underline{v}$, this gives $\hat{x}\cdot g(\underline{t}) = 0$. For a given individual i, let $\underline{s}'$ be any strategy profile such that $s_i' = s_i$, and define $\underline{t}' = t_i\underline{s}'_{-i}$. Since s_i is u_i-dominant for i, $u_i\cdot g(\underline{s}') \geq u_i\cdot g(\underline{t}')$; since t_i is v_i-dominant for i, $v_i\cdot g(\underline{s}') \leq v_i\cdot g(\underline{t}')$. This gives $(v_i - u_i)\cdot(g(\underline{s}') - g(\underline{t}')) \leq 0$. Since $v_i - u_i = b\hat{x}$ for some number $b > 0$, this implies $\hat{x}\cdot g(\underline{s}') \leq \hat{x}\cdot g(\underline{t}')$. That is, if t_i is substituted for s_i in any strategy profile, the probability of x cannot decrease. The profile $\underline{t}$ can be obtained from $\underline{s}$ by n such substitutions. Therefore, $\hat{x}\cdot g(\underline{s}) \leq \hat{x}\cdot g(\underline{t})$, which gives $\hat{x}\cdot g(\underline{s}) = 0$. This proves the claim that $g(\underline{s})$ is ex post Pareto optimal for $\underline{u}$.

Let $\underline{\tilde{s}}$ be a dominant selection profile. For any $\underline{u}$, $\underline{\tilde{s}}(\underline{u})$ fits $\underline{u}$; therefore, the argument of the previous paragraph shows that the decision scheme $g \circ \underline{\tilde{s}}$ satisfies the ex post Pareto condition. By earlier arguments, this scheme is also strategy proof. Theorem 1 shows that $g \circ \underline{\tilde{s}}$ is of the form (1). If $\underline{\tilde{t}}$ is another dominant selection profile, the same holds for $g \circ \underline{\tilde{t}}$, but possibly with different weights $a_1,\ldots,a_n$. Let i be a fixed individual, and let x and y be distinct alternatives. Find a utility scale u such that $u(x) > u(z) > u(y)$ for all $z \notin \{x,y\}$, and let the profile $\underline{u}$ satisfy $u_i = u$ and $u_j = -u$ for all $j \neq i$. Let $\underline{s} = \underline{\tilde{s}}(\underline{u})$ and $\underline{t} = \underline{\tilde{t}}(\underline{u})$; then both $\underline{s}$ and $\underline{t}$ fit $\underline{u}$. For any individual k, let $\underline{s}'$ satisfy $s_k' = s_k$, and define $\underline{t}' = t_k\underline{s}'_{-k}$. Both s_k and t_k are u_k-dominant for k, which implies $u_k\cdot g(\underline{s}') = u_k\cdot g(\underline{t}')$. Since u_k is either u or -u, this gives $u\cdot g(\underline{s}') = u\cdot g(\underline{t}')$. The profile $\underline{t}$

can be obtained from $\underline{s}$ in n steps, where each step consists of substituting some t_k for s_k. By the argument above, such a step does not change the utility of the outcome according to u. Hence $u \cdot g(\underline{s}) = u \cdot g(\underline{t})$. Since $T(u_i) = \{x\}$ and $T(u_j) = \{y\}$ for $j \neq i$, previous statements about $g \circ \tilde{\underline{s}}$ and $g \circ \tilde{\underline{t}}$ show that $g(\underline{s})$ and $g(\underline{t})$ are both lotteries in x and y. This implies $g(\underline{s}) = g(\underline{t})$, and the number a_i of (1) must be the same for the schemes $g \circ \tilde{\underline{s}}$ and $g \circ \tilde{\underline{t}}$. Since $\tilde{\underline{t}}$ was an arbitrary dominant selection profile, the conclusion is that the numbers $a_1,\ldots,a_n$ depend only on g and not on $\tilde{\underline{s}}$ or $\tilde{\underline{t}}$.

Finally, let $\underline{u}$ and $\underline{s}$ be such that $\underline{s}$ fits $\underline{u}$. There exists a dominant selection profile $\tilde{\underline{s}}$ such that $\tilde{\underline{s}}(\underline{u}) = \underline{s}$. The scheme $g \circ \tilde{\underline{s}}$ is of the form (1). For every i, define $\rho_i = f_i(\underline{u})$, where f_i is given by (1). The conclusion of Corollary 1 follows. ||

An important point is that the numbers $a_1,\ldots,a_n$ depend only on g. The corollary could also have been formulated as a statement about decision schemes of the form $g \circ \tilde{\underline{s}}$ for dominant $\tilde{\underline{s}}$; any such scheme is a probability mixture of dictatorial schemes, with weights $a_1,\ldots,a_n$. One might ask whether the lottery ρ_i can be viewed as the value at $\underline{s}$ of a function g_i, and whether g can be expressed as a probability mixture of "dictatorial" game forms. This does not follow directly. For one thing, as Corollary 1 is formulated and proved, ρ_i may depend on $\underline{u}$ as well as on $\underline{s}$. Moreover, the corollary says nothing about $g(\underline{s})$ if there exists no $\underline{u}$ such that $\underline{s}$ fits $\underline{u}$. Finally, the concept of dictatorship has not been defined for general game forms. Nevertheless, the following result provides affirmative answers to the questions.

Corollary 1': Assume that the premise of Corollary 1 holds, and let $a_1,\ldots,a_n$ be the numbers given in that corollary. Then there exist game forms $g_1,\ldots,g_n$, defined on the same strategy sets as g, such that:

(i) For all $\underline{s}$, $g(\underline{s}) = a_1g_1(\underline{s}) + \ldots + a_ng_n(\underline{s})$.

(ii) For all i with $a_i > 0$, $P(g_i(\underline{s})) \subseteq T(u)$ whenever s_i is u-dominant for i in g.

The proof is given in the last part of Appendix A. It uses Corollary 1, and it also uses techniques developed earlier in that appendix.

A natural definition of dictatorship is the following:

> The game form g_i is dictatorial for i if and only if, for any utility scale u, there exists a strategy $s_i \in S_i$ such that $P(g_i(s_i\underline{t}_{-i})) \subseteq T(u)$ for all $\underline{t}$.

The strategy s_i will of course depend on u. The definition requires that person i, by using strategy s_i, can secure an optimal outcome (according to u), no matter what other people do.[18]

Since g is straightforward, there will always exist an s_i which is u-dominant for i. Therefore, when $a_i > 0$, part (ii) of the corollary implies that g_i is dictatorial for i. Moreover, part (ii) says that any u-dominant strategy for i can be chosen as s_i in the definition of dictatorship. If there exists no u such that s_i is u-dominant for i, then (ii) places no restriction on $g_i(\underline{s})$. This is as it should be; the strategy s_i is simply redundant for i's ability to dictate the outcome of g_i. (Even when s_i is of this type, restrictions are placed on $g_j(\underline{s})$ for $j \neq i$; this is one of the ways in which Corollary 1' is stronger than Corollary 1.) When $a_i = 0$, the game form g_i is irrelevant. It is possible, of course, to choose a g_i which is dictatorial for i, but there need not exist any g_i satisfying (ii).

Corollary 1*: Let g be a straightforward game form. Assume that g satisfies the attainability condition. Then there exist nonnegative numbers $a_1,\ldots,a_n$ with sum 1 such that the following holds: For any $\underline{u}$ and any $\underline{s}$ that fits $\underline{u}$, there exist lotteries $\rho_1,\ldots,\rho_n$ such that $g(\underline{s}) = a_1\rho_1 + \ldots + a_n\rho_n$, and $P(\rho_i) \subseteq T(u_i)$ for all i.

Proof: Let $\tilde{\underline{s}}$ be a dominant selection profile. Then $g \circ \tilde{\underline{s}}$ satisfies attainability. This is shown in the following way: Let $x \in A$, choose a utility profile $\underline{u}$ such that $T(u_i) = \{x\}$ for all i, define $\underline{s} = \tilde{\underline{s}}(\underline{u})$, and find a $\underline{t}$ with $g(\underline{t}) = \hat{x}$. (Such a $\underline{t}$ exists since g satisfies attainability.) Since s_1 is u_1-dominant for 1, $u_1 \cdot g(s_1\underline{t}_{-1}) \geq u_1 \cdot g(\underline{t}) = u_1(x)$. For all $\rho \neq \hat{x}$, $u_1 \cdot \rho < u_1(x)$; therefore, $g(s_1\underline{t}_{-1}) = \hat{x}$. A similar argument shows that the outcome will still be $\hat{x}$ if s_2 is substituted for t_2 in $s_1\underline{t}_{-1}$, etc. After n such substitutions, the conclusion $g(\underline{s}) = \hat{x}$ follows. The claim has been proved.

Theorem 1* shows that $g \circ \tilde{\underline{s}}$ is of the form (1). The numbers $a_1,\ldots,a_n$ are the same for all dominant selection profiles $\tilde{\underline{s}}$, that is, these numbers depend only on g. The proof of this statement is equal to an argument used in the proof of Corollary 1. The last paragraph of that proof also applies unchanged, completing the proof of Corollary 1*. ||

The conclusion of Corollary 1* can be strengthened, exactly as in Corollary 1'. The proof is also the same.

The following result corresponds to Theorem 2:

Corollary 2: Let g be a straightforward game form. Assume that for each utility profile $\underline{u}$ there exists a strategy profile $\underline{s}$ such that $\underline{s}$ fits $\underline{u}$ and $g(\underline{s})$ is ex ante Pareto optimal for $\underline{u}$. Then there exists an individual i such that the following holds: For any u and $\underline{s}$ such that s_i is u-dominant for i, $P(g(\underline{s})) \subseteq T(u)$.

Proof: When $g(\underline{s})$ is ex ante Pareto optimal for $\underline{u}$, then $g(\underline{s})$ is ex post Pareto optimal for $\underline{u}$. (See the proof of Theorem 2 in Section 3.) Therefore, the premise of Corollary 2 implies the premise of Corollary 1, and there exist numbers $a_1,\ldots,a_n$ with the properties described in the conclusion of Corollary 1.

Suppose $0 < a_i < 1$ for some i, let $\underline{u}$ be the utility profile constructed in the proof of Theorem 2, and choose $\underline{s}$ such that $\underline{s}$ fits $\underline{u}$ and $g(\underline{s})$ is ex ante Pareto optimal for $\underline{u}$. By Corollary 1, $g(\underline{s}) = a_i\hat{x} + (1-a_i)\hat{y}$. Then $g(\underline{s})$ is not ex ante Pareto optimal for $\underline{u}$, contrary to assumption. Hence each a_i is either 0 or 1.

In other words, there exists an i such that $a_i = 1$ and $a_j = 0$ for $j \neq i$. Let u and $\underline{s}$ be given, and assume that s_i is u-dominant for i. If $T(u) = A$, then $P(g(\underline{s})) \subseteq T(u)$ obviously holds. Otherwise, let the utility scale v satisfy $v(x) = 0$ for $x \in T(u)$ and $v(y) = 1$ for $y \notin T(u)$. For all $j \neq i$, let t_j be v-dominant for j. Complete the strategy profile $\underline{t}$ by setting $t_i = s_i$. Define the utility profile $\underline{v}$ by $v_i = u$ and $v_j = v$ for $j \neq i$. Then $\underline{t}$ fits $\underline{v}$. Corollary 1 gives $P(g(\underline{t})) \subseteq T(u)$, which implies $v \cdot g(\underline{t}) = 0$.

If $j \neq i$ and s_j is substituted for t_j in some strategy profile, the utility of the outcome according to v cannot increase. Since $\underline{s}$ can be obtained from $\underline{t}$ by a series of such substitutions, $v \cdot g(\underline{s}) \leq v \cdot g(\underline{t}) = 0$. This is only possible if $P(g(\underline{s})) \subseteq T(u)$, and the proof of Corollary 2 is complete. ||

The conclusion of Corollary 2 corresponds to that of Corollary 1'. Therefore, the statement implies that g is dictatorial for i. Since the proof of this result is relatively easy, there is no reason to formulate the weaker statement corresponding to Corollary 1. Note that the proof depends on Corollary 1 and therefore on Theorem 1. A proof depending only on Theorem 2 would have been preferable, but there seems to be no easy way to construct such a proof without strengthening the premise.[19]

When dominant strategies are required, nothing is gained by considering general game forms instead of decision schemes. This is the principal conclusion of this section. In particular, there is no point in making the strategy set S_i larger than the set of utility scales or letting a strategy be a more complicated object than a utility scale. But the corollaries also hold when S_i is smaller than the set of utility scales; for example, S_i can be finite.

For a given utility scale, an individual may have several dominant strategies. In game forms covered by the corollaries, it makes essentially no difference which of these strategies is chosen. To be precise, if each individual has a unique top-ranked alternative, all profiles of dominant strategies give the same outcome. This follows from the fact that the numbers $a_1,\ldots,a_n$ of Corollary 1 depend only on the game form and not on the chosen strategies. This result can also be applied to strategy-proof decision schemes; there may exist other dominant strategies than reporting one's preferences correctly, but on strict utility profiles nothing changes if such strategies are used. (The opposite would not necessarily have been an advantage; if the outcome depends in an essential way on the choice of dominant strategy, the utility profile does not uniquely determine the decision, and new problems arise.)

Most of the comments made after Theorem 1 apply to the corollaries as well. For example, Corollary 1 does not provide a complete characterization of game forms satisfying the premise. If only strict utility scales and corresponding dominant strategies are considered, a full characterization is given. On this domain, the results also show that there is no loss of generality in using game forms with $S_i = A$. That is, the participants vote directly for alternatives, and each person's ballot is chosen with a predetermined probability. Of course, it is a dominant strategy to vote for one's top-ranked alternative. This is the "random dictatorship" as described earlier. When both strict and non-strict utility profiles are considered, a game form may require more information than people's top-ranked alternatives. Some examples were given in Section 3. Even the most complicated decision scheme mentioned there, namely the "serial dictatorship," depends only on information about individuals' ordinal preferences over non-random alternatives. Hence such a scheme can be implemented as a game form in which a strategy is a preference ordering over A.

A strategy may be a very complicated object, containing a lot of information. The premise of the corollaries places no explicit restrictions on the strategies. But the conclusion is that very little information is actually used, and there is no reason to let the strategies be complicated objects.

5. Some Generalizations

5.1. Infinite Sets of Alternatives or Individuals

In Sections 3 and 4, the sets A and N were assumed to be finite. The consequences of relaxing these conditions are considered next.

In applications of the model, there may be good reasons for using an infinite set of non-random alternatives. For example, this is the case if the decision involves transfer or allocation of a divisible commodity.[20]

The theorems in Section 3 hold even if A is infinite. This should come as no surprise; there is no reason why the introduction of infinitely many alternatives should increase the class of decision schemes satisfying the premise of the theorems. The proof of Theorem 1, as given in Appendix A, depends on the assumption that A is finite, but it is not difficult to remove this dependence. The issue is discussed in detail in Appendix C. The proof of the corollaries in Section 4 can then be applied unchanged for infinite A.

The discussion in Appendix C assumes that decision schemes are defined on all utility profiles, and that the outcome always is a discrete probability distribution on A. If the decision includes the allocation of a divisible good, it may be possible to restrict the class of admissible utility scales. On an infinite A, there will also exist probability distributions which are not discrete, for example, there exist continuous distributions, which give every single element of A probability 0. These possibilities are not considered in this paper and should be the subject of further studies.

If N is infinite, there exist counterexamples to Theorems 1, 1* and 2. Such examples are constructed in Appendix C. The principal reason why the theorems fail is that strategy proofness is a relatively weak condition when N is infinite. It is possible to find decision schemes in which the outcome depends on the total system of individual preferences (which is necessary to guarantee Pareto optimality), but at the same time no single individual can affect the outcome by a unilateral change of action. The fact that the theorems do not generalize to the case of infinite N corresponds to well-known results in the "traditional" social choice theory; in particular, Arrow's impossibility theorem is not true for infinite sets of voters.[21]

One can ask whether there is any point in studying the case of infinite N. The answer is probably no; the case should be viewed mainly as a mathematical curiosity. Obviously, there are never infinitely many decision makers. Infinity can be seen as an approximation to large numbers. The following conclusion can perhaps be drawn: If there are many individuals, there exists a decision scheme f such that f satisfies the ex post Pareto condition, f is not a probability mixture of dictatorial schemes, and f is "almost" strategy proof. This issue is discussed further in Section 7.

5.2. Decision Schemes Defined on Dense Sets

The theorems assume that decision schemes are defined on all utility profiles. It could clearly make a difference if substantive restrictions were placed on people's preferences. The issue to be discussed here, however, is whether more formal restrictions can expand the class of schemes satisfying the premise of the theorems.

Assume, for example, that utility scales are vectors of rational numbers rather than vectors of arbitrary real numbers. The authority which administers the decision making now knows something it did not know in the earlier model; a lot of previously possible utility scales can be ruled out. This increased knowledge could conceivably enlarge the class of acceptable decision schemes. If this were the case, there could hardly be any reason not to impose this condition. There is certainly some limitation on people's ability to discriminate between almost equally good lotteries, and any utility scale in the original model can be approximated arbitrarily closely by utility scales from the restricted set. Therefore, the assumption places no real restriction on individual preferences.

The fact is, however, that nothing can be gained by making this kind of assumption. If a dense subset of the appropriate Euclidean space is given and utility scales are restricted to this set, all theorems and corollaries still hold.[22] It is easy to modify the earlier proofs so that they apply in this case. Examples of dense subsets are the set of utility scales with rational values, the set of strict utility scales, and the set of scales u satisfying $u(x) \neq u(y)$ for all $x \neq y$.

The argument used above concerning individuals' discriminating ability could perhaps justify restricting the utility scales to some finite set.[23] A utility scale u is said to be <u>normalized</u> if either $u(x) = 0$ for all $x \in A$ or $\min_{x \in A} u(x) = 0$ and $\max_{x \in A} u(x) = 1$. For every utility scale, there exists a normalized scale which represents the same preferences.[24] The set of normalized scales is a bounded subset of Euclidean space. (Here A is assumed to be finite.) Suppose that there is given a finite but large set of utility scales, spread uniformly around in the set of normalized scales. All possible preferences can be approximated by a utility scale from this finite set; the accuracy of the approximation depends on the size of the set.

Therefore, it can be argued that no substantive restriction is placed on individual preferences by requiring that utility scales lie in this set.

The argument of the previous paragraph depends on there being a uniform lower bound on people's ability to discriminate between almost equally good lotteries. Only when such a bound exists can a finite set with the required properties be found. Any specific uniform bound will be difficult to justify; therefore, the argument above is not necessarily persuasive. On the other hand, restricting utility scales to a dense subset of Euclidean space is consistent with an assumption that each individual has a limited discriminating ability, but the decision procedure must be prepared to encounter persons for whom this limit is arbitrarily low.

When utility scales are restricted to a finite set as described above, one cannot expect the theorems to hold exactly. There are reasons to believe, however, that they will hold in an approximate sense, and that the approximation can be made arbitrarily good by choosing the set of allowable utility scales large enough. (It is always assumed that this set is spread uniformly around in the set of normalized utility scales.) I believe that this conjecture can be proved by modifying the earlier proofs, but precise formulation of the statement and formal proof will not be given here.[25]

5.3. Removing the Pareto Condition

Gibbard [14] proves a representation theorem for all straightforward game forms with finite strategy sets, without imposing the Pareto (or attainability) condition. The result can be described in this way: Let g be a straightforward game form defined on strategy sets $S_1, \ldots, S_n$ and assume that each S_i is finite. For each i, a set $T_i \subseteq S_i$ is constructed. The set T_i has the property that for every utility scale u, there exists a strategy $t_i \in T_i$ such that t_i is u-dominant for i. On the domain $T_1 \times \ldots \times T_n$, g is a probability mixture of game forms each of which is either unilateral or duple. A game form is unilateral if the outcome only depends on one individual's strategy; it is duple if there exist two alternatives such that the outcome is always a lottery in these two alternatives.

Can this result, or a result of this type, be proved for decision schemes and for game forms with arbitrary strategy sets? I am not able to give a definite answer to this question, but I believe that the answer is yes. As a step towards proving such a theorem, the case of A having exactly three elements will be discussed below.

Assume, therefore, that A has three elements, and let f be a strategy-proof decision scheme. Let i and j be different individuals and fix the utility scales for all individuals except i and j. Then f can be viewed as a function only of i's and j's utility scales, and it can be written $f(\underline{u}) = f(u_i,u_j)$. Suppose that u_i and u_j represent neither exactly equal nor exactly opposite preferences over lotteries, and that neither u_i nor u_j represents total indifference.[26] Moreover, let u_i' and u_j' have the following property: There exist a continuous path from u_i to u_i' and a continuous path from u_j to u_j' such that, whenever u_i'' lies on the first path and u_j'' lies on the second path, neither u_i'' nor u_j'' represents total indifference, and u_i'' and u_j'' do not represent equal or opposite preferences.[27] Then the effect of changing i's utility scale from u_i to u_i' is the same whether j's utility scale is u_j or u_j'. In symbols, $f(u_i',u_j) - f(u_i,u_j) = f(u_i',u_j') - f(u_i,u_j')$. In a limited sense, therefore, f is additively separable in u_i and u_j. This statement will not be proved here, but a special case is considered in Appendix A, Lemma A11. Similar techniques can be used to prove the general statement. The proof uses a result from the author's note [15].

The statement of the last paragraph corresponds to Lemma 5 of Gibbard [14]. Without having worked through all details, I conjecture that the rest of Gibbard's proof can now be simulated, using infinite sums and integrals instead of finite sums. The conclusion is that f is a probability mixture of unilateral and duple schemes, except on a set of measure 0.

Further studies are needed to determine if this argument can be extended to the case of A having more than three elements.

6. Nash Equilibria

The informal concept of strategy proofness can be made precise in different ways. In this section, Nash equilibria will be studied. This is generally a weaker concept than dominant strategies. Therefore, there is a hope that the class of Pareto-optimal decision procedures that can be "implemented" in Nash equilibria is larger than the classes described in Theorems 1 and 2. (The concept of "implementation" will be defined later.)

Let a game form g and a utility profile $\underline{u}$ be given. A strategy profile $\underline{s}$ is a Nash equilibrium for $\underline{u}$ in g if, for all i and all strategies t_i for i, $u_i \cdot g(\underline{s}) \geq u_i \cdot g(t_i \underline{s}_{-i})$. This means that as long as all individuals $j \neq i$ stick to their equilibrium strategies s_j, there is no reason for i to deviate from s_i. If $\underline{s}$ fits $\underline{u}$, then $\underline{s}$ is a Nash equilibrium for $\underline{u}$, but the converse need not hold. In this sense, Nash equilibrium is a weaker concept than dominant strategy. The game form g admits Nash equilibria if, for all $\underline{u}$, there exists an $\underline{s}$ such that $\underline{s}$ is a Nash equilibrium for $\underline{u}$ in g. Here $\underline{s}$ may depend on $\underline{u}$ in an arbitrary way; in particular, the definition does not require that s_i depend only on u_i.

