DISCUSSION PAPER NO. 224

SOCIAL DECISION, STRATEGIC BEHAVIOR, AND BEST OUTCOMES: AN IMPOSSIBILITY RESULT

by

Allan Gibbard */

*/ Department of Philosophy, University of Pittsburgh

June 1976

ABSTRACT: A modified version of Arrow's theorem shows that if people act strategically, then no system of group decision-making can guarantee a best feasible outcome in all circumstances. The assumptions are these: (1) What is best is determined by an ordering which is independent of considerations of feasibility. The way this ordering is determined satisfies two minimal conditions: Unanimity and No Weak Dictator. (2) When people act strategically, the outcome is independent of preferences involving non-feasible alternatives and of positive linear transformations of their utility scales. (3) There are at least four alternatives.

1. Introduction

What ought to happen depends at least in part on what the people involved prefer. I shall take that as a truism, though it may need qualifications: perhaps what ought to happen depends not on what people actually prefer, but on what they would prefer if they were fully informed and clearheaded, and perhaps it depends not on what people prefer on the whole, but on what each person prefers as regards himself. Accepting these qualifications, though, would only make the problem I raise in this paper more difficult. I shall assume here that what ought to happen depends at least partly on the preferences the people involved actually have. If the reader thinks that only informed preferences matter, he can think of the paper as addressing the special case where actual preferences are fully informed, and if he thinks that only self-regarding preferences matter, he can think of the paper as tackling the special case where everyone's preferences are self-regarding.

My question is how, at least in these special cases, to design a system that will ensure that what ought to happen always does. The systems I consider may be constitutions, they may be ways for a well-meaning government to base its decisions on plebiscites or public opinion polls, or they may be systems of economic incentives. I shall talk at times as if I were discussing only one of these cases, but what I say should apply to all systems through which individuals interact to produce an outcome.

2. Strategic Behavior

Suppose a well-meaning government tries to base its decisions on the preferences of the people involved. It may well not be able to learn what those preferences are. If it asks people their preferences — through plebiscites or public opinion polls — people who understand the system by which decisions are made may have an incentive to misreport their preferences.

Suppose, for instance, the government selects a policy on the basis of a Borda rule, as follows. Each person lists the alternative policies on his ballot in order of preference. Then, where there are m alternatives, each person's first choice gets $m-1$ points, his second choice gets $m-2$ points, and in general, his ι th choice gets $m-\iota$ points (so that his last choice gets zero). The points each alternative gets are added up, and the alternative with the most points wins; ties are broken by chance. Now suppose three people vote among alternatives w, x, y, and z: person i votes ordering xyzw, person j votes wxyz, and the true preference ordering of person k is wxyz. Simple arithmetic shows that if k votes his true preferences, the score is w 6, x 7, y 5, and z 1, so that x wins. k, though, can scuttle x by voting the ordering wyzx; in that case the score is w 6, x 5, y 5, and z 2, so that k's first choice wins. k thus gains by misreporting his preferences.

People who act so as to secure the result they like best will be said to behave strategically. This characterization of strategic behavior needs some refinement: a strategic agent will ordinarily not know for sure how others will act, and so he acts in a way that in some sense holds out the best prospect, on the basis of his limited information, of advancing his interests. Exactly how he does this will not matter for what I have to say in this paper. In particular, a person who votes not in order honestly to reveal his preferences, but in order most effectively to advance his interests, will be said to vote strategically. To say this is not to say that he misreports his preferences, but that however he votes, truthfully or otherwise, he does so because no other way of voting holds out a better prospect of advancing his interests.

Strategic voting is not inevitable. It might be that each voter wants the system to work properly to produce an ethically best result. If he trusts the others to vote honestly, he may himself vote honestly because otherwise, he reasons, he would subvert the way the system works to produce an ethically best result.

In many communities, though, strategic behavior will be impossible or costly to prevent. This paper takes up a problem for such communities: whether a satisfactory system of group decision-making can be designed for them — whether, that is, a system can be designed that will give the results it ought even when everyone behaves strategically.

3. Strategy-Proofness

One way to approach the question I have asked is to inquire whether there are reasonable systems of group decision-making which, by the very way they are designed, are guaranteed never to reward an individual's misreporting his preferences. Such a system will be called strategy-proof. If a government uses a strategy-proof system to make its decisions and people do not misreport their preferences when they have no incentive to do so, then the government will indeed base its decisions on peoples' genuine preferences. The remaining question is whether it will base its decisions on their preferences in a reasonable way.