Assume that f is a decision scheme, and assume that, for all $\underline{u}$, $\underline{u}$ is a Nash equilibrium for $\underline{u}$ in f. Then f is strategy proof. This follows directly from the definitions. Let g be a game form, and let $\tilde{\underline{s}}$ be a selection profile as described in Section 4. That is, for each utility profile $\underline{u}$, $\tilde{\underline{s}}(\underline{u})$ is a strategy profile, and the ith coordinate of $\tilde{\underline{s}}(\underline{u})$ depends only on u_i. If, for all $\underline{u}$, $\tilde{\underline{s}}(\underline{u})$ is a Nash equilibrium for $\underline{u}$, then the decision scheme $g \circ \tilde{\underline{s}}$ is strategy proof. The theorems of Section 3 can now be applied to f and $g \circ \tilde{\underline{s}}$, provided that they satisfy the appropriate Pareto or attainability condition.[28] In these cases, therefore, there is nothing to be gained by using Nash equilibria instead of dominant strategies.[29]

The purpose of using Nash equilibria instead of dominant strategies, is to enlarge the class of "reasonable" decision procedures. The argument above shows that this is only possible if each individual's strategy in the equilibrium is allowed to depend on the entire utility profile. Then an additional problem arises: How can the equilibrium be found? The individuals are not supposed to know each other's preferences, therefore, they cannot choose the equilibrium directly. Perhaps it is possible to design an iteration process which will converge to the equilibrium. In such a procedure, new questions will arise concerning possibilities for strategic behavior. Although nobody can gain by a unilateral change of strategy in the equilibrium, the same is not necessarily true when the total iteration procedure is considered. These problems will not be discussed here, but they will have to be addressed if game forms admitting Nash equilibria shall actually be implemented.

It is possible to construct game forms with these properties: There will always exist a Nash equilibrium. The outcome produced by one such equilibrium is, in many respects, more desirable than anything obtainable in strategy-proof decision schemes. To give an example, let each strategy set S_i consist of all pairs of the form (x,ρ), where x is an alternative and ρ is a lottery. For $\underline{s} = ((x_1,\rho_1),\ldots,(x_n,\rho_n))$, $g(\underline{s})$ is defined in this way: If $\rho_1 = \rho_2 = \ldots = \rho_n$, then $g(\underline{s}) = \rho_1$; otherwise, $g(\underline{s}) = (1/n) \sum_{i=1}^{n} \hat{x}_i$.
Intuitively, the strategy (x,ρ) represents a statement like "I want x, but I am willing to accept ρ as a compromise." If the second component of the

strategy is the same for everybody, a compromise has been found. Otherwise, the decision is made by the random dictatorship procedure. The strategy $(x,\hat{x})$ represents complete unwillingness to compromise.

Let a utility profile $\underline{u}$ be given. For each i, find an alternative x_i in $T(u_i)$ and a normalized utility scale v_i representing the same preferences as u_i. (See definition in Section 5.2.) Define $\rho = (1/n)\sum_{i=1}^{n} \hat{x}_i$, and choose the lottery σ to maximize $\sum_{i=1}^{n} v_i \cdot \sigma$, subject to $u_i \cdot \sigma \geq u_i \cdot \rho$ for all i. Consider the strategy profile $\underline{s} = ((x_1,\sigma),\ldots,(x_n,\sigma))$. Then $g(\underline{s}) = \sigma$, and $\underline{s}$ is a Nash equilibrium for $\underline{u}$. To prove the last claim, assume that individual i unilaterally deviates from the strategy profile $\underline{s}$ by using strategy (x_i',σ'). If $\sigma' = \sigma$, this does not change the outcome; if $\sigma' \neq \sigma$, g is defined by the random dictatorship procedure, and the outcome cannot be better than ρ in i's view. (The outcome is as good as ρ if $x_i' \in T(u_i)$ and worse than ρ otherwise.) By the construction of σ, i cannot gain by making this change. The lottery σ is ex ante Pareto optimal for $\underline{u}$ in the strong sense. A decision scheme can be defined by regarding σ as a function of $\underline{u}$.[30] This decision scheme is not dictatorial, and it is not a probability mixture of dictatorial schemes. It satisfies the ex ante Pareto condition. Moreover, the discussion above shows that the game form g, in a certain sense, implements the decision scheme.

There is a problem, however. The game form g has many Nash equilibria in addition to the ones described above. Let $\underline{u}$ be given, and choose x_i as before. Then $((x_1,\hat{x}_1),\ldots,(x_n,\hat{x}_n))$ is a Nash equilibrium. To prove this, consider the possibility of individual i changing strategy. If the alternatives x_j for $j \neq i$ are not all equal, random dictatorship will be used no matter what i does; therefore, i acts optimally by choosing a strategy in which x_i is the first component. If all x_j for $j \neq i$ are equal, i can secure the outcome x_j, but this cannot be better than the result of using the equilibrium strategy $(x_i, \hat{x}_i)$. When this Nash equilibrium is used, g behaves like a decision scheme of the type described in Theorem 1.[31]

What will happen when there are several Nash equilibria? In general, there is no way of answering this question. There is no reason to believe that one equilibrium has a special status so that it will be chosen ahead of the others. The situation is different when dominant strategies and strategy-proof decision schemes are considered: If "reporting one's preferences correctly" is a dominant strategy, it can be argued that this strategy is the most natural choice even if other dominant strategies exist.[32] (In addition, when the corollaries in Section 4 apply, the choice of dominant strategy will usually not matter.)

There is not necessarily anything wrong with multiple equilibria. Suppose that society somehow has established a norm for what ought to be decided. This norm will be a function; for each configuration of individual preferences, it specifies what ought to happen.[33] For a given system of preferences, there may be several outcomes that are equally good from a social point of view, although the individuals are not indifferent to the choice among these outcomes. Moreover, suppose that the game form g admits Nash equilibria, and that $g(\underline{s})$ is a socially optimal outcome for $\underline{u}$ whenever $\underline{s}$ is a Nash equilibrium for $\underline{u}$. Then g can be said to implement the given social norm.[34] If there are many equilibria for a certain $\underline{u}$, it does not matter which one is eventually chosen, since they all produce optimal outcomes.

On the other hand, every actual decision procedure somehow reaches a definite outcome. This outcome may in itself be a stalemate, but then "stalemate" should be regarded as an alternative and included in the set A. Therefore, a game form with multiple equilibria cannot be a complete description of a decision process. When there is more than one equilibrium outcome, a fully adequate model must explain how the final decision is reached. Therefore, game forms with unique equilibrium outcomes are of special interest, and these game forms will be studied from now on.

A requirement that the equilibrium outcome always be unique, is, however, too strong. Choose the utility profile $\underline{u}$ such that $u_i(x) = 0$ for all i and x. In any game form, every strategy profile is a Nash equilibrium for $\underline{u}$. Hence the outcome is unique only if the game form is a constant function. To avoid this problem, one can require that all equilibrium outcomes be equally good for all individuals.[35] This condition holds in the example just given. The condition can be formalized in this way: In the game form g, Nash equilibria are <u>essentially unique</u> if, for any utility profile $\underline{u}$ and all strategy profiles $\underline{s}$ and $\underline{t}$ that are Nash equilibria for $\underline{u}$, $u_i \cdot g(\underline{s}) = u_i \cdot g(\underline{t})$ for all i. This is also a very strong condition, as shown by the following result.

<u>Lemma</u>: Suppose that the game form g admits Nash equilibria, and that Nash equilibria are essentially unique in g. Let i, $\underline{u}$ and v_i be given, define $\underline{v} = v_i \underline{u}_{-i}$, and let $\underline{s}$ and $\underline{t}$ be Nash equilibria for $\underline{u}$ and $\underline{v}$, respectively. Then $u_j \cdot g(\underline{s}) = u_j \cdot g(\underline{t})$ for all $j \neq i$.

<u>Proof</u>: Let u_i' represent total indifference, and define $\underline{u}' = u_i' \underline{u}_{-i}$. Then $\underline{s}$ is a Nash equilibrium for $\underline{u}'$. Individual i can never gain by choosing

a strategy different from s_i, since i regards all lotteries as equally good. For $j \neq i$, the equilibrium condition follows from the assumption that $\underline{s}$ is an equilibrium for $\underline{u}$.[36] The definitions give $\underline{u}' = u_i'\underline{v}_{-i}$; therefore, a similar argument shows that $\underline{t}$ is a Nash equilibrium for $\underline{u}'$. Since equilibria are essentially unique, $u_j' \cdot g(\underline{s}) = u_j' \cdot g(\underline{t})$ for all j. The conclusion follows, since $u_j = u_j'$ for $j \neq i$. ||

Let g satisfy the premise of the lemma. In words, the conclusion says that a change in one individual's utility scale can neither harm nor benefit other persons. This has strong consequences. Choose two alternatives x and y, let the utility scale u satisfy $u(x) > u(z) > u(y)$ for all $z \notin \{x,y\}$, and define $\underline{u}$ and $\underline{v}$ by $u_i = u$ and $v_i = -u$ for all i. A weak version of the Pareto condition requires that there exist strategy profiles $\underline{s}$ and $\underline{t}$ such that $\underline{s}$ is an equilibrium for $\underline{u}$, $\underline{t}$ is an equilibrium for $\underline{v}$, $g(\underline{s}) = \hat{x}$, and $g(\underline{t}) = \hat{y}$. It is possible to go from $\underline{u}$ to $\underline{v}$ in n steps, each time changing one individual's utility scale from u to -u. If $n \geq 2$, the utility of the outcome, measured by u, will remain constant when these changes are made. This contradicts $g(\underline{s}) = \hat{x}$ and $g(\underline{t}) = \hat{y}$.

The conclusion is that no reasonable decision procedure can be represented by a game form that satisfies the premise of the lemma. In fact, the lemma is more restrictive than Theorem 1. This may appear to contradict the statement that Nash equilibrium is a weaker concept than dominant strategy. The explanation is that Theorem 1 contains no uniqueness condition; there is less need for such a condition when strategy-proof decision schemes are studied, since people can be assumed to choose the natural dominant strategy of reporting preferences truthfully.

The utility scale u_i', which is used in the proof of the lemma, represents total indifference. The conclusion may change if the class of utility scales is restricted, in such a way that total indifference is not allowed. From now on, it is assumed that individual preferences are represented by strict utility scales. That is, each individual has a unique top-ranked alternative. This is not a very stringent restriction, since any utility scale has arbitrarily close approximations in the class of strict scales. Note that Theorems 1 and 2 hold on this restricted domain; see Section 5.2.

On this domain, equilibria are essentially unique in the random dictatorship procedures of Theorem 1. Reporting one's top-ranked alternative incorrectly is disadvantageous no matter what other people do (except for individuals with weight 0, but they do not affect the outcome). Hence the equilibrium outcome is unique for any configuration of preferences. This shows that the class of Pareto-optimal decision procedures that can be implemented in Nash equilibria is not smaller than the class described in Theorem 1. It is larger, as will be shown below.

The discussion relies on a theorem proved by Maskin [18]. The result needed here is a special case of Maskin's theorem; it is presented in the notation of this paper and under the assumption that all utility profiles are strict.

A decision scheme f satisfies <u>no veto power</u> if the following implication holds for all i, ρ and $\underline{u}$: If $u_j \cdot \rho \geq u_j \cdot \sigma$ for all $j \neq i$ and all σ, then $f(\underline{u}) = \rho$. Since each u_j is strict, the premise can only hold when ρ is a degenerate lottery. Therefore, the condition is equivalent to: For all i, x and $\underline{u}$, if $T(u_j) = \{x\}$ for all $j \neq i$, then $f(\underline{u}) = \hat{x}$. That is, the condition requires that if at least n-1 individuals have the same top-ranked alternative, the lone dissenter (if there is one) shall not be able to prevent this alternative from being chosen with certainty. (Random dictatorship does not satisfy this condition.) The scheme f satisfies monotonicity if the following implication holds for all $\underline{u}$, $\underline{u}'$ and ρ: If $f(\underline{u}) = \rho$, and $u_i \cdot \rho \geq u_i \cdot \sigma$ implies $u_i' \cdot \rho \geq u_i' \cdot \sigma$ for all i and σ, then $f(\underline{u}') = \rho$. The condition can be explained in this way: Suppose that ρ is the outcome for certain preferences. Then preferences change, but in nobody's opinion does ρ fall below a lottery that originally was no better than ρ. After the change, ρ shall still be the outcome.

The premise of the monotonicity condition has this consequence: Let $\underline{u}$, $\underline{u}'$ and ρ satisfy the premise, and assume $\rho(x) > 0$. For any i, x cannot fall in i's ordinal ranking of A when u_i' is substituted for u_i. If this is wrong, there must exist i and y such that either $u_i(x) \geq u_i(y)$ and $u_i'(x) < u_i'(y)$, or $u_i(x) > u_i(y)$ and $u_i'(x) = u_i'(y)$. In the former case, the lottery $\sigma = \rho - \rho(x) \cdot \hat{x} + \rho(x) \cdot \hat{y}$ will satisfy $u_i \cdot \rho \geq u_i \cdot \sigma$ and $u_i' \cdot \rho < u_i' \cdot \sigma$, contradicting the premise. In the latter case, there exists a z with $u_i'(z) > u_i'(x)$, because u_i' is strict. Let $\sigma = \rho - \rho(x) \cdot \hat{x} + (1-\varepsilon)\rho(x) \cdot \hat{y} + \varepsilon\rho(x) \cdot \hat{z}$, for some ε with $0 < \varepsilon \leq 1$. Then $u_i' \cdot \rho < u_i' \cdot \sigma$ and, for sufficiently small ε, $u_i \cdot \rho > u_i \cdot \sigma$. Again, the premise is contradicted. This proves the claim.

If there are at least three individuals and f satisfies no veto power and monotonicity, then there exists a game form g such that g admits Nash equilibria, and $g(\underline{s}) = f(\underline{u})$ whenever $\underline{s}$ is an equilibrium for $\underline{u}$ in g. This is Theorem 5 in Maskin's paper.[37] The conclusion implies that g implements f in a strong sense. For any configuration of preferences, there is a unique equilibrium outcome in g, namely the outcome prescribed by f. This does not imply that the decision scheme f is strategy proof in the sense of Section 2. The strategy sets for g may be larger and more complicated than the set of utility scales. In Maskin's proof, each strategy set is equal to the set of utility profiles. That is, in g each person "votes" by specifying a complete configuration of preferences for all the individuals.

In the examples below, there must be at least three individuals. For simplicity, n is also assumed to be odd. (If n is even, some tie-breaking rule is needed.)

<u>Example 1</u>: The decision scheme f^1 is defined as follows: Let $\underline{u}$ be a strict utility profile. If there exists an x such that $\#\{i|T(u_i) = \{x\}\} > n/2$, then $f^1(\underline{u}) = \hat{x}$. Otherwise, $f^1(\underline{u})$ is defined by the random dictatorship procedure, with all individuals having equal weight.

In informal terms, f^1 works like this: People vote for their top-ranked alternatives. If one alternative gets an absolute majority, it is chosen with certainty; otherwise, a randomly chosen ballot decides.

No veto power is obviously satisfied by f^1. Concerning monotonicity, let $\underline{u}$, $\underline{u}'$ and ρ satisfy the premise of that condition. Assume that x is top-ranked in $\underline{u}$ for a majority of the individuals, which gives $f^1(\underline{u}) = \rho = \hat{x}$. By an earlier argument, x cannot fall in anybody's ordinal ranking when $\underline{u}$ is changed to $\underline{u}'$. Hence x is also top-ranked in $\underline{u}'$ for a majority, and $f^1(\underline{u}') = \hat{x}$. Then assume that $f^1(\underline{u}) = \rho$ is defined by random dictatorship. If x is top-ranked in u_i for some i, then $\rho(x) > 0$. The earlier argument shows that x is top-ranked in u_i'. Hence $T(u_i) = T(u_i')$ for all i, and $f^1(\underline{u}') = f^1(\underline{u})$. Monotonicity has been proved, and Maskin's theorem shows that f^1 can be implemented in Nash equilibria, as described above.

The decision scheme f^1 satisfies the ex post Pareto condition. It is not a probability mixture of dictatorial decision schemes. This proves that the class of ex post Pareto-optimal decision procedures that can be implemented in Nash equilibria is larger than the class described in Theorem 1.

The ex ante Pareto condition is not satisfied in Example 1. In the following example, it is assumed that A has exactly three elements. The other assumptions made above still apply.

Example 2: The decision scheme f^2 is defined as follows: Let $\underline{u}$ be a strict utility profile. If there exists an x such that $\#\{i | u_i(x) > u_i(y)\} > n/2$ for all $y \neq x$, then $f^2(\underline{u}) = \hat{x}$. Otherwise, $f^2(\underline{u})$ is defined by the random dictatorship procedure, with all individuals having equal weight.

To compute f^2, one must know people's ordinal rankings of the alternatives. If one alternative beats each of the others in pairwise majority votes, this alternative is chosen. (Absolute majorities are needed.) Otherwise, a randomly chosen individual decides.

It is easy to see that no veto power holds. To prove monotonicity, assume that $\underline{u}$, $\underline{u}'$ and ρ satisfy the premise of the condition. First, suppose that x is a majority winner in $\underline{u}$, so that $\rho = \hat{x}$. When $\underline{u}'$ is substituted for $\underline{u}$, x cannot fall in anybody's ordinal ranking of A. Hence x is a majority winner in $\underline{u}'$, and $f^2(\underline{u}') = \rho$. Then suppose that $f^2(\underline{u}) = \rho$ is defined by the random dictatorship procedure. All three elements of A must occur as the top-ranked alternative for some u_i. Otherwise, one alternative would top u_i for more than half the individuals, contradicting the assumption that there is no majority winner in $\underline{u}$. Hence $\rho(x) > 0$ for all $x \in A$. For all i and all x and y, an earlier argument shows that x cannot fall compared to y when u_i is changed to u_i'. Hence u_i and u_i' induce the same ordinal preferences on A. When this holds for all i, the definition of f^2 gives $f^2(\underline{u}') = f^2(\underline{u})$. Monotonicity has been proved, and f^2 can be implemented in Nash equilibria.

The scheme f^2 satisfies the strong ex ante Pareto condition. To prove this, first assume that x is a majority winner in $\underline{u}$. Then there exists an i with $T(u_i) = \{x\}$; otherwise, some $y \neq x$ would be the top-ranked alternative for a majority of the individuals, and x could not be a majority winner. For all $\sigma \neq \hat{x}$, $u_i \cdot \sigma < u_i(x)$; hence $f^2(\underline{u}) = \hat{x}$ is ex ante Pareto optimal for $\underline{u}$. Then suppose that $f^2(\underline{u}) = \rho$ is defined by the random dictatorship procedure, so that there is no majority winner in $\underline{u}$. Moreover, assume that the Pareto condition is violated; that is, there exists a lottery σ such that $u_i \cdot \rho \leq u_i \cdot \sigma$ for all i, with strict inequality for some i. If $\sigma(x) < \rho(x)$, $\sigma(y) > \rho(y)$ and $\sigma(z) \geq \rho(z)$, then $T(u_i) = \{x\}$ implies $u_i \cdot \rho > u_i \cdot \sigma$. There must exist an i with $T(u_i) = \{x\}$; see the proof of monotonicity. Hence the assumptions are contradicted. Apart from permutations of the alternatives, the only remaining possibility is $\sigma(x) > \rho(x)$, $\sigma(y) < \rho(y)$, and $\sigma(z) < \rho(z)$. If $u_i(y) > u_i(z) \geq u_i(x)$, then $u_i \cdot \rho > u_i \cdot \sigma$, contradicting the assumptions. Similarly, $u_i(z) > u_i(y) \geq u_i(x)$ is impossible. Since u_i is strict, the following three possibilities are exhaustive: (a) $T(u_i) = \{x\}$; (b) $u_i(y) > u_i(x) > u_i(z)$; and (c) $u_i(z) > u_i(x) > u_i(y)$. If (b) is true for more than half the individuals, y is a majority winner, and similarly for (c) and z. But if neither (b) nor (c) holds for a majority, x is a majority winner. In each case, the assumptions are contradicted, and the strong ex ante Pareto condition has been proved.

The decision scheme f^2 is not dictatorial; hence it falls outside the class described in Theorem 2. I do not know whether similar examples can be constructed when A has more than three elements. If the Pareto condition is weakened slightly, however, such examples will exist. In the next example, A is an arbitrary finite set, and ρ_0 is the even-chance lottery on A.

Example 3: Let η be a number with $0 < \eta < 1$. The decision scheme f^3 is defined as follows: Let $\underline{u}$ be a strict utility profile. If there exists an x such that $\#\{i|T(u_i) = \{x\}\} > n/2$, then $f^3(\underline{u}) = \hat{x}$. Otherwise, choose $\underline{v}$ such that, for every i, v_i is normalized and represents the same preferences as u_i. (See definition in Section 5.2.) Choose x such that $\sum_{i=1}^{n} v_i(x)$ is maximized. (If two or more alternatives maximize this expression, choose the one that comes first in some fixed ordering of A.) Let $f^3(\underline{u}) = \eta\hat{x} + (1-\eta)\rho_0$.

As before, no veto power is easily proved. Let $\underline{u}$, $\underline{u}'$ and ρ satisfy the premise of the monotonicity condition. If x is top-ranked in $\underline{u}$ for a majority of the individuals, then the same is true in $\underline{u}'$, and $f^3(\underline{u}') = \rho$. (See the discussion of Example 1.) Otherwise, $\rho(x) > 0$ for all $x \in A$. For every i, earlier arguments show that u_i and u_i' represent the same ordinal preferences on A. Let x be the best and let y be a worst alternative in this ordering. Since u_i is strict, $u_i(x) > u_i(y)$, and there exist numbers $c > 0$ and d such that $u_i'(w) = cu_i(w) + d$ for $w = x$ and $w = y$. Suppose that $u_i'(z) > cu_i(z) + d$ for some $z \in A$. Then there exists a number b with $0 < b < 1$ such that the lottery $\rho' = b\hat{x} + (1-b)\hat{y}$ satisfies $u_i \cdot \rho' > u_i(z)$

and $u_i'\cdot\rho' < u_i'(z)$. Let $\sigma = \rho - \varepsilon\rho' + \varepsilon\hat{z}$, with ε positive but so small that σ is a lottery. That is, $\varepsilon b \leq \rho(x)$ and $\varepsilon(1-b) < \rho(y)$. Then $u_i\cdot\rho > u_i\cdot\sigma$ and $u_i'\cdot\rho < u_i'\cdot\sigma$, contradicting the premise of the monotonicity condition. A similar argument shows that there does not exist any z with $u_i'(z) < cu_i(z) + d$. Therefore, u_i' is a positive linear transformation of u_i. But then the normalization $\underline{v}$ of $\underline{u}$ is equal to the similar normalization of $\underline{u}'$, and $f^3(\underline{u}) = f^3(\underline{u}')$. This proves monotonicity.

When $f^3(\underline{u})$ is defined by the first clause in Example 3, the outcome is strongly ex ante Pareto optimal. When the second clause applies, the lottery $\hat{x}$ is strongly ex ante optimal. By choosing η close to 1, it is possible to guarantee that the outcome is arbitrarily close to the Pareto frontier. In this sense, the ex ante Pareto condition will almost hold.[38]

The general conclusion to be drawn from these examples is that there is something to be gained from using Nash equilibria instead of dominant strategies.[39] Whether the decision procedures described in the examples are desirable or reasonable, may depend on the circumstances and will perhaps be a matter of disagreement. But at least in some connections, they are preferable to the procedures of Theorems 1 and 2.

Which Pareto-optimal decision procedures are implementable in Nash equilibria? The examples give only a very limited answer to this question, and further studies are needed to find more complete characterizations.

7. Concluding Remarks

The problem studied in this paper can be summarized as follows: Collective decisions ought to depend on individual preferences, which are generally not known to the authority administering the decision process. (The decision may also depend on other factors, but that is not the issue here.) The preferences must be elicited, by having people vote or participate in some other way in a decision procedure. The participants must be expected to take account of the effect their behavior has on the outcome; they will choose the actions that give the best possible result, according to their own preferences. The problem is to construct the procedure so that the outcome will depend on the preferences in a desirable way.

How should the outcome depend on the preferences? Two factors are obviously relevant, namely efficiency and distributional objectives. The decision should be efficient or Pareto optimal, so that there exists no possible outcome which is preferred by everybody to the actual decision. The choice of one outcome from the Pareto-optimal set is a distributional issue. Ideally, there should exist decision procedures corresponding to all conceivable systems of distributional objectives. But certain objectives, such as equal treatment of the individuals, are of special interest.

In the principal model discussed in Sections 3 - 5, each person is supposed to have a dominant strategy. This strategy depends only on the person's own preferences, and it is optimal no matter what other people do. The results show that the ideal outlined above is far from achievable. If the efficiency condition is imposed on the lottery produced by the decision procedure, influence has to be distributed in an extremely uneven way; one individual unilaterally chooses the outcome.[40] If the efficiency requirement is applied to the final non-random outcome, it is possible to distribute influence more evenly. But this must be done in a very special (and inefficient) way, namely by means of a "random dictatorship" procedure.

If these results are considered unsatisfactory, what can be done? One possibility is to weaken the condition that dominant strategies exist. The important point is that the outcome depends on the preferences in the correct way, when people take account of how their behavior influences the outcome and act accordingly. Perhaps other and weaker conditions than the existence of dominant strategies are consistent with this objective. One alternative condition, namely the existence of Nash equilibria, is discussed in some detail in Section 6. Another possibility is discussed below.

People do not know each other's preferences. But perhaps they have some probabilistic expectations about other people's preferences, on which their choice of strategy can be based. To formalize this idea, let g be a game form with strategy sets $S_1,\ldots,S_n$, and let $\tilde{s}_1,\ldots,\tilde{s}_n$ be strategy selections, as defined in Section 4. The vector $(\tilde{s}_1,\ldots,\tilde{s}_n)$ is a <u>Bayesian equilibrium</u> if the following holds for all individuals i and all utility scales u: Consider the choice of action for person i, whose preferences are represented by u, assuming that each person $j \neq i$ uses the strategy selection $\tilde{s}_j$. From i's probabilistic beliefs about other people's preferences can be deduced a probability distribution for other people's actions. (In general, i's beliefs may depend on i's own preferences.) For each strategy in S_i, i can compute the expected utility of the outcome, according to u, of using this strategy. The assumptions that have been made about individual preferences imply that i wants to maximize this expected value.[41] The condition for a Bayesian equilibrium is that such a maximum is achieved by choosing the strategy $\tilde{s}_i(u)$.[42]

If the selection profile $\underline{\tilde{s}}$ is a Bayesian equilibrium in the game form g, then truthful reporting of preferences is a Bayesian equilibrium in the decision scheme $g \circ \underline{\tilde{s}}$. This follows directly from the definitions.[43] A similar result holds for dominant strategies; see Section 4. In $g \circ \underline{\tilde{s}}$, there may exist Bayesian equilibria that do not correspond to any equilibrium in g. There are reasons to believe that truthful reporting will be chosen whenever it is an equilibrium strategy; if this is true, other equilibria can be ignored.[44] Under this assumption, there is no loss of generality in studying decision schemes in which truthful reporting is a Bayesian equilibrium, rather than dealing with arbitrary game forms.

In a strategy-proof decision scheme, truthful reporting is a Bayesian equilibrium, regardless of people's beliefs about each other's preferences. If nothing is known about these beliefs when the decision scheme is constructed, the converse is also true: Only in strategy-proof decision schemes will truthful reporting always be a Bayesian equilibrium. To prove this, assume that f is not strategy proof. Then there exist i, $\underline{u}$ and v_i such that $u_i \cdot f(v_i \underline{u}_{-i}) > u_i \cdot f(\underline{u})$. If person i has preferences represented by u_i and is convinced that u_j for $j \neq i$ represents other people's preferences, then reporting utility scale u_i cannot be an optimal action for i. Therefore, truthful reporting is not a Bayesian equilibrium under this system of beliefs.[45] The argument does not depend on beliefs being non-probabilistic; it also holds if i is almost certain that u_j represents j's preferences.

The situation is different if people's beliefs are known and can be taken into account when the decision procedure is constructed. This will be illustrated by an example. In the decision scheme used in the example, the outcome depends only on the individuals' ordinal preferences over A. Hence it is only necessary to specify beliefs concerning such preferences.

It is assumed that nobody is indifferent between two elements of A; this is not essential, but it simplifies the argument. For any two individuals i and j, i considers it equally likely that j has any one of the strict preferences over A. The decision scheme is the Borda rule. For a utility profile $\underline{u}$ representing strict preferences over A, define, for any x, $\Gamma(x) = \#\{(i,y)|u_i(x) > u_i(y)\}$. That is, for each individual, the number of alternatives ranked below x is counted, and $\Gamma(x)$ is the sum of these numbers for all individuals. Let B be the set of alternatives with maximal Γ-value, that is, $x \in B$ if and only if $\Gamma(x) \geq \Gamma(y)$ for all y. The outcome $f(\underline{u})$ is the even-chance lottery over B. For any i, x and y, suppose that the utility scales u and v satisfy $u(x) > u(y)$, $u(x) = v(y)$, $u(y) = v(x)$, and $u(z) = v(z)$ for all $z \notin \{x,y\}$. Let ρ and σ be the expected outcome, as i sees it, if i reports utility scales u and v, respectively. It is easy to prove that $\rho(x) > \sigma(x)$, $\rho(y) < \sigma(y)$, and $\rho(z) = \sigma(z)$ for all $z \notin \{x,y\}$. Let u_i represent i's true preferences, and let v_i be any utility scale. The scale v_i can be obtained from u_i in this way: Take the top-ranked alternative in v_i and interchange it with the alternative immediately above it in u_i. Repeat this step, until this alternative has reached the top. Then take the second-ranked alternative in v_i and move it upwards until it reaches second place, etc. After a finite number of steps, the ordinal ranking of v_i is obtained. Each of these steps is of the same type as the change from u to v, and each step reduces the utility of the outcome according to u_i. Since only ordinal preferences matter for the Borda rule, this proves that truthful reporting is a Bayesian equilibrium.[46] The Borda rule satisfies the ex post Pareto condition, and it is not a random dictatorship.

The example shows that at least for some systems of beliefs, Bayesian equilibrium is a weaker concept than dominant strategy. The class of decision schemes in which truthful reporting is a Bayesian equilibrium, will depend on the beliefs. An interesting problem, which should be the topic of further studies, is to characterize this class under various sets of assumptions. Some related results can be found in Myerson [19], see also Dasgupta et al. [7, Section 5] with further references.

Other alternatives to requiring the existence of dominant strategies are studied by Pattanaik [20] and Dasgupta et al. [7, Section 6]. In the model of this paper, however, these possibilities seem less natural, and they will not be discussed here.

Returning to the assumption that dominant strategies exist, one can ask whether the problem posed by the theorems can be solved by dropping the Pareto condition. Such a "solution" will not be a very satisfactory one, but the question is nevertheless of interest. There are reasons to believe that even without the Pareto condition, the class of straightforward game forms is severely restricted. When strategy sets are finite, a result to this effect is proved by Gibbard [14]; the general case is discussed in Section 5.3 above. Even within this class, however, a number of "nice" decision schemes can be constructed. This is discussed, in an ordinal model, by Barberá [3,4]. All the schemes constructed by Barberá are also strategy proof in the model of this paper. The representation of preferences by utility scales rather than ordinal rankings over A will open for further possibilities, some of which are considered by Zeckhauser [25, p. 943].

The premise of the theorems requires that strategy proofness and Pareto optimality hold exactly. If there is even the slightest deviation, the proof is no longer correct. For practical applications, it is perhaps more reasonable to say that these conditions shall hold in an approximate sense.

To make this idea precise, let f be a decision scheme and let ε be a (small) positive number. For a utility scale u, define $\phi(u) = \max_{x \varepsilon A} u(x) - \min_{x \varepsilon A} u(x)$. The scheme f is <u>$\varepsilon$-strategy proof</u> if $u_i \cdot f(v_i \underline{u}_{-i}) \leq u_i \cdot f(\underline{u}) + \varepsilon \cdot \phi(u_i)$, for all $\underline{u}$, i and v_i. That is, incorrect reporting of preferences can at most lead to a small gain. The factor $\phi(u_i)$ makes the definition independent of positive linear transformations of u_i. The lottery ρ is <u>ε-Pareto optimal</u> for $\underline{u}$ if there exists a σ such that σ is Pareto optimal for $\underline{u}$ and $\sum_{x \varepsilon A} |\rho(x) - \sigma(x)| \leq \varepsilon$. This applies both to the ex post and the ex ante condition. It is now obvious what it means to say that f satisfies the ex post or ex ante ε-Pareto condition.

Consider Theorem 1 or Theorem 2, with the premise weakened to the ε-conditions for some $\varepsilon > 0$. Then the theorems are not true; counterexamples are easily constructed. If ε is small, however, I believe that the conclusion will be approximately true. More precisely, I make the following conjecture, concerning Theorem 1: There exists a number M, which depends on the number of individuals and perhaps on the number of alternatives in A. For any $\varepsilon > 0$ and any decision scheme f that satisfies ε-strategy proofness and the ex post ε-Pareto condition, there exists a decision scheme f_0 such that f_0 is a probability mixture of dictatorial schemes and $\sum_{x \varepsilon A} |\hat{x} \cdot f(\underline{u}) - \hat{x} \cdot f_0(\underline{u})| \leq M\varepsilon$ for all $\underline{u}$. It should be possible to prove this statement by modifying the earlier proof, though this may lead to a very large value of M. Concerning Theorem 2, a similar conjecture can be formulated, and I believe that it can be proved in the same way.

The significance of the statement depends on the value of M; the smaller M can be made, the more severe are the restrictions imposed on the decision schemes under consideration. This suggests topics for further studies: First, the conjecture must be proved formally. Second, the lowest possible value of M, as a function of n, should be found.[47] When n increases, M will certainly increase; therefore, the statement is weaker when there are many individuals. Regardless of the size of M, however, the conjecture is not void. If ε is so small that $M\varepsilon$ is small, the decision scheme must be almost equal to a probability mixture of dictatorial schemes.[48]

Finally, some comments are in order on the relationship between the theorems of this paper and results proved by others. Corollaries 1 and 2 in Gibbard [14] are of special interest; they correspond closely to the corollaries in this paper, except that Gibbard assumes that the strategy sets are finite. One can ask whether this is a significant restriction; all practical applications will probably have to be based on finite sets. Even if this argument is accepted, the theorems proved here provide additional information. Gibbard's results do not rule out the following possibility: There exists a strategy-proof decision scheme which satisfies the ex post Pareto condition and is "reasonable," in particular, it is not a probability mixture of dictatorial schemes. This decision scheme can be approximated arbitrarily closely by game forms that are defined on finite strategy sets and are arbitrarily close to being straightforward. As the approximations become better, the strategy sets must be made larger.[49] If this were the case, the significance of the theorems would not be too great; for all practical purposes, the given scheme could be implemented with dominant strategies and finite strategy sets. But Theorem 1 of this paper rules out this possibility; the given "reasonable" decision scheme cannot be strategy proof, and then it cannot be approximated arbitrarily closely by game forms that are arbitrarily close to being straightforward. (Note that this follows from the theorem; the conjecture made earlier in this section need not be invoked.) Hence the results in this paper give more information than the previous theorems, even if game forms in practical applications must have finite strategy sets.[50]

Theorem 2 can also be viewed as a generalization of Theorem V in Zeckhauser [25]. The premise is the same in the two theorems. Zeckhauser's conclusion is that a decision scheme satisfying this premise cannot guarantee a nondictatorial outcome, in the following sense: There exists a utility profile in which the individuals are divided into two groups with strictly conflicting preferences, but the resulting decision represents a total victory for one of the groups.[51] This leaves open the possibility that other utility profiles give less dictatorial outcomes, and that the winning group is different in different situations. Hence it is a relatively weak form of dictatorship.[52] Theorem 2 of this paper strengthens the conclusion by proving that there exists one individual who always wins.

NOTES

[1]This paper was written while the author was a doctoral candidate at the John F. Kennedy School of Government, Harvard University. His work was supported by research fellowships from the Norwegian Research Council for Science and the Humanities and from the University of Oslo.

[2]Vickrey made a more specific conjecture, namely that strategy proofness is equivalent to the condition "independence of irrelevant alternatives" from Arrow's impossibility theorem. He was essentially right; see Satterthwaite [22, Theorems 2 and 2'] and Kalai and Muller [16, Theorem 3].

[3]Several other proofs of the theorem have been published; see, for example, Gärdenfors [11] and Schmeidler and Sonnenschein [23].

[4]Instead of introducing chance, one can drop the assumption that the outcome be unique. That is, when the votes are given, a set of possible outcomes is computed; the choice of one outcome from this set is not specified. The consequences of making this assumption are discussed by Gärdenfors [10] and Barberá [2].

[5]If all preferences over risky prospects are possible, then these prospects can be regarded as the basic outcomes, the Gibbard-Satterthwaite theorem can be applied, and only dictatorial procedures are strategy proof. In order to reach this conclusion, one must, for example, allow preference relations in which x is preferred to each of y and z, but a lottery in y and z is preferred to x. This seems unreasonable. To make the problem an interesting one, some restrictions must be imposed on the preferences. This is done by the assumption in the text. The theory underlying the assumption will not be discussed in detail. For such a discussion, see,

for example, Raiffa [21] or Arrow [1, Chapter 2]. Note that the existence of utility functions is not an assumption of the theory. It is assumed that an individual's preferences over lotteries satisfy certain rationality conditions; then the existence of a utility function representing the preferences can be proved.

[6]This is not to deny that the study of all strategy-proof procedures has theoretical interest; see comments in Section 5.3 below.

[7]These "random dictatorship" procedures have been studied by Gibbard [12, p. 593] and Zeckhauser [25, p. 938].

[8]The relationship between Gibbard's theorems and the results of this paper is discussed further in Section 7.

[9]Randomized strategies are not explicitly considered; they are not, however, ruled out. If g is a game form with strategy sets S_i, another game form g* with strategy sets S_i^* can be constructed in this way: The elements of S_i^* are the randomized strategies for i in g, that is, S_i^* is the set of lotteries over S_i. For $\underline{s}^* \in S_1^* \times \ldots \times S_n^*$, $g^*(\underline{s}^*)$ is the expected value of $g(\underline{s})$, when $\underline{s}$ varies over $S_1 \times \ldots \times S_n$ according to the probability distribution given by $\underline{s}^*$. The results of this paper hold for the game form g*; hence they hold for g when randomized strategies are permitted.

[10]On the other hand, straightforwardness is not the strongest possible condition, since it takes no account of the possibility that coalitions may form; see remarks at the end of Section 3.

[11]See Dasgupta et al. [7, Theorem 4.1.1].

[12]If the utility scales u and v are positive linear transformations of each other (that is, if there exist numbers $c > 0$ and d such that $v(x) = cu(x) + d$ for all $x \varepsilon A$), then u and v represent the same preferences over lotteries. From the individual's point of view, only preferences exist; the utility scale is merely the theorist's abstraction. It may seem reasonable, therefore, to require that a decision scheme f depend only on the individuals' preferences. Formally, this amounts to requiring $f(\underline{u}) = f(\underline{v})$ whenever v_i is a positive linear transformation of u_i for each i. This condition is not needed for the main results, and it is not imposed. A slightly weakened version will, however, follow from Theorem 1; see note 14 and accompanying text.

[13]The phrase "strategy proof" has two different meanings; it has been used both informally and as a formal property of decision schemes. It will always be clear which interpretation is intended.

[14]See note 12 above. The statement need not be true when $\underline{u}$ is not strict.

[15]Formally, the following statement holds: Let τ be a permutation of A, let $\underline{u}$ and $\underline{v}$ be strict, and assume $v_i(x) = u_i(\tau(x))$ for all i and x. Then $\hat{x} \cdot f(\underline{v}) = \hat{y} \cdot f(\underline{u})$ for all x and y with $y = \tau(x)$.

[16]This proof is borrowed from Gibbard [13, p. 678]; see also Zeckhauser [25, p. 939].

[17]If the strategy selection is supposed to describe individual behavior, it must choose the same strategy for any two utility scales that represent the same preferences; see note 12 above. Nothing will change if such a condition is added to the definition.

[18]An alternative condition, used by Gibbard [14, p. 612], is the following: For any $x \in A$, there exists an s_i such that $g_i(s_i\underline{t}_{-i}) = \hat{x}$ for all $\underline{t}$. The two conditions are equivalent. If the one in the text holds and x is given, choose u such that $T(u) = \{x\}$; then $P(g_i(s_i\underline{t}_{-i})) \subseteq T(u)$ is equivalent to $g_i(s_i\underline{t}_{-i}) = \hat{x}$. If the condition of this note holds and u is given, choose any $x \in T(u)$; then $g_i(s_i\underline{t}_{-i}) = \hat{x}$ implies $P(g_i(s_i\underline{t}_{-i})) \subseteq T(u)$.

[19]The advantage of such a proof is that it would significantly simplify the total proof of Corollary 2, which could then rely on the direct proof of Theorem 2 given in Appendix B. A proof based only on Theorem 2 can be constructed if there exists a dominant selection profile $\tilde{\underline{s}}$ such that $g \circ \tilde{\underline{s}}$ satisfies the ex ante Pareto condition. This represents a strengthening of the premise; see comments to Corollary 1.

[20]One can argue that nothing is perfectly divisible; there is always a smallest possible unit and therefore a finite number of outcomes. Even if this is true, an assumption of perfect divisibility is often introduced to simplify the model.

[21]See Fishburn [9] and Kirman and Sondermann [17].

[22]This is also true if utility profiles are restricted to a dense subset of the space to which these objects belong. That is, it is possible to impose some restrictions on the relationship between different individuals' utility scales.

[23]Note that this condition is different from the finiteness condition in Gibbard's theorem [14]. The assumption here is that an individual's true

utility scale belongs to some finite set; in Gibbard's theorem, all utility scales are possible, but there can be only finitely many strategies with which to express one's preferences.

[24]See note 12.

[25]Related issues are discussed in Section 7. In that section, it is also made clear what it means that the theorems hold in an approximate sense.

[26]Formally, the following is required: Neither u_i nor u_j is a constant vector, and there do not exist numbers c and d such that $u_j(x) = cu_i(x) + d$ for all $x \in A$.

[27]This condition will hold if u_i' is close to u_i and u_j' is close to u_j. That is, for given u_i and u_j, there exists $\varepsilon > 0$ such that the condition holds when $\|u_i - u_i'\| < \varepsilon$ and $\|u_j - u_j'\| < \varepsilon$. But the condition also holds for some u_i' and u_j' that are not close to u_i and u_j.

[28]In the game form g, $\underline{\tilde{s}}$ need not be a dominant strategy selection, and the corollaries of Section 4 do not apply to g. There may also exist other Nash equilibria than those produced by $\underline{\tilde{s}}$. Even if this is the case, $g(\underline{\tilde{s}}(\underline{u}))$ is one possible outcome when preferences are described by $\underline{u}$. Therefore, it is possible that g behaves like $g \circ \underline{\tilde{s}}$, in which case g is of the form given by the theorems.

[29]This is a special case of a more general result; see Dasgupta et al. [7, Theorem 7.1.1].

[30]In the description above, σ need not be uniquely determined, but it is easy to make the choice unique by introducing a tie-breaking rule.

31 In this equilibrium, individual i's strategy depends only on u_i. Hence the remarks in note 28 apply to the game form g discussed here.

32 This argument is only correct if the decision scheme is invariant to positive linear transformations of the utility scales; see note 12.

33 There are, of course, enormous problems involved in specifying such a norm; these problems will not be discussed here.

34 This implementation concept is studied extensively by Maskin [18] and Dasgupta et al. [7].

35 This is a much stronger condition than the possibility discussed above that the different equilibrium outcomes are equally good according to some social norm.

36 This argument is a special case of Theorem 2 in Maskin [18].

37 Conversely, if such a g exists, then f satisfies monotonicity. This is shown in Theorem 2 of the same paper. No veto power is not necessary for the existence of a g with the stated properties, as is shown by the discussion above of random dictatorship.

38 In Example 3, the ex post Pareto condition also holds only in an approximate sense. It is possible to modify the example so that this condition is exactly satisfied: If no alternative is top-ranked in u_i for a majority, let $B = PO(\underline{u})$, and for each i let v_i be a positive linear transformation of u_i such that v_i restricted to B has maximum 1 and minimum 0. Find x as before, and let $f^3(\underline{u}) = \eta\hat{x} + (1-\eta)\rho_B$, where ρ_B is the even-chance lottery on B. No veto power and ex post Pareto optimality are easily proved.

If $\underline{u}$ and $\underline{u}'$ satisfy the premise of monotonicity and $f^3(\underline{u})$ is defined by the second clause, earlier techniques can be used to prove that $PO(\underline{u}) = PO(\underline{u}')$ and that u_i and u_i' are equal (up to a positive linear transformation) on this set. Monotonicity follows.

39 Strictly speaking, Example 3 does not support this conclusion, since the Pareto condition is weakened. But the point is that f^3 is far from being dictatorial, although the Pareto condition almost holds. This combination is probably impossible for strategy-proof decision schemes; see Section 7.

40 Following the tradition of social choice theory, these procedures have been called "dictatorial." Considering its strong negative connotation, this is perhaps not the best name. In some situations, there is nothing wrong with "dictatorial" decisions. For example, one person may be an undisputed expert to whose judgment everybody is willing to yield, or the decision may be of such a nature that one person has a moral claim on making it alone. In many other cases, dictatorship is undesirable, and the results in this paper are by no means vacuous.

41 See note 5.

42 In other words, if every person $j \neq i$ uses strategy selection $\tilde{s}_j$, then it is optimal for i to use $\tilde{s}_i$. Hence a Bayesian equilibrium is a Nash equilibrium in another game, namely the game in which strategies are the original strategy selections. People's beliefs about each other's preferences are a part of the specification of the new game.

43 For a formal proof, see Dasgupta et al. [7, Theorem 5.2].

[44]The same argument was made in Section 6 concerning dominant strategies. It is more convincing in that connection, since the choice among dominant strategies can never affect the person's utility level. In the situation discussed in the text, if everybody is confident that everybody else will tell the truth, then nobody has any reason to do otherwise. But somehow this confidence has to be established.

[45]This result is Theorem 5.1 in Dasgupta et al. [7].

[46]The example is based on Bebchuk [5], where a similar result is proved for a more general class of decision schemes.