The answer to this remaining question seems to be negative: all strategy-proof systems are defective as ways to make a social choice depend on individual preferences. Take first a strategy-proof system which picks the alternative to be put into effect without resort to any element of chance. Any such system, it turns out, will either be dictatorial or be duple, in the sense that the outcome is confined, independently of the way people vote, to a fixed pair of alternatives. (See Gibbard, 1973, and Satterthwaite, 1975.) Take next a strategy-proof system in which chance does play some role in determining the final outcome. It turns out that any such system, for almost all combinations of individual preferences, is a fixed probability mixture of systems, each of which is either duple or unilateral, in the sense that it

denies everyone but a single fixed voter any influence whatsoever on the outcome. If in addition, such a system guarantees that the alternative to be put into effect will be Pareto-optimal, then in the absence of individual indifference between alternatives, the system is a random dictatorship — a fixed probability mixture of dictatorial systems. (See Gibbard, 1977, 1976.) In brief, then, only a narrow class of unappealing systems can preclude advantageous individual misreporting of preferences.

I have implied that a random dictatorship is "unappealing", and I should say something about what is wrong with it. I shall assume without argument that a dictatorial system is unsatisfactory, and talk about random dictatorships that are not fully dictatorial — that are, I shall say, non-degenerately random. One defect of a non-degenerately random dictatorship is that the lotteries it produces may not be Pareto-optimal ex ante: there may be an alternative lottery which everyone prefers. (This problem is discussed by Zeckhauser, 1973, p. 939.) Another defect is more pertinent to the concerns of this paper: a non-degenerately random dictatorship may result in one alternative's being put into effect when another feasible alternative is better.

The assumptions behind this claim are crucial to the argument of the entire remainder of this paper. Some alternatives are feasible and some are infeasible, and for any set M of alternatives, we can ask what ought to be done in

the case where M is the set of feasible alternatives. I shall assume, here and in the rest of the paper, that what ought to be done it this: put into effect one of the best of the feasible alternatives, where a best feasible alternative is a feasible alternative such that no other feasible alternative is better than it. I shall assume three things about the relation is better than: first, that it is an ordering (with ties allowed); second, that if everyone prefers an alternative x to an alternative y, then x is better than y; and third, that what is better than what does not depend on which alternatives are feasible.

It follows from these assumptions that under some possible circumstances, a non-degenerately random dictatorship may have an outcome which is not a best feasible alternative. For suppose the contrary. Consider a society of two people, i and j, with three alternatives, x, y, and z, where i ranks the alternatives in order xyz and j ranks them zxy. Suppose first that all three alternatives are feasible. A lottery between x and z results, and so either x or z may be put into effect. x and z are therefore both best feasible alternatives, and therefore equally good. Everyone prefers x to y, and hence x is better than y. Since z and x are equally good and x is better than y, it follows that z is better than y. Now suppose only y and z are feasible. The non-degenerately random dictatorship yields a lottery between y and z, and so even though

z is better than y, either may be put into effect. That contradicts the supposition that a non-degenerately random dictatorship will always put into effect a best feasible alternative.

4. An Alternative Approach

On the basis of three assumptions about the relation _is better than_ and a theorem which characterizes strategy-proof systems, I have shown that no strategy-proof system can ensure that the alternative put into effect will always be a best feasible alternative. That leaves open the question of whether there could be a system which was not strategy-proof, but under which the effects of strategic manipulation were always benign.[1] Strategic misreporting might be benign in that it switched the outcome from one best feasible alternative to another, or from a non-best feasible alternative to a best one. The question I now want to ask, then, is this: could there be a system of voting that ensures that whatever peoples' preferences are, the outcome of their strategic voting is always a best feasible alternative?

The question can be broadened to include systems that are not systems of voting. By a system of _voting_, I have meant a system in which people somehow report their preferences, and a decision is based in some way on their reported preferences. Think now of systems in which people do not

necessarily report their preferences, but do take actions of some kind, and thereby interact to produce an outcome. Economic systems are prime examples of a system of interaction which does not consist of voting. What we can now ask is this: Is there any possible system of human interaction that will ensure that whatever peoples' preferences are, if they understand the system and act rationally through it to advance their interests, the outcome will always be a best feasible alternative?[2]