[47]It is not necessary to use the same limit ε in the Pareto condition and in strategy proofness. Perhaps a more natural conjecture can be formulated by fixing the two limits independently.

[48]This contrasts with Example 3 in Section 6. The decision scheme constructed there is not close to being dictatorial, but the ex ante Pareto condition can be approximated arbitrarily well.

[49]Strictly speaking, the game forms cannot approximate the decision scheme. But each game form, together with an "almost dominant" strategy selection, defines a decision scheme; see discussion in Section 4. These derived decision schemes can approximate the original scheme. To make the statement precise, ε-straightforwardness and ε-dominant strategy must be defined; see the definition of ε-strategy proofness above.

[50]There are other differences between Gibbard's results and the ones proved here. In particular, Theorem 1 and Corollary 1' make statements about all utility scales or all strategies, while Corollary 1 in Gibbard [14]

provides a characterization of a game form on a subset of the set of strategy profiles. This difference is not very important. Moreover, Gibbard's Corollary 1 can easily be strengthened so that it applies to all strategy profiles.

[51]The utility profile constructed in Lemma B1 satisfies this condition; hence this lemma proves Zeckhauser's theorem.

[52]For example, majority vote is dictatorial in Zeckhauser's sense, whenever it is defined.

REFERENCES

[1] Arrow, Kenneth J.: Essays in the Theory of Risk-Bearing. Chicago: Markham Publishing Company, 1971.

[2] Barberá, Salvador: "The Manipulation of Social Choice Mechanisms that do not Leave 'Too Much' to Chance," Econometrica, 45 (1977), 1573-1588.

[3] Barberá, Salvador: "Nice Decision Schemes," in Hans W. Gottinger and Werner Leinfellner (eds.), Decision Theory and Social Ethics: Issues in Social Choice. Dordrecht, Holland: D. Reidel Publishing Company, 1978.

[4] Barberá, Salvador: "Majority and Positional Voting in a Probabilistic Framework," Review of Economic Studies, 46 (1979), 379-389.

[5] Bebchuk, Lucian Arye: "Manipulation and Information," mimeograph, Harvard University, 1979.

[6] Dantzig, George B.: Linear Programming and Extensions. Princeton, NJ: Princeton University Press, 1963.

[7] Dasgupta, Partha, Peter Hammond and Eric Maskin: "The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility," Review of Economic Studies, 46 (1979), 185-216.

[8] Dummett, Michael and Robin Farquharson: "Stability in Voting," Econometrica, 29 (1961), 33-43.

[9] Fishburn, Peter C.: "Arrow's Impossibility Theorem: Concise Proof and Infinite Voters," Journal of Economic Theory, 2 (1970), 103-106.

[10] Gärdenfors, Peter: "Manipulation of Social Choice Functions," Journal of Economic Theory, 13 (1976), 217-228.

[11] Gärdenfors, Peter: "A Concise Proof of Theorem on Manipulation of Social Choice Functions," Public Choice, 32 (1977), 137-142.

[12] Gibbard, Allan: "Manipulation of Voting Schemes: A General Result," Econometrica, 41 (1973), 587-601.

[13] Gibbard, Allan: "Manipulation of Schemes that Mix Voting with Chance," Econometrica, 45 (1977), 665-681.

[14] Gibbard, Allan: "Straightforwardness of Game Forms with Lotteries as Outcomes," Econometrica, 46 (1978), 595-614.

[15] Hylland, Aanund: "A Note on Finite Differences and Additive Separability," mimeograph, 1979.

[16] Kalai, Ehud and Eitan Muller: "Characterization of Domains Admitting Nondictatorial Social Welfare Functions and Nonmanipulable Voting Procedures," Journal of Economic Theory, 16 (1977), 457-469.

[17] Kirman, Alan P. and Dieter Sondermann: "Arrow's Theorem, Many Agents, and Invisible Dictators," Journal of Economic Theory, 5 (1972), 267-277.

[18] Maskin, Eric: "Nash Equilibrium and Welfare Optimality," mimeograph, Massachusetts Institute of Technology, 1977.

[19] Myerson, Roger B.: "Incentive Compatibility and the Bargaining Problem," Econometrica, 47 (1979), 61-73.

[20] Pattanaik, Prasanta K.: "Threats, Counter-Threats, and Strategic Voting," Econometrica, 44 (1976), 91-103.

[21] Raiffa, Howard: Decision Analysis. Reading, MA: Addison-Wesley Publishing Company, 1968.

[22] Satterthwaite, Mark Allen: "Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions," Journal of Economic Theory, 10 (1975), 187-217.

[23] Schmeidler, David and Hugo Sonnenschein: "Two Proofs of the Gibbard-Satterthwaite Theorem on the Possibility of a Strategy-Proof Social Choice Function," in Hans W. Gottinger and Werner Leinfellner (eds.), Decision Theory and Social Ethics: Issues in Social Choice. Dordrecht, Holland: D. Reidel Publishing Company, 1978.

[24] Vickrey, William: "Utility, Strategy, and Social Decision Rules," Quarterly Journal of Economics, 74 (1960), 507-535.

[25] Zeckhauser, Richard: "Voting Systems, Honest Preferences and Pareto Optimality," American Political Science Review, 67 (1973), 934-946.

APPENDIX A

PROOF OF THEOREMS 1 AND 1*

A.1. Proof of Theorem 1

Theorem 1: Assume that the set A of alternatives is finite and has at least three elements, while the set N of individuals is finite and non-empty. Let f be a decision scheme. If f is strategy proof and satisfies the ex post Pareto condition, then f is a probability mixture of dictatorial decision schemes.

The proof is comprised of a number of lemmas. Most of the necessary notation is introduced in Section 2, but a couple of additional definitions are given here.

Let I be a subset of N and let x and y be distinct elements of A. A set D(x,I,y) of utility scales is defined in this way: The profile $\underline{u}$ is an element of D(x,I,y) if and only if $T(u_i) = \{x\}$ for all $i \in I$, $T(u_j) = \{y\}$ for all $j \notin I$, and x dominates all $z \notin \{x,y\}$ in $\underline{u}$. Obviously, $\underline{u} \in D(x,I,y)$ implies $PO(\underline{u}) \subseteq \{x,y\}$. If $\underline{u} \in D(x,I,y)$ and $\underline{u}' \in D(y,\tilde{I},x)$, then $\underline{u}$ and $\underline{u}'$ are related in the sense that $T(u_i) = T(u_i')$ for all i; the difference is that x dominates z in $\underline{u}$ while y dominates z in $\underline{u}'$, for any $z \notin \{x,y\}$.

For any set $B \subseteq A$, $\hat{B}$ shall denote the characteristic function of B. That is, $\hat{B}$ is the real-valued function defined on A which satisfies $\hat{B}(x) = 1$ for $x \in B$ and $\hat{B}(y) = 0$ for $y \notin B$. Then $\hat{B}$ can also be viewed as a vector of the same dimension as lotteries and utility scales. If ρ is a lottery, $\hat{B}\cdot\rho = \sum_{x\in B} \hat{x}\cdot\rho = \sum_{x\in B} \rho(x)$; this number is the total probability the lottery ρ assigns to elements of B.

In all the lemmas, it is assumed that the premise of Theorem 1 holds. The Pareto condition will be applied only to profiles which belong to $D(x,I,y)$ for some x,y and I. The first two lemmas establish some properties, which will be useful later, concerning the possible effect of certain changes in one individual's utility scale.

Lemma A1: Let $\underline{u}$ be a utility profile and let v_i be a utility scale, and define $\underline{v} = v_i\underline{u}_{-i}$. Then $(u_i - v_i)\cdot(f(\underline{u}) - f(\underline{v})) \geq 0$.

Proof: Individual i cannot manipulate f at $\underline{u}$, which implies $u_i \cdot f(\underline{u}) \geq u_i \cdot f(\underline{v})$. Neither can i manipulate f at $\underline{v}$, which gives $v_i \cdot f(\underline{v}) \geq v_i \cdot f(\underline{u})$, or $-v_i \cdot f(\underline{u}) \geq -v_i \cdot f(\underline{v})$. Summation gives $(u_i - v_i)\cdot f(\underline{u}) \geq (u_i - v_i)\cdot f(\underline{v})$, and the proof is complete. $\|$

Suppose, for example, that v_i is constructed from u_i by increasing the utility for some elements of A by a fixed amount, leaving everything else unchanged. Then the total probability of the elements whose utility has increased cannot decrease. Formally, let a set $B \subseteq A$ and a number $b > 0$ be given, and assume $v_i(x) = u_i(x) + b$ for $x \in B$ and $v_i(y) = u_i(y)$ for $y \notin B$. Then $u_i - v_i = -b\hat{B}$, and the lemma gives $b\hat{B}\cdot(f(\underline{u}) - f(\underline{v})) \leq 0$, or $\hat{B}\cdot f(\underline{u}) \leq \hat{B}\cdot f(\underline{v})$. This is exactly what was asserted; the total probability of the elements of B has not decreased. Note the assumption that $v_i(x) - u_i(x)$ is the same for all $x \in B$; Lemma A1 does not rule out the possibility $v_i(x) > u_i(x)$ for all $x \in B$, $v_i(y) = u_i(y)$ for all $y \notin B$, and $\hat{B}\cdot f(\underline{u}) > \hat{B}\cdot f(\underline{v})$.

Lemma A2: Let B be a subset of A and let the number ε satisfy $0 < \varepsilon < 1/2$. Moreover, let a utility profile $\underline{u}$ and a utility scale v_i be given, and write $\underline{v} = v_i\underline{u}_{-i}$. Assume $0 \leq u_i(x) \leq \varepsilon$ for all $x \in B$ and $1 - \varepsilon \leq u_i(y) \leq 1$ for all $y \notin B$. Then $\hat{B}\cdot f(\underline{u}) < \hat{B}\cdot f(\underline{v}) + 2\varepsilon$.

Proof: Write $\rho = f(\underline{u})$ and $\sigma = f(\underline{v})$. Let $b_1 = \sum_{y \notin B} \rho(y) = 1 - \hat{B}\cdot\rho$ and $b_2 = 1 - \hat{B}\cdot\sigma$. Then $u_i\cdot\rho \leq b_1 + \varepsilon(1-b_1)$, while $u_i\cdot\sigma \geq (1-\varepsilon)b_2$. Strategy proofness requires $u_i\cdot\rho \geq u_i\cdot\sigma$; hence $(1-\varepsilon)b_2 \leq b_1 + \varepsilon(1-b_1) = (1-\varepsilon)b_1 + \varepsilon$. Since $0 < \varepsilon < 1/2$, this implies $\hat{B}\cdot\rho - \hat{B}\cdot\sigma = b_2 - b_1 \leq \varepsilon/(1-\varepsilon) < 2\varepsilon$, which proves the lemma. ||

If u_i satisfies the premise of Lemma A2, it ranks all alternatives in B almost at the bottom and all alternatives outside B almost at the top. If individual i's preferences change from some v_i to this u_i, one should expect that the total probability assigned to elements of B would not increase, or at least not increase by much. This is exactly what the lemma says, and it also gives a precise upper bound on the increase.

Lemmas A3 and A4 are concerned with utility profiles which have at most two Pareto-optimal alternatives. Specifically, profiles from the sets $D(x,I,y)$ will be considered.

Lemma A3: Let x and y be distinct elements of A, and let I be a subset of N. Assume that $\underline{u},\underline{u}' \in D(x,I,y) \cup D(y,\tilde{I},x)$. Then $f(\underline{u})$ is a lottery in x and y, and $f(\underline{u}) = f(\underline{u}')$.

Proof: The Pareto condition implies that $f(\underline{u})$ is a lottery in x and y whenever $\underline{u}$ belongs to $D(x,I,y)$ or $D(y,\tilde{I},x)$. Choose a utility profile $\underline{v}$ such that $v_i(x) > v_i(y) > v_i(z)$ and $v_j(y) > v_j(x) > v_j(z)$ for all $i \in I$, $j \notin I$ and $z \notin \{x,y\}$. The profile $\underline{v}$, which is an element both of $D(x,I,y)$ and of $D(y,\tilde{I},x)$, will be kept fixed throughout the proof of this lemma.

Define $\underline{v}^{(0)} = \underline{v}$, and $\underline{v}^{(k)} = u_k \underline{v}^{(k-1)}_{-k}$ for $k = 1,\ldots,n$. Then $\underline{v}^{(n)} = \underline{u}$. For any $k \in N$, if x (or y) dominates all $z \notin \{x,y\}$ in $\underline{u}$, then x (or y) dominates all these z in $\underline{v}^{(k)}$. The utility scale $v_i^{(k)}$ is either u_i or v_i; hence $T(v_i^{(k)})$ is $\{x\}$ if $i \in I$ and $\{y\}$ otherwise. Therefore, $\underline{v}^{(k)} \in D(x,I,y) \cup D(y,\tilde{I},x)$ for $k = 1,\ldots,n$. The same obviously holds for $k = 0$. Hence $f(\underline{v}^{(k)})$ is a lottery in x and y, for all $k = 0,\ldots,n$. For any $k \in N$, let $\eta = f(\underline{v}^{(k)}) - f(\underline{v}^{(k-1)})$; then $\eta = b\hat{x} - b\hat{y}$ for some real number b. If $\eta \neq 0$, then $v_k^{(k-1)} \cdot \eta$ and $v_k^{(k)} \cdot \eta$ are either both strictly positive or both strictly negative. (The sign depends on the sign of b and on whether $k \in I$.) In any case, k can manipulate f at either $\underline{v}^{(k-1)}$ or $\underline{v}^{(k)}$. Hence $\eta = 0$. Since $k \in N$ was arbitrary, this implies $f(\underline{v}) = f(\underline{v}^{(0)}) = f(\underline{v}^{(n)}) = f(\underline{u})$.

A similar argument can be used to prove $f(\underline{u}') = f(\underline{v})$. Hence $f(\underline{u}) = f(\underline{u}')$, and the proof is complete. ||

For any x, y and I with $x \neq y$, a number $a(x,I,y)$ is defined by $a(x,I,y) = \hat{x} \cdot f(\underline{u})$ for some $\underline{u} \in D(x,I,y)$. Lemma A3 shows that the choice of $\underline{u}$ does not matter. The definition implies $0 \leq a(x,I,y) \leq 1$. Moreover, $a(y,\tilde{I},x) = \hat{y} \cdot f(\underline{u}')$, where $\underline{u}'$ is some element of $D(y,\tilde{I},x)$; therefore, Lemma A3 gives $a(y,\tilde{I},x) = 1 - a(x,I,y)$. If $\underline{u} \in D(x,N,y)$, then $PO(\underline{u}) = \{x\}$; hence the Pareto condition gives $a(x,N,y) = 1$. A similar argument gives $a(x,\phi,y) = 0$.

The next lemma says that $a(x,I,y)$ depends only on I, not on x and y.

Lemma A4: Let x, y, z and w be alternatives satisfying $x \neq y$ and $z \neq w$, and let I be a subset of N. Then $a(x,I,y) = a(z,I,w)$.

Proof: The first step is to prove $a(x,I,y) = a(x,I,z)$, where x, y and z are distinct alternatives. Let $b = a(x,I,y) - a(x,I,z)$, and assume $b > 0$. (The case $b < 0$ can be treated similarly.) Choose a positive number ε such that $2n\varepsilon < b$, which implies $\varepsilon < 1/2$. Construct utility profiles $\underline{u}$ and $\underline{v}$ such that the following holds: For $i \in I$, $u_i(x) > u_i(y) > u_i(z) > u_i(w)$ for $w \notin \{x,y,z\}$, and $v_i = u_i$. For $j \notin I$, $u_j(y) = 1$, $u_j(z) = 1 - \varepsilon/2$, $u_j(x) = 0$ and $u_j(w) = 1 - \varepsilon$, while $v_j(z) > v_j(x) > v_j(y) > v_j(w)$, where w again denotes any element except x, y and z. In $\underline{u}$, y dominates all alternatives except x and y; therefore, $\underline{u} \in D(y,\tilde{I},x)$, and $\hat{x}\cdot f(\underline{u}) = 1 - a(y,\tilde{I},x) = a(x,I,y)$. In $\underline{v}$, x dominates all alternatives but x and z, which implies $\underline{v} \in D(x,I,z)$ and $\hat{x}\cdot f(\underline{v}) = a(x,I,z)$. Hence $\hat{x}\cdot f(\underline{u}) - \hat{x}\cdot f(\underline{v}) = b$. Define $\underline{v}^{(0)} = \underline{v}$ and $\underline{v}^{(k)} = u_k \underline{v}^{(k-1)}_{-k}$ for $k = 1,\ldots,n$. This gives $\underline{v}^{(k-1)} = v_k \underline{v}^{(k)}_{-k}$. If $k \in I$, then $u_k = v_k$ and $f(\underline{v}^{(k-1)}) = f(\underline{v}^{(k)})$. For $k \notin I$, u_k and v_k satisfy the premise of Lemma A2, with $B = \{x\}$, and $\underline{v}^{(k-1)}$ and $\underline{v}^{(k)}$ playing the roles of $\underline{v}$ and $\underline{u}$. Hence $\hat{x}\cdot f(\underline{v}^{(k)}) < \hat{x}\cdot f(\underline{v}^{(k-1)}) + 2\varepsilon$. Since $\underline{v}^{(0)} = \underline{v}$ and $\underline{v}^{(n)} = \underline{u}$, an induction argument gives $\hat{x}\cdot(f(\underline{u}) - f(\underline{v})) < 2n\varepsilon < b$. This contradicts a previous statement and proves $a(x,I,y) = a(x,I,z)$.

A similar argument can be used to prove $a(x,I,y) = a(z,I,y)$.

The facts just proved will now be used repeatedly, with different triples of alternatives playing the roles of x, y and z, to prove the lemma. In particular, if x, y, z and w are all different, then $a(x,I,y) = a(z,I,y) = a(z,I,w)$. If $x = z$ or $y = w$, the conclusion of the lemma follows immediately from the facts just stated. If $x = w$ and $y \neq z$, then $a(x,I,y) = a(z,I,y) = a(z,I,x)$, and similarly if $x \neq w$ and $y = z$. Finally, suppose that $x = w$ and $y = z$. For any alternative $z' \notin \{x,y\}$, $a(x,I,y) = a(x,I,z') = a(y,I,z') = a(y,I,x)$ follows. (Here the assumption that A has at least three elements is needed.) This covers all possibilities and the proof of Lemma A4 is complete. ||

For any $I \subseteq N$, a number $a(I)$ can now be defined by $a(I) = a(x,I,y)$, where x and y are arbitrary distinct elements of A. Earlier remarks imply $a(N) = 1$, $a(\phi) = 0$, and $a(\tilde{I}) = a(x,\tilde{I},y) = 1 - a(y,I,x) = 1 - a(I)$.

The utility profiles considered in the next lemma can have an arbitrary number of Pareto-optimal elements. It is assumed, however, that each individual has a unique top-ranked alternative, and that only two alternatives occur at the top of anybody's ranking. It turns out that only these two alternatives get positive probability.

<u>Lemma A5</u>: Let I be a subset of N, and let x and y be distinct elements of A. Assume that the utility profile $\underline{u}$ satisfies $T(u_i) = \{x\}$ for all $i \in I$, and $T(u_j) = \{y\}$ for all $j \notin I$. Then $f(\underline{u})$ is a lottery in x and y, $\hat{x} \cdot f(\underline{u}) = a(I)$, and $\hat{y} \cdot f(\underline{u}) = 1 - a(I) = a(\tilde{I})$.

Proof: First, the inequality $\hat{x} \cdot f(\underline{u}) \leq a(I)$ will be proved. To this end, construct a utility profile $\underline{v}$ in the following way: For $i \in I$, let $v_i = u_i$. For $j \notin I$, v_j is constructed from u_j by moving x upwards until it becomes the second-ranked alternative. Formally, $v_j(z) = u_j(z)$ for all $z \neq x$, and $v_j(x)$ is any number which satisfies $v_j(x) < u_j(y)$ and $v_j(x) > u_j(w)$ for all $w \neq y$ (including $w = x$). Since $T(u_j) = \{y\}$ and A is finite, such a number exists. In $\underline{v}$, x dominates every alternative except x and y; therefore, $\underline{v} \in D(x,I,y)$ and $\hat{x} \cdot f(\underline{v}) = a(I)$. Let $j \notin I$ and consider the effect of substituting v_j for u_j in some utility profile. Since $v_j - u_j = b\hat{x}$ for some number $b > 0$, Lemma A1 implies that this move cannot decrease the probability assigned to x. (See comments after the proof of that lemma.) When $i \in I$, nothing changes if v_i is substituted for u_i. Hence $\hat{x} \cdot f(\underline{v}) \geq \hat{x} \cdot f(\underline{u})$, which implies $\hat{x} \cdot f(\underline{u}) \leq a(I)$.

The situation is symmetrical, and a similar argument can be used to prove $\hat{y} \cdot f(\underline{u}) \leq a(\tilde{I}) = 1 - a(I)$.

Now assume that the conclusion of the lemma is wrong. Then at least one of $\hat{x} \cdot f(\underline{u}) < a(I)$ or $\hat{y} \cdot f(\underline{u}) < a(\tilde{I})$ must hold. Together with the inequalities proved above, this implies $(\hat{x} + \hat{y}) \cdot f(\underline{u}) < 1$, and there must exist an element $z \notin \{x,y\}$ with $\hat{z} \cdot f(\underline{u}) > 0$. Choose one such z; this alternative will be kept fixed in the rest of the proof of Lemma A5.

Now the profile $\underline{u}'$ shall be constructed by moving z upwards in everybody's utility scale until it becomes the second-ranked alternative; everything else is left unchanged. Formally, u_i' for $i \in I$ shall satisfy $u_i'(w) = u_i(w)$ for all $w \neq z$, $u_i'(z) < u_i(x)$, and $u_i'(z) > u_i(w)$ for all $w \neq x$. The utility scale u_j' for $j \notin I$ is constructed in the same way, but with y substituted for x. As above, Lemma A1 can be used to prove $\hat{z} \cdot f(\underline{u}') \geq \hat{z} \cdot f(\underline{u})$, which implies $\hat{z} \cdot f(\underline{u}') > 0$. Therefore, $(\hat{x} + \hat{y}) \cdot f(\underline{u}') < 1$, and at least one of $\hat{x} \cdot f(\underline{u}') < a(I)$ and $\hat{y} \cdot f(\underline{u}') < 1 - a(I) = a(\tilde{I})$ must hold. By symmetry, these two cases can be treated similarly; hence there is no loss of generality in assuming $\hat{x} \cdot f(\underline{u}') < a(I)$.

Let $b = a(I) - \hat{x} \cdot f(\underline{u}')$ and choose a positive number ε such that $2n\varepsilon < b$. A utility profile $\underline{v}$ shall be constructed in this way: For $i \in I$, $v_i = u_i'$. For $j \notin I$, $v_j(z) = 1$, $v_j(x) = 0$, and $v_j(w) = 1 - \varepsilon$ for all $w \notin \{x,z\}$. For all $i \in I$, $v_i = u_i'$ ranks x highest and z second highest. Hence all alternatives but x and z are dominated by z in $\underline{v}$. Therefore, $\underline{v} \in D(z,\tilde{I},x)$, and $\hat{x} \cdot f(\underline{v}) = 1 - a(\tilde{I}) = a(I)$. Consider the effect of going from $\underline{u}'$ to $\underline{v}$ by changing one individual's utility scale at a time. A change from u_j' to v_j for some $j \notin I$ satisfies the premise of Lemma A2; here v_j and u_j' correspond to u_i and v_i, respectively, and $B = \{x\}$. The lemma implies that if this change increases the probability assigned to x, the increase is less than 2ε. Combined with the fact that $v_i = u_i'$ for $i \in I$, this implies $\hat{x} \cdot f(\underline{v}) < \hat{x} \cdot f(\underline{u}') + 2n\varepsilon < \hat{x} \cdot f(\underline{u}') + b = a(I)$. This contradicts an earlier observation about $\underline{v}$, and the proof of Lemma A5 is complete.

Note that this proof can be applied even if $I = \phi$ or $I = N$. In these cases, however, the lemma follows immediately from earlier statements. For example, if $I = N$ and $\underline{u}$ satisfies the premise of the lemma, then $\underline{u} \in D(x,N,y)$ for any $y \neq x$, and the conclusion follows from the definition of $a(x,I,y)$. ||

Suppose that every individual in a set I has x as the unique top-ranked alternative. Presumably, x will then have at least as strong a position as it has in the profile described in Lemma A5. (The position may be stronger, since nothing is now assumed about persons outside I; for example, some of these may also rank x at the top.) This is exactly the statement proved in the next lemma.

Lemma A6: Let I be a subset of N, let $\underline{u}$ be a utility profile, and let x be an alternative. Assume $T(u_i) = \{x\}$ for all $i \in I$. Then $\hat{x} \cdot f(\underline{u}) \geq a(I)$.

Proof: Suppose that I, $\underline{u}$ and x provide a counterexample to the lemma, and define $b = a(I) - \hat{x} \cdot f(\underline{u})$. Then b is positive, and it is possible to find a positive number ε such that $2n\varepsilon < b$. Fix an alternative $y \neq x$, and define a utility profile $\underline{v}$ by $v_i = u_i$ for all $i \in I$, and $v_j(x) = 0$, $v_j(y) = 1$, and $v_j(z) = 1 - \varepsilon$ for all $j \notin I$ and all $z \notin \{x,y\}$. Lemma A2 can be applied to a change from u_j to v_j for some $j \notin I$; v_j and u_j play the roles of u_i and v_i, and $B = \{x\}$. By an argument similar to the one used above, this implies $\hat{x} \cdot f(\underline{v}) < \hat{x} \cdot f(\underline{u}) + 2n\varepsilon < \hat{x} \cdot f(\underline{u}) + b = a(I)$. But $\underline{v}$ satisfies the premise of Lemma A5, which implies $\hat{x} \cdot f(\underline{v}) = a(I)$. This contradiction proves Lemma A6. ||

The function a(I) will now be investigated. This function turns out to be a probability measure on N. That is, each individual has a non-negative "weight" or probability, the weights sum to 1, and a(I) is equal to the sum of the weights for the members of I.

For any $i \in N$, a number a_i is defined by $a_i = a(\{i\})$. These numbers are the weights mentioned above. Clearly, $0 \leq a_i \leq 1$ for all i.

Lemma A7:

(a) If $I \subseteq N$, $J \subseteq N$, and $I \cap J = \phi$, then $a(I \cup J) = a(I) + a(J)$.

(b) For all $I \subseteq N$, $a(I) = \sum_{i \in I} a_i$.

(c) $\sum_{i \in N} a_i = 1$.

Proof: Let I and J satisfy the premise of part (a), and define $K = N \setminus (I \cup J)$. Let $b = 1 - (a(I) + a(J) + a(K))$. Choose a number ε which satisfies $0 < \varepsilon < 1$ if $b \leq 0$, and $0 < \varepsilon$ and $6n\varepsilon < b$ if $b > 0$. Moreover, choose three distinct elements x, y and z from A and define $B = A \setminus \{x,y,z\}$. (The assumption that A has at least three elements is needed in the proof of this lemma.) Construct a utility profile $\underline{u}$ which satisfies $u_i(x) = u_j(y) = u_k(z) = 1$, $u_i(y) = u_j(z) = u_k(x) = 1 - \varepsilon$ and $u_i(z) = u_j(x) = u_k(y) = \varepsilon$ for $i \in I$, $j \in J$ and $k \in K$, and $u_i(w) = 0$ for all $i \in N$ and $w \in B$. Lemma A6 can be used to conclude $\hat{x} \cdot f(\underline{u}) \geq a(I)$, $\hat{y} \cdot f(\underline{u}) \geq a(J)$ and $\hat{z} \cdot f(\underline{u}) \geq a(K)$. Since $f(\underline{u})$ is a lottery, $\hat{x} \cdot f(\underline{u}) + \hat{y} \cdot f(\underline{u}) + \hat{z} \cdot f(\underline{u}) \leq 1$, which implies $a(I) + a(J) + a(K) \leq 1$ and $b \geq 0$. Then assume $b > 0$. Choose a utility profile $\underline{v}$ which satisfies $v_i = u_i$ for $i \in I \cup K$, and $T(v_j) = \{z\}$ for $j \in J$. For $j \in J$, Lemma A2 can be applied to a change in individual j's utility scale from u_j to v_j. The set B of Lemma A2 corresponds to $B \cup \{x\}$ here. This can be used to conclude $\hat{x} \cdot f(\underline{u}) + \hat{B} \cdot f(\underline{u}) \leq \hat{x} \cdot f(\underline{v}) + \hat{B} \cdot f(\underline{v}) + 2n\varepsilon < \hat{x} \cdot f(\underline{v}) + \hat{B} \cdot f(\underline{v}) + b/3$. The profile $\underline{v}$ satisfies the premise of Lemma A5, with z substituted by y. Hence $\hat{x} \cdot f(\underline{v}) + \hat{B} \cdot f(\underline{v}) = \hat{x} \cdot f(\underline{v}) = a(I)$, which implies $\hat{x} \cdot f(\underline{u}) + \hat{B} \cdot f(\underline{u}) < a(I) + b/3$. Since the situation is symmetrical, similar inequalities can be proved with y and J substituted for x and I, and with z and K in these roles. The left-hand sides of these strict inequalities sum to at least 1, while the right-hand sides sum to $a(I) + a(J) + a(K) + b = 1$. This is impossible. Hence the assumption $b > 0$ must be wrong, and $b = 0$ follows. (For later reference, the Pareto condition has not been used here. If it were used, the argument could be simplified slightly, since it would imply $\hat{B} \cdot f(\underline{u}) = 0$.)

In other words, $a(I) + a(J) + a(K) = 1$. By earlier remarks, $a(\tilde{K}) = 1 - a(K)$; which implies $a(\tilde{K}) = a(I) + a(J)$. Since $\tilde{K} = I \cup J$, part (a) is proved.

Part (b) is proved by induction on the number of elements in I. The conclusion follows from earlier remarks when $I = \phi$ and from the definition of a_i when $I = \{i\}$. Let the integer m satisfy $1 \leq m < n$ and assume, as an induction hypothesis, that part (b) holds for all sets with m or fewer elements. Suppose that I has m+1 elements, and let i be an element of I. Part (a) gives $a(I) = a(I \setminus \{i\}) + a(\{i\})$. The induction hypothesis can be applied to $I \setminus \{i\}$ and $\{i\}$, and it follows that the conclusion holds for the given I. This completes the induction step and proves part (b).

As observed earlier, $a(N) = 1$. Part (c) follows by applying part (b) with $I = N$. ||

If the domain of definition for f is restricted to the set of profiles where each individual has a unique alternative with maximal utility, the theorem can now easily be proved. Let $\underline{u}$ be such a profile and define, for each x, $I_x = \{i | T(u_i) = \{x\}\}$. Then Lemma A6 gives $\hat{x} \cdot f(\underline{u}) \geq a(I_x)$. If these inequalities are added for all x, the left-hand sides sum to 1 since $f(\underline{u})$ is a lottery, while the right-hand sides sum to 1 by Lemma A7. Hence equality holds for all x. On the set of profiles considered here, there is a unique decision scheme which is dictatorial for i. If this scheme is called f_i, it is now easy to see that $f(\underline{u}) = \sum_{i \varepsilon N} a_i f_i(\underline{u})$ for any profile $\underline{u}$ of this type. Since the numbers a_i are non-negative and sum to 1, this is exactly the conclusion of the theorem.

Some more work is needed, however, to remove the condition that every individual has a unique top-ranked alternative.

Lemma A8: Let $\underline{u}$ be a utility profile and let I be a subset of N. Define $B \subseteq A$ by $B = \bigcup_{i \varepsilon I} T(u_i)$. Then $\hat{B} \cdot f(\underline{u}) \geq a(I)$.

Proof: Suppose that the lemma is incorrect for $\underline{u}$ and I, let $b = a(I) - \hat{B} \cdot f(\underline{u})$, and choose ε such that $0 < \varepsilon < b/n$. Let C denote the set $A \setminus B$. A utility profile $\underline{v}$ shall be constructed in the following way: For $j \notin I$, let $v_j = u_j$. Then assume $i \varepsilon I$, let x_i be an arbitrary element of $T(u_i)$, and define $c_i = u_i(x_i) - \max_{y \varepsilon C} u_i(y)$. Since C is finite and C and $T(u_i)$ are disjoint sets, c_i is strictly positive. The utility scale v_i shall satisfy $v_i(x_i) = u_i(x_i) + c_i \varepsilon$, $v_i(x) = u_i(x)$ for $x \varepsilon B \setminus \{x_i\}$, and $v_i(y) = u_i(y) + c_i$ for $y \varepsilon C$. (These definitions require $C \neq \phi$. If $C = \phi$, then $\hat{B} \cdot f(\underline{u}) = 1$, and the lemma is trivially true.)

In order to discuss the effect of substituting v_i for u_i for some $i \varepsilon I$, let $\underline{u}'$ be a profile with $u_i' = u_i$, and define $\eta = f(\underline{u}') - f(v_i \underline{u}'_{-i})$. Lemma A1 implies $(v_i - u_i) \cdot \eta \leq 0$. Since $v_i - u_i = c_i \varepsilon \hat{x}_i + c_i \hat{C}$ and $c_i > 0$, this implies $\varepsilon \hat{x}_i \cdot \eta + \hat{C} \cdot \eta \leq 0$. The number $\hat{x}_i \cdot \eta$ is the difference between the probability assigned to x_i in two different lotteries; hence $|\hat{x}_i \cdot \eta| \leq 1$. This implies $\hat{C} \cdot \eta \leq \varepsilon$. The lotteries $f(\underline{u})$ and $f(\underline{v})$ differ by at most n changes of this type. This gives $\hat{C}\cdot(f(\underline{u}) - f(\underline{v})) \leq n\varepsilon$ and, since $(\hat{B} + \hat{C})\cdot f(\underline{u}) = (\hat{B} + \hat{C})\cdot f(\underline{v}) = 1$, $\hat{B}\cdot f(\underline{v}) \leq \hat{B}\cdot f(\underline{u}) + n\varepsilon < \hat{B}\cdot f(\underline{u}) + b = a(I)$.

For $i \varepsilon I$, $T(v_i)$ has one element, namely x_i. Define, for each $x \varepsilon B$, $I_x = \{i \varepsilon I | T(v_i) = \{x\}\}$. The sets I_x are disjoint. For all $i \varepsilon I$, x_i is an element of B, which implies $\bigcup_{x\varepsilon B} I_x = I$. Then Lemma A7 gives $\sum_{x\varepsilon B} a(I_x) = a(I)$. For each $x \varepsilon B$, Lemma A6 can be applied, with I_x and $\underline{v}$ playing the roles of I and $\underline{u}$, to conclude that $\hat{x}\cdot f(\underline{v}) \geq a(I_x)$ for all $x \varepsilon B$. These inequalities add up to $\hat{B}\cdot f(\underline{v}) \geq a(I)$, which contradicts the conclusion of the previous paragraph and proves Lemma A8. ||

The next lemma will complete the proof of Theorem 1. This last lemma depends only on Lemmas A7 and A8; its proof does not make direct use of the premise of the theorem or of the earlier lemmas.

<u>Lemma A9</u>: Assume that the non-negative numbers a_i for $i \varepsilon N$ and $a(I)$ for $I \subseteq N$ satisfy Lemma A7. Let f be a decision scheme, and assume that Lemma A8 holds. Then there exist decision schemes $f_1,\ldots,f_n$ such that, for each $i \varepsilon N$, f_i is dictatorial for i, and $f = a_1f_1 + \ldots + a_nf_n$.

<u>Proof</u>: Since $a_1 + \ldots + a_n = 1$, the conclusion of the lemma is exactly the conclusion of the theorem. If $a_i = 0$ for some i, f_i can be any decision scheme which is dictatorial for i, and i can be eliminated from further consideration. Hence there is no loss of generality in assuming $a_i > 0$ for all i, and this assumption is made from now on.

The statement to be proved does not require any special relationship between $f_i(\underline{u})$ and $f_i(\underline{v})$ for $\underline{u} \neq \underline{v}$ (except what follows from the condition that f_i be dictatorial for i). Hence the construction of $f_1,\ldots,f_n$ can be done separately for each utility profile. Let a profile $\underline{u}$ be given, and let $\rho = f(\underline{u})$. For each i, a lottery shall be chosen as the value of $f_i(\underline{u})$. These lotteries must satisfy $\rho = \sum_{i\epsilon N} a_i f_i(\underline{u})$, and $\hat{x}\cdot f_i(\underline{u}) = 0$ if $x \notin T(u_i)$. If $T(u_i)$ has only one element, this leaves only one possibility, namely $f_i(\underline{u}) = \hat{x}_i$, where x_i is the unique element of $T(u_i)$. Otherwise, $f_i(\underline{u})$ may depend on all of $\underline{u}$ and not just on u_i. If $T(u_i)$ has one element for all i, the required conclusion is easily proved; see comments after Lemma A7. (Lemma A8 can be used instead of Lemma A6.) In this case, the discussion below is correct but redundant.

Consider the following linear programming problem:

Variables: b_{xi}, for $x \epsilon A$, $i \epsilon N$

Objective: Minimize $\sum_{x\notin T(u_i)} b_{xi}$

Constraints:

$$\sum_{x\epsilon A} b_{xi} = a_i \qquad \text{for all } i \tag{1}$$

$$\sum_{i\epsilon N} b_{xi} = \rho(x) \qquad \text{for all } x \tag{2}$$

$$b_{xi} \geq 0 \qquad \text{for all } x \text{ and } i.$$

The dual of this problem can be written in this way:

Variables: c_i for $i \epsilon N$

d_x for $x \epsilon A$

Objective: Maximize $\sum_i a_i c_i - \sum_x \rho(x) d_x$

$$\text{Constraints:}\quad c_i - d_x \leq 0 \quad \text{when} \quad x \in T(u_i) \tag{3}$$

$$c_i - d_x \leq 1 \quad \text{when} \quad x \notin T(u_i)\ .$$

Note that c_i and d_x are unrestricted in sign; this corresponds to the fact that (1) and (2) are equality constraints. (The sign of d_x has been changed compared to the "standard" form of the dual problem; this makes no difference, since d_x is an unrestricted variable.) For a discussion of linear programming problems and duality theory, see, for example, Dantzig [6 , Chapter 6].

Now assume that c_i for $i \in N$ and d_x for $x \in A$ represent a feasible solution to the dual problem. The numbers c_i can be viewed as the possible values of a random variable where a_i is the probability of c_i occurring. Similarly, the numbers d_x can be regarded as the observations of a random variable whose probability distribution is given by ρ. Define functions F and G by

$F(\alpha) = \sum_{c_i \leq \alpha} a_i$ and $G(\alpha) = \sum_{d_x \leq \alpha} \rho(x)$, for all real numbers α. Then F and G are the cumulative distribution functions of c_i and d_x, respectively.

The next step is to prove $F(\alpha) \geq G(\alpha)$ for all α. Let α be given, define $I = \{i \in N | c_i > \alpha\}$, and define B as in Lemma A8. For each $x \in B$, there exists an $i \in I$ such that $x \in T(u_i)$, and constraint (3) gives $d_x \geq c_i > \alpha$. If B' is defined by $B' = \{x \in A | d_x > \alpha\}$, this implies $B \subseteq B'$. Lemma A8 now gives $a(I) \leq \hat{B} \cdot f(\underline{u}) = \sum_{x \in B} \rho(x) \leq \sum_{x \in B'} \rho(x)$. But $F(\alpha) = 1 - a(I)$ and

$G(\alpha) = 1 - \sum_{x \in B'} \rho(x)$, hence the inequality $F(\alpha) \geq G(\alpha)$ has been established. This means that the d_x-distribution, in a probabilistic sense, lies uniformly above the c_i-distribution (or at least not below). Therefore, the former has at least as high an expected value as the latter. In symbols, $\sum_x \rho(x)d_x \geq \sum_i a_i c_i$. (It is a straightforward but tedious exercise to convert this argument into a formal and purely algebraic proof.)

The conclusion of the last paragraph is that the value of the objective of the dual problem is always non-positive. The value can be 0, since a feasible solution is given by $c_i = d_x = 0$ for all i and x. Therefore, the optimal value for the dual problem is 0.

There exist feasible solutions to the primal problem. For example, a solution is given by $b_{ix} = \rho(x)a_i$. These numbers are non-negative and (1) and (2) hold because $\sum_x \rho(x) = \sum_i a_i = 1$. According to the duality theorem of linear programming, the optimal value of the objective is then the same for the primal and the dual problem. Let b^*_{xi} be an optimal solution to the primal problem. Then the objective has value 0, which implies $b^*_{xi} = 0$ for all x and i with $x \notin T(u_i)$. The lotteries $f_i(\underline{u})$ can now be defined by

$$\hat{x} \cdot f_i(\underline{u}) = b^*_{xi}/a_i \quad ,$$

for all x and i. Constaint (1) shows that $f_i(\underline{u})$ actually is a lottery, while (2) gives $\rho = \sum_i a_i f_i(\underline{u})$. Moreover, $\hat{x} \cdot f_i(\underline{u}) = 0$ if $x \notin T(u_i)$. All the requirements are satisfied, and the proof of Lemma A9 and Theorem 1 is complete. ||

A.2. Proof of Theorem 1*

Theorem 1*: Assume that the set A of alternatives is finite and has at least three elements, while the set N of individuals is finite and non-empty.

Let f be a decision scheme. If f is strategy proof and satisfies the attainability condition, then f is a probability mixture of dictatorial decision schemes.

In the proof of Theorem 1, the Pareto condition is only applied to utility profiles which belong to $D(x,I,y)$ for some x, y and I. Hence it is sufficient to prove the Pareto conditions for these profiles. This is done in the following series of lemmas. All the time, it is assumed that the premise of Theorem 1* holds. Lemmas A1 and A2 can be used, since they are proved above without reference to the Pareto condition.

Lemma A10: Let x and $\underline{u}$ satisfy $T(u_i) = \{x\}$ for all $i \in N$. Then $f(\underline{u}) = \hat{x}$.

Proof: By attainability, there exists a profile $\underline{v}$ such that $f(\underline{v}) = \hat{x}$. Strategy proofness implies $u_1 \cdot f(u_1\underline{v}_{-1}) \geq u_1 \cdot f(\underline{v})$; otherwise, individual 1 can manipulate f at the profile $u_1\underline{v}_{-1}$. Since $u_1 \cdot \rho < u_1 \cdot \hat{x}$ for all $\rho \neq \hat{x}$, this gives $f(u_1\underline{v}_{-1}) = \hat{x}$. If now u_2 is substituted for v_2 in $u_1\underline{v}_{-1}$, a similar argument shows that the outcome must remain equal to $\hat{x}$. After n such substitutions the conclusion $f(\underline{u}) = \hat{x}$ follows. ||

Any profile which belongs to some set $D(x,N,y)$ or $D(x,\phi,y)$, is of the form considered in Lemma A10. Therefore, the Pareto condition has been proved for these profiles. Choose a set $I \subseteq N$ such that $I \neq \phi$ and $I \neq N$, choose distinct alternatives x and y, and write $B = A \setminus \{x,y\}$. These entities will be kept fixed throughout Lemmas A11 - A13.

<u>Lemma A11</u>: Let i and j be distinct individuals, and let u_k for $k \notin \{i,j\}$ be fixed utility scales. For any numbers α and β with $\alpha < 1$ and $0 < \beta < 1$, choose utility scales u_i and u_j such that $u_i(x) = u_j(y) = 1$, $u_i(y) = \alpha$, $u_j(x) = \beta$, and $u_i(z) = u_j(z) = 0$ for all $z \in B$, and define $h^B(\alpha,\beta) = \hat{B}\cdot f(\underline{u})$. Then the function h^B is additively separable in α and β. That is

$$h^B(\alpha_1,\beta_1) - h^B(\alpha_1,\beta_2) - h^B(\alpha_2,\beta_1) + h^B(\alpha_2,\beta_2) = 0 , \tag{4}$$

for all α_1, α_2, β_1, β_2 with $\alpha_1 < 1$, $\alpha_2 < 1$, $0 < \beta_1 < 1$ and $0 < \beta_2 < 1$.

<u>Proof</u>: By definition, $h^B(\alpha,\beta)$ is additively separable if it can be written as the sum of a function of α and a function of β. Separability obviously implies (4). If α_0 and β_0 are fixed numbers with $\alpha_0 < 1$ and $0 < \beta_0 < 1$, equation (4) implies $h^B(\alpha,\beta) = h^B(\alpha,\beta_0) + [h^B(\alpha_0,\beta) - h^B(\alpha_0,\beta_0)]$, which proves separability. Hence equation (4) and the separability condition are equivalent.

Functions h^x and h^y are defined in analogy with h^B; that is, when α, β and $\underline{u}$ are of the form described in the lemma, $h^x(\alpha,\beta) = \hat{x}\cdot f(\underline{u})$ and $h^y(\alpha,\beta) = \hat{y}\cdot f(\underline{u})$. Since $f(\underline{u})$ is a lottery, $h^x(\alpha,\beta) + h^y(\alpha,\beta) + h^B(\alpha,\beta) = 1$ for all α and β.

Let β and u_j be fixed and satisfy the premise of the lemma, and choose α_1 and α_2 such that $\alpha_1 < \alpha_2 < 1$. Suppose that individual i has preferences represented by a utility scale u_i as described in the lemma, with $u_i(y) = \alpha_1$. If i reports this utility scale, the outcome will have a utility value for i of $u_i\cdot f(\underline{u}) = h^x(\alpha_1,\beta) + \alpha_1 h^y(\alpha_1,\beta)$. On the other hand, if i reports preferences corresponding to $u_i(y) = \alpha_2$, the utility of the outcome, evaluated at i's true preferences, is $h^x(\alpha_2,\beta) + \alpha_1 h^y(\alpha_2,\beta)$. By strategy proofness, the latter number cannot exceed the former. Straightforward algebraic manipulations give

$$h^B(\alpha_2,\beta) - h^B(\alpha_1,\beta) \geq (\alpha_1-1)(h^y(\alpha_2,\beta) - h^y(\alpha_1,\beta)) .$$

The argument can be repeated with α_1 and α_2 interchanged. This gives

$$h^B(\alpha_2,\beta) - h^B(\alpha_1,\beta) \leq (\alpha_2-1)(h^y(\alpha_2,\beta) - h^y(\alpha_1,\beta)) .$$

The "multiplier" α-1 in the right-hand side of these two inequalities is a continuous function of α. Therefore, the inequalities imply that there exists a number α' with $\alpha_1 \leq \alpha' \leq \alpha_2$ and

$$h^B(\alpha_2,\beta) - h^B(\alpha_1,\beta) = (\alpha'-1)(h^y(\alpha_2,\beta) - h^y(\alpha_1,\beta)) . \tag{5}$$

Similarly, let fixed α and u_i be as described in the lemma, let β_1 and β_2 satisfy $0 < \beta_1 < \beta_2 < 1$, and apply strategy proofness to individual j. An argument similar to the one used above can be applied; the conclusion is that there exists a number β' with $\beta_1 \leq \beta' \leq \beta_2$ and

$$h^B(\alpha, \beta_2) - h^B(\alpha,\beta_1) = [(1-\beta')/\beta'](h^y(\alpha,\beta_2) - h^y(\alpha,\beta_1)) . \tag{6}$$

Together with certain monotonicity conditions to be detailed below, equations (5) and (6) imply that h^B is additively separable. To motivate this statement, assume for a moment that h^B and h^y are twice continuously differentiable functions. Divide equation (5) by $\alpha_2 - \alpha_1$ and let α_1 and α_2 converge to a common limit α. This gives

$$h_1^B(\alpha,\beta) = (\alpha-1)h_1^y(\alpha,\beta) ,$$

where h_1^B denotes the derivative of h^B with respect to the first argument, etc. Similarly, equation (6) implies

$$h_2^B(\alpha,\beta) = [(1-\beta)/\beta]h_2^y(\alpha,\beta) .$$

Differentiate the first of these equations with respect to β and the second with respect to α, and use the fact that the cross derivatives of a twice continuously differentiable function are equal. Since $\alpha-1 \neq (1-\beta)/\beta$, the resulting equations imply $h_{12}^y(\alpha,\beta) = 0$ and $h_{12}^B(\alpha,\beta) = 0$. Hence h^B is additively separable.

The functions h^B and h^y need not be differentiable; they need not even be continuous. One might expect that the argument of the previous paragraph can be carried through by using finite differences instead of derivatives. This is essentially true, but the proof is complicated, and some added assumptions are needed. The proof is given in [15].

Assume $0 < \alpha_1 < \alpha_2$, and consider the difference between the utility profiles which define $h^y(\alpha_1,\beta)$ and $h^y(\alpha_2,\beta)$. The only difference is that the utility of y in individual i's utility scale increases when α_2 is substituted for α_1. By Lemma A1, this change cannot decrease the probability assigned to y. That is, $h^y(\alpha,\beta)$ is a non-decreasing function of α, for any β. In equation (5), $\alpha'-1$ is a negative number; hence $h^B(\alpha,\beta)$ is non-increasing in α for any β. Equation (6) implies that $h^B(\alpha,\beta_2) - h^B(\alpha,\beta_1)$ and $h^y(\alpha,\beta_2) - h^y(\alpha,\beta_1)$ are either both negative, both 0, or both positive. As above, Lemma A1 can be used to prove that $h^x(\alpha,\beta)$ is non-decreasing in β. Since h^x, h^y and h^B sum to 1, this implies that both h^B and h^y are non-increasing functions in β, for any α. The "multiplier" α-1 in equation (5) is a continuously differentiable and increasing function on the relevant range $(-\infty, 1)$. Similarly, the term $(1-\beta)/\beta$ from (6) is continuously differentiable and decreasing on (0, 1). Finally, $\alpha-1 \neq (1-\beta)/\beta$ for all possible α and β; in fact, $\alpha-1 < 0 < (1-\beta)/\beta$.

The statements of the last paragraph, together with equations (5) and (6), establish the premise of the Proposition of [15], as generalized in Section 7 of that note. The conclusion of the Proposition is that h^B is additively separable on any closed and bounded rectangle included in its domain of definition. There will always exist such a rectangle which contains all the points mentioned in (4). This completes the proof of Lemma A11. ||

Lemma A12: Let α_i for $i \in I$ and β_j for $j \notin I$ be real numbers, and assume $\alpha_i < 1$ and $0 < \beta_j < 1$ for all $i \in I$ and $j \notin I$. Suppose that $\underline{u}$ is a utility profile which satisfies $u_i(x) = u_j(y) = 1$, $u_i(y) = \alpha_i$, $u_j(x) = \beta_j$, and $u_i(z) = u_j(z) = 0$, for $i \in I$, $j \notin I$ and $z \in B$. Then $\hat{B}\cdot f(\underline{u}) = 0$.

Proof: Let $\underline{\alpha}$ and $\underline{\alpha}'$ denote real vectors with dimension equal to the number of elements in I, while $\underline{\beta}$ and $\underline{\beta}'$ denote vectors with dimension equal to the number of elements in $\tilde{I}$. Each coordinate of $\underline{\alpha}$ corresponds to an element of I. For $i \in I$, α_i denotes the ith coordinate of $\underline{\alpha}$. Similar remarks apply to $\underline{\beta}$, $\tilde{I}$ and β_j.

For any $\underline{\alpha}$ and $\underline{\beta}$ satisfying $\alpha_i < 1$ and $0 < \beta_j < 1$ for all $i \in I$ and $j \notin I$, let $\underline{u}$ be the profile described in the lemma, and define $h(\underline{\alpha},\underline{\beta}) = \hat{B}\cdot f(\underline{u})$. The function h corresponds to h^B in Lemma A11. It is necessary to prove that h is identically 0.

The first step is to prove that $h(\underline{\alpha},\underline{\beta})$ is additively separable in $\underline{\alpha}$ and $\underline{\beta}$. That is,

$$h(\underline{\alpha},\underline{\beta}) - h(\underline{\alpha},\underline{\beta}') - h(\underline{\alpha}',\underline{\beta}) + h(\underline{\alpha}',\underline{\beta}') = 0 \quad , \tag{7}$$

for all $\underline{\alpha}$, $\underline{\alpha}'$, $\underline{\beta}$ and $\underline{\beta}'$ for which the expression is defined. If $\underline{\alpha} = \underline{\alpha}'$ or $\underline{\beta} = \underline{\beta}'$, this is obvious. Then assume that $\underline{\alpha}$ and $\underline{\alpha}'$ differ on exactly one coordinate, and that $\underline{\beta}$ and $\underline{\beta}'$ also differ on exactly one coordinate. Choose $i \in I$ and $j \notin I$ such that $\alpha_i \neq \alpha_i'$ and $\beta_j \neq \beta_j'$; these choices are unique by assumption. Lemma A11 can be applied, with i and j chosen this way; then (7) follows from (4). Next, let $\underline{\alpha}$ and $\underline{\alpha}'$ differ on one coordinate while no special assumptions are made about $\underline{\beta}$ and $\underline{\beta}'$. For some positive integer m, it is possible to find $\underline{\beta}^{(0)},\ldots,\underline{\beta}^{(m)}$, such that $\underline{\beta}^{(0)} = \underline{\beta}$, $\underline{\beta}^{(m)} = \underline{\beta}'$, $h(\underline{\alpha},\underline{\beta}^{(k)})$ and $h(\underline{\alpha}',\underline{\beta}^{(k)})$ are defined for $k = 0,\ldots,m$, and $\underline{\beta}^{(k-1)}$ and $\underline{\beta}^{(k)}$ differ on only one coordinate for $k = 1,\ldots,m$. By the previous result, (4) holds when $\underline{\beta}^{(k-1)}$ and $\underline{\beta}^{(k)}$ are substituted for $\underline{\beta}$ and $\underline{\beta}'$. These m equalities can be summed, and the result proves equation (4) for the given $\underline{\alpha}$, $\underline{\alpha}'$, $\underline{\beta}$ and $\underline{\beta}'$. Finally, let $\underline{\alpha}$, $\underline{\alpha}'$, $\underline{\beta}$ and $\underline{\beta}'$ be any vectors for which (4) is defined. A sequence $\underline{\alpha} = \underline{\alpha}^{(0)},\underline{\alpha}^{(1)},\ldots,\underline{\alpha}^{(m)} = \underline{\alpha}'$ can be constructed, such that $\underline{\alpha}^{(k-1)}$ and $\underline{\alpha}^{(k)}$ differ on only one coordinate. The previous result shows that (4) holds with $\underline{\alpha}^{(k-1)}$ and $\underline{\alpha}^{(k)}$ substituted for $\underline{\alpha}$ and $\underline{\alpha}'$. Equation (4) follows by summing these equalities.

Let a number $\varepsilon > 0$ be given, and let $\underline{\alpha}$ and $\underline{\beta}$ be such that $h(\underline{\alpha},\underline{\beta})$ is defined. In addition, assume $1 - \varepsilon < \beta_j < 1$ for all $j \notin I$. Let $\underline{u}$ be the utility profile which defines $h(\underline{\alpha},\underline{\beta})$ and let $\underline{v}$ be a profile which satisfies $v_i = u_i$ for $i \in I$ and $T(v_j) = \{x\}$ for $j \notin I$. Lemma A10 implies $f(\underline{v}) = \hat{x}$. For any $j \notin I$, consider the effect of substituting u_j for v_j in some utility profile. Lemma A2 can be used to prove that the substitution cannot increase the total probability assigned to elements of B by as much as 2ε. The profile $\underline{u}$ can be obtained from $\underline{v}$ by fewer than n such substitutions; hence $\hat{B}\cdot f(\underline{u}) < \hat{B}\cdot f(\underline{v}) + 2n\varepsilon = 2n\varepsilon$, and $h(\underline{\alpha},\underline{\beta}) < 2n\varepsilon$. By a similar argument, if $h(\underline{\alpha},\underline{\beta})$ is defined and $1 - \varepsilon \leq \alpha_i \leq 1$ for all $i \in I$, then $h(\underline{\alpha},\underline{\beta}) < 2n\varepsilon$.

If the lemma is wrong, there exist $\underline{\alpha}$ and $\underline{\beta}$ such that $h(\underline{\alpha},\underline{\beta}) \neq 0$. By definition, $h(\underline{\alpha},\underline{\beta})$ cannot be negative; hence $h(\underline{\alpha},\underline{\beta}) > 0$. Choose $\varepsilon > 0$ such that $4n\varepsilon < h(\underline{\alpha},\underline{\beta})$, and let $\underline{\alpha}'$ and $\underline{\beta}'$ satisfy $\alpha_i' = \beta_j' = 1 - \varepsilon$ for all $i \in I$ and $j \notin I$. The argument of the previous paragraph gives $h(\underline{\alpha},\underline{\beta}') < 2n\varepsilon$ and $h(\underline{\alpha}',\underline{\beta}) < 2n\varepsilon$. Since $h(\underline{\alpha}',\underline{\beta}') \geq 0$, equation (7) now implies $h(\underline{\alpha},\underline{\beta}) < 4n\varepsilon$. This contradiction proves Lemma A12. ||

<u>Lemma A13</u>: If $\underline{u} \in D(x,I,y)$, then $f(\underline{u})$ is a lottery in x and y.

<u>Proof</u>: Assume $\underline{u} \in D(x,I,y)$. Then $T(u_i) = \{x\}$ for all $i \in I$. For $j \notin I$, $u_j(y) > u_j(x) > u_j(z)$ for all $z \in B$. A utility profile $\underline{v}$ shall be constructed as follows: For $i \in I$, find numbers c_i and d_i such that $c_i u_i(x) + d_i = 1$ and $c_i u_i(z) + d_i < 0$ for all $z \in B$. Since $u_i(x) > u_i(z)$ for all $z \in B$, such numbers exist, and c_i is positive. The utility scale v_i shall satisfy $v_i(x) = 1$, $v_i(y) = c_i u_i(y) + d_i$, and $v_i(z) = 0$ for $z \in B$. For $j \notin I$, numbers c_j and d_j shall be chosen such that $c_j u_j(y) + d_j = 1$ $> c_j u_j(x) + d_j > 0 > c_j u_j(z) + d_j$ for all $z \in B$. Since $u_j(y) > u_j(x) > u_j(z)$ for all $z \in B$, such numbers exist, and c_j is positive. The utility scale v_j is given by $v_j(y) = 1$, $v_j(x) = c_j u_j(x) + d_j$, and $v_j(z) = 0$ for $z \in B$.

The profile $\underline{v}$ satisfies the premise of Lemma A12; therefore, $\hat{B}\cdot f(\underline{v}) = 0$. Define $\underline{v}^{(0)} = \underline{v}$ and $\underline{v}^{(k)} = u_k \underline{v}_{-k}^{(k-1)}$ for $k = 1,\ldots,n$. Assume $\hat{B}\cdot f(\underline{u}) > 0$, and find the smallest k such that $\hat{B}\cdot f(\underline{v}^{(k)}) > 0$. Such a k exists, since $\underline{v}^{(n)} = \underline{u}$, and k is positive, since $\underline{v}^{(0)} = \underline{v}$. Let $\rho = f(\underline{v}^{(k-1)})$ and $\sigma = f(\underline{v}^{(k)})$; then $\hat{B}\cdot\rho = 0$ and $\hat{B}\cdot\sigma > 0$ by the choice of k. Strategy proofness implies $u_k\cdot\sigma \geq u_k\cdot\rho$ and $v_k\cdot\sigma \leq v_k\cdot\rho$. Since σ and ρ both are lotteries, $\hat{A}\cdot\sigma = \hat{A}\cdot\rho = 1$. These statements can be put together to give $(v_k - c_k u_k - d_k\hat{A})\cdot(\sigma-\rho) \leq 0$. (The fact $c_k > 0$ is used here.) The vector $v_k - c_k u_k - d_k\hat{A}$ has value 0 in the directions of x and y, and it has strictly positive value in the direction of z for all $z \in B$. For all these z, $\rho(z) = 0$ and $\sigma(z) - \rho(z) \geq 0$. Hence the inner product $(v_k - c_k u_k - d_k\hat{A})\cdot(\sigma-\rho)$ is a sum of non-negative terms. In order for the product to be non-positive, each term must be 0. This, in turn, implies $\sigma(z) - \rho(z) = 0$ for all z. Hence $\hat{B}\cdot\rho = \hat{B}\cdot\sigma$, contradicting statements made above. Therefore, the assumption $\hat{B}\cdot f(\underline{u}) > 0$ must be wrong, and the conclusion of Lemma A13 follows. ||

Lemmas A3 - A9 can now be applied, and the proof of Theorem 1* is complete.

A.3. Proof of Corollary 1'

Corollary 1': Assume that the set A of alternatives is finite and has at least three elements, while the set N of individuals is finite and nonempty. Let g be a straightforward game form with strategy sets $S_1,\ldots,S_n$. Assume that for each utility profile $\underline{u}$ there exists a strategy profile $\underline{s}$ such that $\underline{s}$ fits $\underline{u}$ and $g(\underline{s})$ is ex post Pareto optimal for $\underline{u}$. Then there exist nonnegative numbers $a_1,\ldots,a_n$ with sum 1 and game forms $g_1,\ldots,g_n$ defined on the same strategy sets as g, such that:

(i) For all $\underline{s}$, $g(\underline{s}) = a_1g_1(\underline{s}) + \ldots + a_ng_n(\underline{s})$.

(ii) For all i with $a_i > 0$, $P(g_i(\underline{s})) \subseteq T(u)$ whenever s_i is u-dominant for i in g.

Suppose that the premise holds. The proof will use Corollary 1. The premise of Corollaries 1 and 1' is the same, and the conclusion of Corollary 1 is the following: There exist nonnegative numbers $a_1,\ldots,a_n$ with sum 1 such that, for any $\underline{u}$ and any $\underline{s}$ which fits $\underline{u}$, there exist lotteries $\rho_1,\ldots,\rho_n$ with $g(\underline{s}) = a_1\rho_1 + \ldots + a_n\rho_n$, and $P(\rho_i) \subseteq T(u_i)$ for all i. Therefore, the numbers $a_1,\ldots,a_n$ are already specified.

For any i and any $s_i \in S_i$, a set $T_i(s_i)$ of alternatives is defined in this way:

$x \in T_i(s_i)$ if and only if $x \in T(u)$ for all u such that s_i is u-dominant for i in g.

Part (ii) of the conclusion is equivalent to the following statement:

For all i with $a_i > 0$ and all $\underline{s}$, $P(g_i(\underline{s})) \subseteq T_i(s_i)$.

If there exists no u such that s_i is u-dominant for i, then $T_i(s_i)$ is trivially equal to A. In this case, (ii) places no restriction on $g_i(\underline{s})$. Otherwise, it is not clear that $T_i(s_i)$ is non-empty. When $a_i > 0$, however, later arguments will prove $T_i(s_i) \neq \phi$.

The next step is to prove a lemma which corresponds closely to Lemma A8.

Lemma A14: Let $\underline{s}$ be a strategy profile and let I be a subset of N. Define $B \subseteq A$ by $B = \bigcup_{i\varepsilon I} T_i(s_i)$. Then $\hat{B}\cdot g(\underline{s}) \geq \sum_{i\varepsilon I} a_i$.

Proof: If B = A, then $\hat{B}\cdot g(\underline{s}) = 1$ and the conclusion is obviously true. This includes the possibility that $T_i(s_i) = A$ for some $i \varepsilon I$.

Suppose that $\underline{s}$, I and B provide a counterexample to the lemma, and define $a(I) = \sum_{i\varepsilon I} a_i$, $b = a(I) - \hat{B}\cdot g(\underline{s})$, and $C = A \setminus B$. Let ε be a positive number such that $2n\varepsilon < b$, choose an alternative $y \varepsilon C$, and let the utility scale u satisfy $u(x) = 0$ for $x \varepsilon B$, $u(y) = 1$, and $u(z) = 1 - \varepsilon$ for $z \varepsilon C \setminus \{y\}$. Construct a strategy profile $\underline{s}'$ in this way: For $i \varepsilon I$, $s_i' = s_i$. For $j \notin I$, s_j' is any strategy which is u-dominant for j; such a strategy exists, since g is straightforward. Let $j \notin I$, and consider the effect of substituting s_j' for s_j in any strategy profile. By an argument similar to the proof of Lemma A2, it is possible to prove that this change cannot increase the total probability of B by as much as 2ε. (Intuitively, this is obvious: High probability of B is bad according to u. If a switch from s_j to s_j' significantly increases this probability, individual j would be better off using s_j, contradicting the assumption that s_j' is u-dominant for j.) Since $\underline{s}'$ can be obtained from $\underline{s}$ by at most n such substitutions, $\hat{B}\cdot g(\underline{s}') < \hat{B}\cdot g(\underline{s}) + 2n\varepsilon < \hat{B}\cdot g(\underline{s}) + b = a(I)$, which implies $\hat{C}\cdot g(\underline{s}') > 1 - a(I)$.

Choose an alternative $z \varepsilon C$ in the following way: If $\hat{y}\cdot g(\underline{s}') > 1 - a(I)$, let z = y. Otherwise, find some $z \varepsilon C$ with $z \neq y$ and $\hat{z}\cdot g(\underline{s}') > 0$. Such a z must exist, since $\hat{C}\cdot g(\underline{s}') > 1 - a(I)$. Define b' by $b' = \hat{z}\cdot g(\underline{s}') - (1 - a(I))$

if $z = y$, and $b' = \hat{z} \cdot g(\underline{s}')$ if $z \neq y$. Then b' is positive, and a positive number ε' with $n\varepsilon' < b'$ can be found.

For each $i \in I$, $T_i(s_i)$ and C are disjoint sets, and $z \notin T_i(s_i)$. By the definition of $T_i(s_i)$, there exists a utility scale u_i such that s_i is u_i-dominant for i and $z \notin T(u_i)$. Find such a u_i, choose an alternative $x_i \in T(u_i)$, and define $c_i = u_i(x_i) - u_i(z)$. Then c_i is positive. Define v_i by $v_i(x_i) = u_i(x_i) + c_i\varepsilon'$, $v_i(z) = u_i(z) + c_i$, and $v_i(w) = u_i(w)$ for $w \notin \{x_i, z\}$, and let t_i be v_i-dominant for i. In order to discuss the effect of substituting t_i for s_i in some strategy profile, let $\underline{s}''$ be a profile with $s_i'' = s_i$, and let $\eta = g(\underline{s}'') - g(t_i\underline{s}''_{-i})$. Then $u_i \cdot \eta \geq 0$ since s_i is u_i-dominant, and $v_i \cdot \eta \leq 0$ since t_i is v_i-dominant. Since $v_i - u_i = c_i\varepsilon'\hat{x}_i + c_i\hat{z}$, this gives $c_i\varepsilon'\hat{x}_i \cdot \eta + c_i\hat{z} \cdot \eta \leq 0$. The facts $c_i > 0$ and $|\hat{x}_i \cdot \eta| \leq 1$ then imply $\hat{z} \cdot \eta \leq \varepsilon'$. That is, the substitution of t_i for s_i cannot reduce the probability of z by more than ε'.

Complete the strategy profile $\underline{t}$ by defining $t_j = s_j'$ for $j \notin I$. Since $s_i' = s_i$ for $i \in I$, the profile $\underline{t}$ can be obtained from $\underline{s}'$ by at most n substitutions of the type considered in the last paragraph. The result obtained there implies $\hat{z} \cdot g(\underline{s}') - \hat{z} \cdot g(\underline{t}) < n\varepsilon' < b'$. Together with the definition of b', this gives $\hat{z} \cdot g(\underline{t}) > 1 - a(I)$ if $z = y$, and $\hat{z} \cdot g(\underline{t}) > 0$ if $z \neq y$.

Let $v_j = u$ for $j \notin I$. This defines a utility profile $\underline{v}$, since v_i for $i \in I$ has already been defined. The profile $\underline{v}$ is strict; $T(v_i) = \{x_i\}$ for $i \in I$ and $T(v_j) = \{y\}$ for $j \notin I$. Moreover, $\underline{t}$ fits $\underline{v}$. Corollary 1 implies $g(\underline{t}) = \sum_{i \in I} a_i\hat{x}_i + (1 - a(I))\hat{y}$. Each x_i is different from z. Hence this contradicts the conclusion of the previous paragraph, both if $z = y$ and if $z \neq y$. This completes the proof of Lemma A14. $\|$

The proof of Lemma A14 does not assume that the sets $T_i(s_i)$ are non-empty. When $a_i > 0$, however, the lemma implies $T_i(s_i) \neq \phi$ for all $s_i \in S_i$. This is easily seen by choosing $I = \{i\}$. If $a_i = 0$, $T_i(s_i)$ may be empty. For example, if $g(\underline{s})$ does not depend on s_i, then every u satisfies the condition that s_i is u-dominant for i, and $T_i(s_i)$ is obviously empty. But this is of no consequence, since (ii) does not apply to such an i.

The conclusion of Corollary 1' requires no special relationship between $g_i(\underline{s})$ and $g_i(\underline{t})$ for $\underline{s} \neq \underline{t}$, apart from the conditions imposed on $g_i(\underline{s})$ and $g_i(\underline{t})$ separately. Therefore, the lotteries $g_1(\underline{s}),\ldots,g_n(\underline{s})$ can be constructed separately for each strategy profile $\underline{s}$. This can be done by applying the linear programming argument of Lemma A9, with s_i and $T_i(s_i)$ playing the roles of u_i and $T(u_i)$. When the proof of Lemma A9 invokes Lemma A8, Lemma A14 can be used instead. The linear programming argument constructs the lottery $g_i(\underline{s})$ for all i with $a_i > 0$. If $a_i = 0$, any lottery can be chosen as $g_i(\underline{s})$. Conditions (i) and (ii) will hold, and the proof of Corollary 1' is complete.

APPENDIX B

DIRECT PROOF OF THEOREM 2

Theorem 2: Assume that the set A of alternatives is finite and has at least three elements, while the set N of individuals is finite and non-empty. Let f be a decision scheme. If f is strategy proof and satisfies the ex ante Pareto condition, then f is dictatorial.

The proof consists of three lemmas. Throughout the proof, it is assumed that f satisfies the premise of the theorem. The notation introduced in Section 2 still applies, and a few more definitions are given here.

When I is a subset of N and x is an element of A, I is said to be decisive for x if $f(\underline{u}) = \hat{x}$ whenever $T(u_i) = \{x\}$ for all $i \in I$. The set I is decisive if it is decisive for all $x \in A$.

Let x and y be distinct alternatives, and let $I \subseteq N$. A utility profile $\underline{u}$ is said to be (x,I,y)-polarized if $u_i(x) > u_i(z) > u_i(y)$ and $u_j(y) > u_j(z) > u_j(x)$ for all $i \in I$, $j \notin I$, and all $z \notin \{x,y\}$. Clearly, $\underline{u}$ is (x,I,y)-polarized if and only if $\underline{u}$ is $(y,\tilde{I},x)$-polarized.

Lemma B1: For any pair of distinct alternatives x and y, and any set $I \subseteq N$, there exists an (x,I,y)-polarized utility profile $\underline{u}$ such that either $f(\underline{u}) = \hat{x}$ or $f(\underline{u}) = \hat{y}$.

Proof: Let $x \neq y$ and I be given. If $I = N$ or $I = \phi$, the result follows by the Pareto condition. Otherwise, pick an alternative z different from x and y. This alternative is kept fixed in the proof of this lemma. A set S of utility profile is defined as follows: The profile $\underline{u}$ is an element of S if and only if $u_i(x) = 1 > u_i(z) > 0.1 > u_i(w) > 0 = u_i(y)$ and $u_j(y) = 1 > u_j(z) > 0.1 > u_j(w) > 0 = u_j(x)$, for all $i \in I$, $j \notin I$, and all $w \notin \{x,y,z\}$. It is clear that $\underline{u} \in S$ implies that $\underline{u}$ is (x,I,y)-polarized. For $\underline{u} \in S$, define $c^-(\underline{u}) = \min_{i\in I} u_i(z)$, $c^+(\underline{u}) = \max_{i\in I} u_i(z)$, $d^-(\underline{u}) = \min_{j\notin I} u_j(z)$ and $d^+(\underline{u}) = \max_{j\notin I} u_j(z)$. Finally, two subsets L and U of S are defined by $\underline{u} \in L$ if and only if $\underline{u} \in S$ and $c^+(\underline{u}) + d^+(\underline{u}) < 1$, and $\underline{u} \in U$ if and only if $\underline{u} \in S$ and $c^-(\underline{u}) + d^-(\underline{u}) > 1$. Clearly, L and U are non-empty.

Let $\underline{u} \in L$ and $\rho = f(\underline{u})$. If $\rho(w) > 0$ for some $w \notin \{x,y,z\}$, construct a lottery σ by $\sigma = \rho - \rho(w)\cdot\hat{w} + \rho(w)\cdot\hat{z}$. For all i, $u_i(z) > u_i(w)$, which implies $u_i\cdot\sigma > u_i\cdot\rho$. This contradicts the ex ante Pareto condition. If $\rho(z) > 0$, let b satisfy $c^+(\underline{u}) < b < 1 - d^+(\underline{u})$; such a number exists, since $\underline{u} \in L$. If σ is given by $\sigma = \rho - \rho(z)\cdot\hat{z} + b\rho(z)\cdot\hat{x} + (1-b)\rho(z)\cdot\hat{y}$, it is easy to prove that $u_i\cdot\sigma > u_i\cdot\rho$ for all i. The Pareto condition is again contradicted. Hence $f(\underline{u})$ is a lottery in x and y.

Then let two utility profiles $\underline{u} \in L$ and $\underline{u}' \in L$ be given. Construct a profile $\underline{v} \in S$ such that $v_i(z) \leq \min(u_i(z), u_i'(z))$ for all i. Let $\underline{v}^{(0)} = \underline{v}$ and $\underline{v}^{(i)} = u_i\underline{v}_{-i}^{(i-1)}$ for $i = 1,\ldots,n$, which gives $\underline{v}^{(n)} = \underline{u}$. For every i, $c^+(\underline{v}^{(i)}) \leq c^+(\underline{u})$ and $d^+(\underline{v}^{(i)}) \leq d^+(\underline{u})$; hence $\underline{v}^{(i)} \in L$. The argument of the previous paragraph proves that $f(\underline{v}^{(i-1)})$ and $f(\underline{v}^{(i)})$ are lotteries in x and y, for all i.

Let $\eta = f(\underline{v}^{(i)}) - f(\underline{v}^{(i-1)})$. If $\eta \neq 0$, then $u_i\cdot\eta$ and $v_i\cdot\eta$ are either both strictly positive or both strictly negative; the sign depends on whether $i \in I$ and whether $\hat{x}\cdot\eta > 0$. In any case, i can manipulate f at

$\underline{v}^{(i-1)}$ or $\underline{v}^{(i)}$. Hence $\eta = 0$, and since this holds for all i, $f(\underline{u}) = f(\underline{v})$ follows. A similar argument gives $f(\underline{u}') = f(\underline{v})$. The profiles $\underline{u}$ and $\underline{u}'$ were arbitrary elements of L; hence f has the same value on all utility profiles from L. Let this value be ρ_L. Then ρ_L is a lottery in x and y. If it is equal to $\hat{x}$ or $\hat{y}$, the proof is complete. Assume, therefore, that there exists a number α with $0 < \alpha < 1$ and $\rho_L = \alpha\hat{x} + (1-\alpha)\hat{y}$.

Then suppose that $\underline{u} \in U$, and let $\rho = f(\underline{u})$. If $\rho(w) > 0$ for some $w \notin \{x,y,z\}$, a contradiction can be derived as above. If $\rho(x) > 0$ and $\rho(y) > 0$, find numbers b and ε such that $c^-(\underline{u}) > b > 1 - d^-(\underline{u})$ and $0 < \varepsilon$, $b\varepsilon \leq \rho(x)$ and $(1-b)\varepsilon \leq \rho(y)$, and construct σ by $\sigma = \rho + \varepsilon\hat{z} - b\varepsilon\hat{x} - (1-b)\varepsilon\hat{y}$. Such numbers b and ε can be found, since $\rho(x)$ and $\rho(y)$ are positive and $\underline{u} \in U$. It is easy to see that $u_i \cdot \sigma > u_i \cdot \rho$ for all i, and the Pareto condition is contradicted. The conclusion is that ρ is either a lottery in x and z, or a lottery in y and z. For any $\underline{u} \in U$ and $i \in I$, u_i induces the following ordering on the lotteries of this type: Lotteries in x and z are preferred to those in y and z. Among the former, a higher probability of x is preferred; among the latter, a higher probability of z is better. For $j \notin I$, u_j induces the exact opposite ordering on this set of lotteries.

If $\underline{v}$ and $\underline{v}'$ belong to U and only differ on individual i's utility scale, then $f(\underline{v}) = f(\underline{v}')$; otherwise, v_i and v_i' would induce the same strict preference between $f(\underline{v})$ and $f(\underline{v}')$, implying that i could manipulate f at $\underline{v}$ or $\underline{v}'$. For any two profiles $\underline{u}$ and $\underline{u}'$ in U, there exists a profile $\underline{v} \in U$ such that $v_i(z) \geq \max(u_i(z), u_i'(z))$ for all $i \in N$. It is possible to go from $\underline{u}$ to $\underline{v}$ to $\underline{u}'$, changing one individual's utility scale at a time and all the time staying inside U. None of these changes can alter the value of f. Hence $f(\underline{u}) = f(\underline{u}')$, that is, f has the same value for all utility profiles in U.

Let this value be ρ_U. If ρ_U is $\hat{x}$ or $\hat{y}$, the lemma is proved. Otherwise, there exist a number β with $0 \leq \beta < 1$ such that either $\rho_U = \beta\hat{x} + (1-\beta)\hat{z}$ or $\rho_U = \beta\hat{y} + (1-\beta)\hat{z}$. The situation is symmetrical, and there is no loss of generality in assuming $\rho_U = \beta\hat{x} + (1-\beta)\hat{z}$.

Let b be a number which satisfies $0.1 < b < 0.9$ and $b \neq (\alpha-\beta)/(1-\beta)$. Construct a utility scale u such that $u(x) = 1$, $u(y) = 0$, $u(z) = b$ and $0 < u(w) < 0.1$ for $w \notin \{x,y,z\}$. Then $u\cdot\rho_L = \alpha$ and $u\cdot\rho_U = \beta + (1-\beta)b$, which implies $u\cdot\rho_L \neq u\cdot\rho_U$. Define $b' = u\cdot\rho_L - u\cdot\rho_U$, and assume $b' > 0$. (The case $b' < 0$ is treated similarly.) Choose a positive number ε such that $n\varepsilon < b'$, $0.1 < b - \varepsilon$ and $b + \varepsilon < 1$. Then find utility profiles $\underline{u}$ and $\underline{v}$ in S such that $u_i(z) = b - \varepsilon$, $v_i(z) = b + \varepsilon$ and $u_i(w) = v_i(w) = u(w)$ for $i \in I$ and all $w \neq z$, while $u_j(z) = 1 - b$ and $v_j = u_j$ for $j \notin I$. Then $\underline{u} \in L$ and $\underline{v} \in U$. As before, define $\underline{v}^{(0)} = \underline{v}$ and $\underline{v}^{(i)} = u_i\underline{v}_{-i}^{(i-1)}$ for $i = 1,\ldots,n$, which gives $\underline{v}^{(n)} = \underline{u}$. For a fixed $i \in I$, let $\eta = f(\underline{v}^{(i)}) - f(\underline{v}^{(i-1)})$. By strategy proofness, $v_i\cdot\eta \leq 0$. Any component of η is less than 1 in absolute value, and $v_i - u = \varepsilon\hat{z}$. Hence $(u-v_i)\cdot\eta \leq \varepsilon$, which gives $u\cdot\eta \leq \varepsilon$. For $i \notin I$, $f(\underline{v}^{(i)}) = f(\underline{v}^{(i-1)})$. This implies $u\cdot(f(\underline{u}) - f(\underline{v})) \leq n\varepsilon < b'$. Since $f(\underline{u}) = \rho_L$ and $f(\underline{v}) = \rho_U$, this contradicts the definition of b' and concludes the proof of Lemma B1. ||

Lemma B2: Let I be a subset of N. Then either I or $\tilde{I}$ is decisive.

Proof: The set N is decisive by the Pareto condition, which proves the lemma for $I = N$ and $I = \phi$. If I is non-empty and different from N, let x and y be distinct elements of A, and let $\underline{u}$ be an (x,I,y)-polarized utility profile such that $f(\underline{u}) = \hat{x}$ or $f(\underline{u}) = \hat{y}$. The existence of $\underline{u}$ is guaranteed by Lemma B1, and $\underline{u}$ is kept fixed in the proof of this lemma. Moreover, assume that $f(\underline{u}) = \hat{x}$; then the conclusion will be that I is decisive. If $f(\underline{u}) = \hat{y}$, a similar argument can be used to prove that $\tilde{I}$ is decisive.

The first step is to prove that I is decisive for x. Let $\underline{v}$ satisfy $T(v_i) = \{x\}$ for all $i \in I$, and define $\underline{v}^{(0)}, \underline{v}^{(1)}, \ldots, \underline{v}^{(n)}$ as before. If $f(\underline{v}) \neq \hat{x}$, there must exist some $i \in N$ such that $f(\underline{v}^{(i-1)}) \neq \hat{x}$ and $f(\underline{v}^{(i)}) = \hat{x}$. If $i \in I$, this gives $v_i \cdot f(\underline{v}^{(i-1)}) < v_i \cdot f(\underline{v}^{(i)})$; hence i can manipulate f at $\underline{v}^{(i-1)}$. If $i \notin I$, $u_i \cdot f(\underline{v}^{(i-1)}) > u_i \cdot f(\underline{v}^{(i)})$, because x is the unique element with lowest utility in u_i; therefore, i can manipulate f at $\underline{v}^{(i)}$. In any case, strategy proofness is contradicted. This contradiction proves $f(\underline{v}) = \hat{x}$. Hence I is decisive for x.

Then let z and w be distinct alternatives, both different from x. By Lemma B1, there exists a (z,I,w)-polarized utility profile $\underline{u}'$ such that $f(\underline{u}') = \hat{z}$ or $f(\underline{u}') = \hat{w}$. In the latter case, an argument similar to the one in the last paragraph can be used to prove that $\tilde{I}$ is decisive for w. When $x \neq w$, it is obviously impossible that I is decisive for x and $\tilde{I}$ is decisive for w. Hence $f(\underline{u}') = \hat{z}$, and again the argument of the last paragraph can be used to conclude that I is decisive for z. Since z was an arbitrary alternative different from x, I is decisive, and Lemma B2 is proved. ||

<u>Lemma B3</u>: There exists an $i \in N$ such that $\{i\}$ is decisive.

<u>Proof</u>: Let m be the smallest positive integer such that there exists a decisive set I_0 with m elements. Since N is decisive, this is well defined.

If $m = 1$, the proof is complete. Then assume $m > 1$. Let I and J be non-empty disjoint subsets of I_0 such that $I \cup J = I_0$, and let $K = \tilde{I}_0$. Each of I and J has fewer than m elements and is not decisive by the definition of m. Since I_0 is decisive, its complement K cannot be decisive. Moreover, K is non-empty; otherwise, $J = \tilde{I}$ and either I or J would have to be decisive by Lemma B2. The union of any two of I, J and K is a decisive set. For $I \cup J$, this holds by assumption. The set $I \cup K$ is the complement of J, which is not decisive, and the conclusion follows from Lemma B2. A similar argument applies to $J \cup K$.

Let x, y and z be three distinct alternatives, and choose three utility scales u, u' and u'' such that $u(x) = u'(y) = u''(z) = 1$, $u(y) = u'(z) = u''(x) = 0.9$, $u(z) = u'(x) = u''(y) = 0$, and $u(w) < 0$, $u'(w) < 0$, and $u''(w) < 0$ for all $w \notin \{x,y,z\}$. Construct a utility profile $\underline{u}$ such that $u_i = u$, $u_j = u'$, and $u_k = u''$ for all $i \in I$, $j \in J$ and $k \in K$. By the Pareto condition, $f(\underline{u})$ is a lottery in x, y and z. Then it is easy to see that $(u + u' + u'')\cdot f(\underline{u}) = 1.9$; hence at least one of the numbers $u\cdot f(\underline{u})$, $u'\cdot f(\underline{u})$ and $u''\cdot f(\underline{u})$ must be less than 0.7. The situation is symmetrical, and there is no loss of generality in assuming $u\cdot f(\underline{u}) < 0.7$. Let $\underline{v}$ be a utility profile satisfying $T(v_i) = \{y\}$ for $i \in I$ and $v_j = u_j$ for $j \in J \cup K$. Then $T(v_i) = \{y\}$ for all $i \in I \cup J$. Since $I \cup J$ is decisive, this implies $f(\underline{v}) = \hat{y}$ and $u\cdot f(\underline{v}) = 0.9$. Let $\underline{v}^{(0)}, \underline{v}^{(1)}, \ldots, \underline{v}^{(n)}$ be defined as before. Then there must exist some i for which $u\cdot f(\underline{v}^{(i-1)}) > u\cdot f(\underline{v}^{(i)})$. For $i \notin I$, this is impossible, since $\underline{v}^{(i-1)} = \underline{v}^{(i)}$. For $i \in I$, it contradicts strategy proofness, since $u_i = u$. This contradiction completes the proof of Lemma B3. ||

The proof of Theorem 2 can now be completed. Let $i \in N$ be such that $\{i\}$ is a decisive set. If f is not dictatorial for i, there exists a utility profile $\underline{u}$ and an alternative x such that $\hat{x} \cdot f(\underline{u}) > 0$ and $x \notin T(u_i)$. Choose any alternative $y \in T(u_i)$, let $b = u_i(y)$, and construct a utility scale v_i such that $T(v_i) = \{y\}$. By the definition of decisiveness, $f(v_i \underline{u}_{-i}) = \hat{y}$. Hence $u_i \cdot f(v_i \underline{u}_{-i}) = b$, while $u_i \cdot f(\underline{u}) < b$. This contradicts strategy proofness. Hence f is dictatorial for i, and the proof of Theorem 2 is complete.

APPENDIX C

INFINITE SETS OF ALTERNATIVES OR INDIVIDUALS

In Theorems 1, 1* and 2, it is assumed that the sets A and N are finite. The consequences of removing these assumptions are discussed in this appendix.

C.1. Infinite A and Finite N

First assume that A is infinite and N is finite. As will be demonstrated below, the theorems hold in this case.

The definitions given in Section 2 and Appendix A still apply, but a few additional comments are in order. A lottery ρ is a function from A into the set of real numbers, such that $\rho(x) \geq 0$ for all $x \in A$ and $\sum_{x \in A} \rho(x) = 1$. That is, a lottery is a discrete probability distribution on A. The set $P(\rho)$ is finite or countable; otherwise, $\sum_{x \in A} \rho(x)$ would diverge and could not be equal to 1. For any number $\varepsilon > 0$, there exists a finite set $B \subseteq A$ such that $\hat{B} \cdot \rho \geq 1 - \varepsilon$.

On the infinite set A, it is possible to construct probability distributions which are not of the type described above. Formally, such a distribution is a measure on A. In particular, there exist atomless probability measures, that is, distributions which give probability 0 to any single element of A. These possibilities will not be considered here. If f is a decision scheme and $\underline{u}$ is a utility profile, then $f(\underline{u})$ shall be a lottery as defined above.

As before, a utility scale is a real-valued function defined on A. The purpose of introducing utility scales, is to represent individual preferences over lotteries. Certain rationality conditions are imposed on an individual's preferences, and it is proved that the preferences can be represented by a utility scale. The same rationality conditions also imply that the utility scale is bounded; see Arrow [1 , page 69]. To illustrate this, suppose that an individual's preferences are represented by a scale u which is unbounded from above. (Similar remarks apply if u is unbounded from below.) Choose an infinite sequence of alternatives $x_1, x_2, \ldots$, such that $u(x_m) \geq 2^m$ for $m = 1,2,\ldots$, and find an alternative y with $u(y) > u(x_1)$. Let ρ and σ be lotteries with $\rho(x_1) = \sigma(y) = 1/2$ and $\rho(x_m) = \sigma(x_m) = 1/2^m$ for $m \geq 2$. The rationality conditions imply that σ is preferred to ρ; this is intuitively clear, since σ can be obtained from ρ by making an alternative with probability 0.5 strictly better, while leaving everything else unchanged. But the expected utility, according to u, is infinite for both ρ and σ, contradicting the assumption that u represents the preferences. The possibility remains, however, that u represents the preferences on lotteries with finite utility. Another rationality assumption, namely the condition that preferences be continuous, rules this out: It is possible to find infinite sequences $\rho^{(1)}, \rho^{(2)}, \ldots$ and $\sigma^{(1)}, \sigma^{(2)}, \ldots$ of lotteries, such that $\lim_{m\to\infty} \rho^{(m)} = \rho$, $\lim_{m\to\infty} \sigma^{(m)} = \sigma$, and $u \cdot \rho^{(m)}$ and $u \cdot \sigma^{(m)}$ are finite for all m. Since σ is preferred to ρ, the continuity assumption implies that there exists an integer M such that $\sigma^{(m)}$ is preferred to $\rho^{(m')}$ whenever $m, m' \geq M$. But there exist numbers $m, m' \geq M$ such that $u \cdot \sigma^{(m)} < u \cdot \rho^{(m')}$.

The conclusion is that unbounded utility scales do not make sense. They contradict the rationality assumptions imposed on the preferences. If these assumptions are not imposed, preferences cannot in general be represented by utility scales, and the whole analysis breaks down. From now on, it is assumed that all utility scales are bounded.

When A is infinite, there exist utility profiles $\underline{u}$ such that $PO(\underline{u}) = \phi$. In such a case, any lottery is also dominated by some other lottery. (The condition that u_i is bounded is not sufficient to rule out this possibility.) On such profiles, no decision scheme can satisfy the premise of Theorem 1 and Theorem 2, and the theorems will be true in a trivial sense. (In Theorem 1*, it is not obvious that the premise cannot be satisfied, but the discussion below will show that this theorem is also trivially true if such profiles are included.) To make the problem an interesting one, profiles $\underline{u}$ with $PO(\underline{u}) = \phi$ must be excluded from consideration. Actually, a somewhat stronger restriction will be imposed. The utility scale u is said to be normal if u is bounded and $T(u) \neq \phi$; the utility profile $\underline{u}$ is normal if u_i is normal for all i. It is assumed that the decision scheme f is defined only on normal profiles. (It is easy to see that $T(u_i) \subseteq PO(\underline{u})$ for all $\underline{u}$ and i; hence assuming that $\underline{u}$ is normal is actually a stronger restriction than $PO(\underline{u}) \neq \phi$, provided that $n > 1$.) The possibility of defining f on a larger domain is discussed below.

Expressions like $u \cdot \rho$, $\hat{B} \cdot \rho$, etc., where u is a utility scale, ρ is a lottery, and B is a subset of A, are now defined by infinite sums. All these sums will converge; indeed, they are all absolutely convergent. (The sum $\sum_{x \varepsilon A} c_x$ is absolutely convergent if $\sum_{x \varepsilon A} |c_x|$ converges.) This can be proved by using the definition of a lottery and the fact that u and $\hat{B}$

are bounded functions. All rules of algebra can be applied, as if the sums were finite.

The proof of Theorem 1, given in Appendix A, will now be considered. The proof cannot be applied without change; the assumption that A is finite is used in a couple of places. For example, the condition is used in the following way in the proof of Lemma A5: A utility scale u_j satisfies $T(u_j) = \{y\}$. In order to construct the utility scale v_j and proceed with the proof, it is necessary to find a number which is less than $u_j(y)$ and greater than $u_j(w)$ for all $w \neq y$. When A is infinite, such numbers need not exist, even if u_j is normal. A similar problem occurs in the proof of Lemma A8: The number c_i may be 0, in which case the argument breaks down. The proof of these lemmas can, however, be applied if all utility scales have a "gap" between the utility of the top-ranked alternative(s) and the utility of all other alternatives.

This idea is formalized by the following definition: A utility scale u is <u>separated</u> if u is normal and there exists a number $\gamma > 0$ such that $u(x) \geq u(y) + \gamma$ for any $x \in T(u)$ and all $y \notin T(u)$. A utility profile $\underline{u}$ is separated if u_i is separated for all i.

The definition of the set D(x,I,y) will be changed slightly. A profile $\underline{u}$ belongs to D(x,I,y) if and only if $T(u_i) = \{x\}$ and $T(u_j) = \{y\}$ for all $i \in I$ and $j \notin I$, and there exist numbers $\gamma_i > 0$ for all $i \in N$ such that $u_i(x) \geq u_i(z) + \gamma_i$ for all $i \in N$ and all $z \notin \{x,y\}$. The definition implies that $\underline{u}$ is separated, but it is a little stronger: For $j \notin I$, there must also be a "gap" below the second-ranked alternative x in u_j. (This refinement is needed only for the proof of Theorem 1*; for the proof of Theorem 1, D(x,I,y) could have been defined as the set of separated profiles satisfying the definition in Appendix A.)

The proof of Lemmas A1 and A2 does not depend on the finiteness condition. Lemmas A3 and A4 hold for the set $D(x,I,y)$ as defined above; the profile $\underline{v}$ in the proof of Lemma A3 must be chosen such that, for all $i \in N$, there is a gap below the second-ranked alternative in the scale v_i. Then the numbers $a(x,I,y)$ and $a(I)$ can be defined. Lemma A5 holds for separated profiles. In the proof, when an alternative is moved upwards to become the second-ranked alternative in a certain scale, it is necessary that this alternative has a gap both above and below it in the scale which is being constructed. (This is possible, since the original scale is separated by assumption.) Then all statements of the form $\underline{v} \in D(x,I,y)$ will still be correct, with the new and more restrictive definition of $D(x,I,y)$. Lemmas A6 - A8 are true for all separated utility profiles; the earlier proof can be applied directly. (This implies that Lemma A7 holds without qualifications, since it is a statement about the numbers a_i and $a(I)$ and not about utility profiles. Its proof uses only separated profiles.)

Now consider Lemma A8 as it is stated in Appendix A. It is no longer required that profiles be separated. (The assumption that all profiles are normal is still in effect.) As noted above, the proof of Lemmas A1 and A2 does not depend on profiles being separated; therefore, these lemmas can be used. Suppose that Lemma A8 is wrong, and let $\underline{u}$ and I provide a counterexample. Define B as in the lemma, and let $b = a(I) - \hat{B} \cdot f(\underline{u})$; then b is positive by assumption. Choose a positive number ε such that $4n\varepsilon < b$. Let $C = A \setminus B$, and let C' be a finite subset of C such that $\hat{C}' \cdot f(\underline{u}) \geq \hat{C} \cdot f(\underline{u}) - b/2$; such a set can always be found, since there exists a finite set whose total probability in $f(\underline{u})$ is at least $1 - b/2$. Moreover, let $B' = A \setminus C'$; then $B \subseteq B'$, and $a(I) - \hat{B}' \cdot f(\underline{u}) \geq b/2$.

Then a utility profile $\underline{v}$ shall be constructed. For $i \in I$, v_i is constructed as in the proof of Lemma A8, with B' and C' substituted for B and C. Since C' is finite and C' and $T(u_i)$ are disjoint, c_i is strictly positive. When v_i is substituted for u_i in any utility profile, the total probability of the elements of B' cannot increase by more than ε. The proof of this claim exactly parallels the argument used in the proof of Lemma A8. Let x be a fixed element of C'. For $j \notin I$, v_j shall satisfy $v_j(x) = 1$, $v_j(y) = 1 - \varepsilon$ for all $y \in C' \setminus \{x\}$, and $v_j(z) = 0$ for all $z \in B'$. By Lemma A2, a change from u_j to v_j cannot increase the probability of B' by as much as 2ε.

The statements in the last paragraph imply $\hat{B}' \cdot f(\underline{v}) \leq \hat{B}' \cdot f(\underline{u}) + 2n\varepsilon < \hat{B}' \cdot f(\underline{u}) + b/2 \leq a(I)$. The profile $\underline{v}$ is separated. Let $B'' = \bigcup_{i \in I} T(v_i)$. Then $B'' \subseteq B \subseteq B'$. Lemma A8 can be applied to $\underline{v}$, which gives $\hat{B}'' \cdot f(\underline{v}) \geq a(I)$. Hence $\hat{B}' \cdot f(\underline{v}) \geq a(I)$, contradicting an earlier statement. This contradiction completes the proof of the unrestricted Lemma A8.

Lemma A9 remains. Let $\underline{u}$ be a normal utility profile, and define $\rho = f(\underline{u})$. As in Appendix A, it is necessary to find lotteries $f_i(\underline{u})$ for $i = 1,\ldots,n$, such that $\rho = \sum_{i \in N} a_i f_i(\underline{u})$ and $\hat{x} \cdot f_i(\underline{u}) = 0$ when $x \notin T(u_i)$. Two alternatives x and y are said to be <u>equivalent</u> if they belong to $T(u_i)$ for the same individuals i, that is, if $\{i | x \in T(u_i)\} = \{i | y \in T(u_i)\}$. This is obviously an equivalence relation, and it divides A into a finite number of non-empty equivalence classes. (There are at most 2^n classes.) Let χ be a typical equivalence class, and define $\rho(\chi) = \sum_{x \in \chi} \rho(x)$. If B is the set defined in Lemma A8 for some $I \subseteq N$ and if χ is an equivalence class, then either $\chi \subseteq B$ or $\chi \cap B = \phi$. Hence B can be regarded as a collection of equivalence classes rather than as a set of alternatives, and Lemma A8 will hold in this interpretation.

In the linear programming problems used in the proof of Lemma A9, the following changes shall be made: Equivalence classes are substituted for alternatives. Hence, for example, the primal problem shall have a variable $b_{\chi i}$ for every equivalence class χ and every $i \in N$. This guarantees that the number of variables and constraints is finite. The conditions $x \in T(u_i)$ and $x \notin T(u_i)$ shall be replaced by $\chi \subseteq T(u_i)$ and $\chi \cap T(u_i) = \phi$, respectively; it is obvious that exactly one of these conditions holds for every χ and i. As in Appendix A, Lemma A8 can be used to prove that the optimal value of the dual objective is 0. The primal problem has a feasible solution, and the duality theorem implies that there exists a feasible solution to the primal problem for which the objective has value 0. Let $b^*_{\chi i}$ be such a solution. The lottery $f_i(\underline{u})$ can now be defined by

$$\hat{x} \cdot f_i(\underline{u}) = \begin{cases} 0 & \text{if } \rho(\chi) = 0 \\ [\rho(x)/\rho(\chi)] \cdot [b^*_{\chi i}/a_i] & \text{if } \rho(\chi) > 0 , \end{cases}$$

where χ is the equivalence class to which x belongs. (This assumes $a_i > 0$; if $a_i = 0$, $f_i(\underline{u})$ can be any lottery on $T(u_i)$.) It is now easy to prove that $f_i(\underline{u})$ satisfies the conditions. This completes the proof of Theorem 1 for the case of infinite A.

Concerning Theorem 1*, the proof of Lemmas A10 - A13 does not depend on A being finite. In Lemma A13, the new definition of D(x,I,y) is essential for the existence of number c_i and d_i (c_j and d_j) with the required properties. Theorem 1* follows from these lemmas and the proof of Theorem 1.

Is it possible to relax the condition that utility profiles be normal? This is the issue to be discussed next. The logic of the argument is the following: Let f satisfy the premise of Theorem 1 or Theorem 1* on the set of normal profiles. The argument above shows that f, on this domain, is a probability mixture of dictatorial schemes. The question is: Can f be defined on a larger domain and still satisfy the premise of the theorem? For Theorem 1, the enlarged domain can certainly not contain any profile $\underline{u}$ with $PO(\underline{u}) = \phi$; for Theorem 1*, there is no equally obvious limitation on the domain.

To be specific, consider the case that individual 1 has preferences represented by the utility scale u which is not normal. That is, the scale u is bounded and $T(u) = \phi$. If $a_1 = 0$, f may be independent of individual 1's utility scale, in which case it is trivially possible to extent the domain of definition. But this is of no consequence, since f is a probability mixture of $f_2,\dots,f_n$, and f_1 can be ignored.

Then assume $a_1 > 0$. In this case, the theorem will be wrong if the domain can be extended, since no decision scheme can be dictatorial for 1 on the extended domain. Let x be a fixed alternative, and choose a utility profile $\underline{u}$ such that $u_1 = u$ and $T(u_i) = \{x\}$ for $i \geq 2$. Define $\rho = f(\underline{u})$, $b = u(x)$ and $b' = \sup_{y \in A} u(y)$. By assumption, b' is well defined and finite, and $b < b'$. Assume $\rho(x) \geq 1 - a_1$. Then $u_1 \cdot \rho < a_1 b' + (1 - a_1)b$. (Note that strict inequality holds even if $\rho(x) = 1 - a_1$. Then there exists a $y \neq x$ with $\rho(y) > 0$, and $u_1 \cdot \rho \leq a_1 b' + (1-a_1)b - \rho(y)(b' - u(y))$, and $\rho(y)(b' - u(y))$ is a positive number.) By the definition of b', there exists a $y \in A$ such that $u_1 \cdot \rho < a_1 u(y) + (1-a_1)u(x)$. If individual 1 reports a utility scale with y as the unique top-ranked alternative, earlier results imply that the outcome will be $a_1\hat{y} + (1-a_1)\hat{x}$. Hence person 1 can manipulate f at $\underline{u}$. Then assume $\rho(x) < 1 - a_1$, and let $\underline{v}$ be the utility profile which satisfies $v_1(x) = 0$, $v_1(y) = 1$ for all $y \neq x$, and $v_i = u_i$ for $i \geq 2$. The previous results can be applied to the normal profile $\underline{v}$, to conclude that $\hat{x} \cdot f(\underline{v}) = 1 - a_1$ and $v_1 \cdot f(\underline{v}) < v_1 \cdot f(\underline{u})$. That is, individual 1 can manipulate f at $\underline{v}$. This implies that $f(\underline{u})$ cannot be given any value without violating the premise of Theorem 1 (or Theorem 1*).

The conclusion can be summed up as follows: Suppose that the preferences of the different individuals are determined independently of each other; then the domain of definition of f must be a Cartesian product. Moreover, assume that this domain contains all normal profiles. If i is an individual with $a_i > 0$, it is not possible to define f on profiles whose ith coordinate is bounded but not normal, without violating the premise of Theorem 1. The same holds for Theorem 1*.

This completes the discussion of Theorems 1 and 1*. Concerning Theorem 2, it is easy to see that neither the argument in Section 3 that Theorem 1 implies Theorem 2 nor the direct proof in Appendix B depends on the assumption that A is finite. Hence Theorem 2 holds for infinite A, with f defined on the set of normal profiles. If strategy proofness is to be preserved, the scheme f cannot be defined on profiles in which the "dictator" has no top-ranked alternative.

C.2. Infinite N

If N is infinite, Theorems 1, 1*, and 2 are not true. It does not matter whether A is finite or infinite. Throughout most of this section, however, it is assumed that A is finite. The case of an infinite A is considered briefly at the end. Theorem 1* is not discussed separately; all counterexamples to Theorem 1 also contradict Theorem 1*.

The fact that the theorems fail for infinite N should come as no surprise. Similar results have been proved in the "traditional" social choice theory. In particular, Arrow's impossibility theorem does not hold when there are infinitely many voters; see Fishburn [9] and Kirman and Sondermann [17]. These authors use the mathematical concept of an ultrafilter to construct procedures based on a kind of "weighted majority vote" in which transitive majorities are guaranteed. When N is finite, this can only be achieved by essentially giving all weight to one individual; in the infinite case, however, there are other possibilities. The same principles can be used to construct counterexamples to Theorem 2; see comments below.

In the framework of this paper, there is another important difference between the cases of finite and infinite sets of individuals. The condition of strategy proofness loses most of its strength when N is infinite, because it is possible to construct decision schemes in which the outcome depends on the total system of individual preferences (which is a necessary condition for Pareto optimality), but at the same time no one individual can affect the outcome by a unilateral change of action. This is not possible when N is finite. Because of this weakening of the condition, counterexamples to Theorems 1 and 2 can be constructed without using advanced mathematical concepts.

When N is infinite, a utility profile $\underline{u}$ must be interpreted as a function from N into the set of utility scales; u_i is the value of this function at the element $i \varepsilon N$. (If N is countable, a utility profile can alternatively be thought of as an infinite sequence of utility scales.) The other definitions of Section 2 and Appendix A apply unchanged.

Theorem 1 is discussed first. For any utility profile $\underline{u}$, define a set T_1 by $x \in T_1$ if and only if $x \in T(u_i)$ for infinitely many individuals i. Since A is finite and N is infinite, T_1 is non-empty. Let $f^1(\underline{u})$ be the even-chance lottery over T_1. No single individual can bring about a change in T_1, therefore f^1 is a strategy-proof decision scheme. If $\hat{x} \cdot f^1(\underline{u}) > 0$, then $x \in T(u_i)$ for some i, which implies $x \in PO(\underline{u})$. That is, f^1 satisfies the ex post Pareto condition. Let I, J and K be disjoint, infinite subsets of N with $I \cup J \cup K = N$. Moreover, let x, y and z be distinct elements of A. Find utility profiles $\underline{u}$ and $\underline{v}$ such that $T(u_i) = T(v_i) = \{x\}$ for all $i \in I$, $T(u_j) = T(v_j) = \{y\}$ for all $j \in J$, and $T(u_k) = \{z\}$ and $T(v_k) = \{y\}$ for all $k \in K$. If f^1 satisfies the conclusion of Theorem 1, it must be a probability mixture of dictatorial schemes. In particular, $\hat{x} \cdot f^1(\underline{u})$ and $\hat{x} \cdot f^1(\underline{v})$ must both be equal to the total probability assigned to the set I. But $\hat{x} \cdot f^1(\underline{u}) = 1/3$ and $\hat{x} \cdot f^1(\underline{v}) = 1/2$; therefore, f^1 is a counterexample to Theorem 1. (The example does not depend on A having three elements. The profile $\underline{v}$ uses only two alternatives and proves that if f^1 satisfies the conclusion of Theorem 1, then I must have total probability 1/2. Similar profiles can be used to prove the same statement for J and K. But disjoint sets cannot have probabilities which sum to more than 1.)

The scheme f^1 does not satisfy the strong ex post Pareto condition. That is, there exist $\underline{u}$, x and y such that $\hat{x} \cdot f(\underline{u}) > 0$, $u_i(y) \geq u_i(x)$ for all i, and $u_i(y) > u_i(x)$ for some i. There seems to be no easy way of modifying f^1 so that it satisfies the strong condition; see, however, the discussion of f^3 below.

The ex ante Pareto condition is not satisfied by f^1; therefore, it does not contradict Theorem 2. In order to construct a counterexample to that theorem, let Q be a fixed linear ordering of A. For any $\underline{u}$, construct T_1 as above, let x be the first element of T_1 according to the ordering Q, and define $f^2(\underline{u}) = \hat{x}$. Again, strategy proofness is guaranteed, since no single individual can affect the outcome. If $f^2(\underline{u}) = \hat{x}$, there must exist an individual i with $x \in T(u_i)$. For this i, there exists no lottery ρ with $u_i \cdot \rho > u_i(x)$. Hence f^2 satisfies the ex ante Pareto condition. For an arbitrary individual i, let x and y be distinct alternatives, and let $\underline{u}$ satisfy $T(u_i) = \{x\}$ and $T(u_j) = \{y\}$ for all $j \neq i$. Then $T_1 = \{y\}$ and $f^2(\underline{u}) = \hat{y}$, which shows that f^2 is not dictatorial for i. Since i was arbitrary, f^2 is not dictatorial in the sense of this paper. This proves that f^2 is a counter-example to Theorem 2.

The decision scheme f^2 does not satisfy the strong (ex post or ex ante) Pareto condition. It is possible, however, to construct a scheme which satisfies this strong condition and contradicts Theorem 2. Let a fixed well-ordering of N be given. References below to "the first individual" with a certain property, will always refer to this well-ordering. For a given utility profile $\underline{u}$, let T_1 be defined as above. Define a relation R on T_1 by xRy if either $u_i(x) = u_i(y)$ for all i, or $u_i(x) > u_i(y)$ for the first individual i with $u_i(x) \neq u_i(y)$. A set T_2 shall consist of all R-maximal elements of T_1; that is, $x \in T_2$ if and only if $x \in T_1$ and xRy for all $y \in T_1$. It is easy to prove that R is transitive, reflexive and complete, and that T_2 is non-empty. Moreover, if x and y are elements of T_2, then $u_i(x) = u_i(y)$ for all i. Let $f^3(\underline{u})$ be the even-chance lottery over T_2. (Any lottery over T_2 could equally well have been chosen as the value of $f^3(\underline{u})$; for example, it is possible to choose $\hat{x}$ where x is the first element of T_2 according to Q.)

If f^3 is not strategy proof, there exist i, $\underline{u}$ and $\underline{v}$ such that $u_j = v_j$ for all $j \neq i$ and $u_i \cdot f^3(\underline{u}) < u_i \cdot f^3(\underline{v})$. There must exist x and y such that

$\hat{x}\cdot f^3(\underline{u}) > 0$, $\hat{y}\cdot f^3(\underline{v}) > 0$ and $u_i(x) < u_i(y)$. The set T_1 is the same whether it is constructed from $\underline{u}$ or $\underline{v}$, and this set must contain x and y. Moreover, $\hat{x}\cdot f^3(\underline{u}) > 0$ implies $x \in T_2$ and xRy (with R and T_2 constructed from $\underline{u}$). Since x and y do not have equal utility for all individuals, the first individual j for whom $u_j(x) \neq u_j(y)$ must have $u_j(x) > u_j(y)$. This j must come before i in the well-ordering. These statements imply that yRx is not true, and hence $y \notin T_2$, where R and T_2 are now defined from $\underline{v}$. This contradicts $\hat{y}\cdot f^3(\underline{v}) > 0$ and proves strategy proofness. If f^3 does not satisfy the strong ex ante Pareto condition, there exist $\underline{u}$, ρ and σ such that $f^3(\underline{u}) = \rho$, $u_i\cdot\sigma \geq u_i\cdot\rho$ for all i, and $u_i\cdot\sigma > u_i\cdot\rho$ for some i. Let x be any alternative with $\rho(x) > 0$; then $x \in T_2$. For any i, u_i is constant on T_2, hence $u_i\cdot\rho = u_i(x)$. If i satisfies $x \in T(u_i)$, then the assumption $u_i\cdot\sigma \geq u_i\cdot\rho$ implies $P(\sigma) \subseteq T(u_i)$. The construction of f^3 gives $x \in T_1$, and the definition of T_1 implies $P(\sigma) \subseteq T_1$. Let J be the set $\{j \in N | \text{there exists a } y \in P(\sigma) \text{ with } u_j(y) > u_j(x)\}$. If $u_i\cdot\sigma > u_i(x)$, then $i \in J$; this holds by assumption for at least one i, which proves that J is non-empty. Let j be the first element of J, and fix $y \in P(\sigma)$ with $u_j(y) > u_j(x)$. Since $y \in T_1$, $x \in T_2$ implies xRy, and there must exist an i such that i precedes j in the well-ordering, and $u_i(x) > u_i(y)$. By the definition of j, i is not in J. This implies $u_i(z) \leq u_i(x)$ for all $z \in P(\sigma)$. Since $\sigma(y) > 0$, this gives $u_i\cdot\sigma < u_i(x) = u_i\cdot\rho$, contrary to assumption. The strong ex ante Pareto condition is proved. The argument used to show that f^2 is not dictatorial can also be applied to f^3. This completes the proof that f^3 has the properties announced above.

In fact, f^3 is also a counterexample to Theorem 1. The premise is clearly satisfied. Let 1 be the first individual, and find infinite, disjoint subsets I and J of N such that $I \cup J = N \setminus \{1\}$. Consider a utility profile $\underline{u}$ which satisfies $T(u_1) = \{z\}$, $T(u_i) = \{x\}$ and $T(u_j) = \{y\}$ for $i \varepsilon I$ and $j \varepsilon J$, where x, y and z are distinct alternatives. If f^3 satisfies the conclusion of Theorem 1, the value of $f^3(\underline{u})$ can only depend on the probability assigned to individual 1 and the sets I and J. But $f^3(\underline{u}) = \hat{x}$ if $u_1(x) > u_1(y)$ and $f^3(\underline{u}) = \hat{y}$ if $u_1(x) < u_1(y)$.

Counterexamples to Theorem 2 can also be constructed by using the methods of Fishburn [9] and Kirman and Sondermann [17]. The construction is sketched here; for definitions and details, see the cited papers. A non-principal (or free) ultrafilter is given. For any profile $\underline{u}$, a set T_1^* is defined by $x \ \varepsilon \ T_1^*$ if and only if $\{i | x \ \varepsilon \ T(u_i)\}$ belongs to the ultrafilter. The set T_1^* will be non-empty (when A is finite, as is assumed here), and it will consist of the optimal elements in the social ordering constructed by Fishburn or Kirman and Sondermann. The decision scheme whose outcome is the even-chance lottery over T_1^* will be a counterexample to Theorem 2. Another example can be constructed from T_1^* in the same way as f^2 was constructed from T_1. The ex ante Pareto condition is satisfied in both cases, since there exist individuals i with $T_1^* \subseteq T(u_i)$. It is also possible to apply the technique which was used above to construct f^3, starting from T_1^* instead of T_1. The resulting scheme contradicts Theorem 2 and satisfies the strong ex ante Pareto condition. (The schemes discussed in this paragraph do not provide counterexamples to Theorem 1. An ultrafilter is equivalent to a probability measure; a set has measure 1 if it belongs to the ultrafilter and measure 0 otherwise. Therefore, all these schemes are probability mixtures of dictatorial schemes, that is, they essentially satisfy the conclusion of Theorem 1.)

This completes the discussion of the case where A is finite and N is infinite. Now assume that A and N are both infinite. Let Q be a fixed well-ordering of A, and let individual 1 be a designated member of N. Counterexamples to Theorems 1 and 2 will be constructed by modifying f^1 and f^2.

Only normal utility profiles will be considered; see discussion in Section C.1. For any such profile $\underline{u}$, define T_1 as above. If $T_1 = \phi$, let x be the first element of $T(u_1)$ according to the ordering Q, and let $\hat{x}$ be the outcome. If T_1 has one, two or three elements, let the outcome be the even-chance lottery over T_1. If T_1 is infinite, or if T_1 is finite and has four or more elements, assign probability 1/3 to each of the three elements of T_1 which come first in the ordering Q. No single individual can affect T_1. When $T_1 = \phi$, individual 1 can influence the outcome, but not in a way which contradicts strategy proofness. Otherwise, the proof that this decision scheme contradicts Theorem 1 is similar to the arguments used for f^1 above.

The scheme f^2 can be modified in the following way: If $T_1 \neq \phi$, let x be the first element of T_1. If $T_1 = \phi$, let x be the first element of $T(u_1)$. In both cases, let $\hat{x}$ be the outcome. (As above, "first" refers to the ordering Q.) This scheme is not dictatorial for individual 1 and not for any other person. It is easy to prove that it contradicts Theorem 2.

It is not possible to modify f^3 so that it can be applied when A is infinite. There exist normal utility profiles in which no alternative or lottery satisfies the strong (ex post or ex ante) Pareto condition. Therefore, if a scheme similar to f^3 should be constructed, the domain of definition would have to be further restricted.