5. Systems of Interaction

One way to represent a system through which people interact is by what I shall call a "game form with variable feasibility", or "GFWVF". Let there be n players, a non-empty set L of alternatives, and for each player i, a finite non-empty set S_i of pure strategies for i. A GFWVF g for alternative set L and pure strategy sets $S_1, \ldots, S_n$ is defined as follows. A pure strategy profile $\underset{\sim}{s}$ for g is an n-tuple $\langle s_1, \ldots, s_n \rangle$ where $s_1 \in S_1$, $\ldots, s_n \in S_n$. g is a function whose domain consists of all pairs $\{M, \underset{\sim}{s}\}$ of a finite non-empty subset M of L (called the feasible set) and a pure strategy profile $\underset{\sim}{s}$, and whose value $g(M, \underset{\sim}{s})$, for any such M and $\underset{\sim}{s}$, is a lottery over members of M (that is, an assignment to the members of M of non-negative real numbers adding up to one). The lottery is to be interpreted as giving the probability each alternative

has of being put into effect when the set of feasible alternatives is M and players play the pure strategies given by $\underset{\sim}{s}$.

The theory of non-cooperative games can be thought of as the theory of what happens when rational agents interact strategically through a GFWVF. For each set M of feasible alternatives, a GFWVF g determines a <u>game form</u> — a function whose arguments are all pure strategy profiles and whose values are lotteries over a fixed set of feasible alternatives. (See Gibbard, 1973, 1976.) A combination of a game form and a utility scale for each player is a <u>game</u>, in the sense of standard game theory. Where g is a GFWVF, M a feasible set, $U_1, \ldots, U_n$ are utility scales, and $\underset{\sim}{U} = \langle U_1, \ldots, U_n \rangle$, we can designate the resulting game as $\langle g, M, \underset{\sim}{U} \rangle$. For any such game, non-cooperative game theory tells us (or aspires to tell us) what the players might do, and hence which of the alternatives might be put into effect as a result of the play of the game.

Given a GFWVF, then, which alternatives might be put into effect depends on the set of feasible alternatives and the utility scales of the players. I shall call the function that expresses this dependence a "social choice function" or "SCF". Given a GFWVF g, we can informally characterize the consequent SCF as the function c such that, for any $\underset{\sim}{U} = \langle U_1, \ldots, U_n \rangle$ and finite non-empty set M of alternatives, $c(M, \underset{\sim}{U})$ is the set of feasible alternatives which might be put into effect as a result of the play of the game $\langle g, M, \underset{\sim}{U} \rangle$.

In the rest of this paper, I shall talk not about GFWVF's, but about the corresponding SCF's. A social choice function is defined as follows. Let there be n people and a set L of alternatives. A utility scale U_i over L is a function that assigns a real number to each alternative in L, and an n-person utility profile over L is an n-tuple of utility scales over L. An n-person social choice function (or SCF) over L is a function c whose domain consists of all pairs $\langle M,\underset{\sim}{U}\rangle$ consisting of a finite non-empty subset M of L and an n-person utility profile $\underset{\sim}{U}$ over L, where for each such M and $\underset{\sim}{U}$, $c(M,\underset{\sim}{U})$ is a non-empty subset of M. The set $c(M,\underset{\sim}{U})$ is called the choice set of c for M and $\underset{\sim}{U}$.

Note the difference between the way a GFWVF is used to represent a system of interaction and the way a SCF is used to represent it. A GFWVF is to be interpreted as giving a lottery as a function of what people actually do. A SCF will be interpreted here as giving possible outcomes as a function of peoples' true utility scales; it expresses the end result when players who are guided by their true utilities interact strategically. As an illustration of the difference, take a system which consists of each person's writing down a utility (from one to a hundred) for each alternative, with the reported utilities being aggregated in some way to determine a social decision or lottery. The GFWVF that represents this system will show how the outcome depends on what people

report. Now if the system rewards strategic misreporting of $\hat{u}$tilities, what people report may well not be their true utilities. The SCF that represents the system will tell which alternatives might be put into effect as a function not of utilities as reported, but of peoples' true utilities.

The choice sets of an SCF may contain more than one alternative. There are a number of reasons for this. In the first place, the GFWVF that underlies the SCF may yield non-degenerate lotteries. In that case, the choice set of the resulting SCF will consist of all alternatives that get non-zero probability in a given situation. In the second place, even if a GFWVF always yields a single alternative with probability one for any pure strategy profile and feasible set, players may adopt mixed strategies, so that again, more than one alternative has a non-zero probability of being adopted. Finally, a GFWVF may yield games with multiple equilibria. For any feasible set M and utility profile $\underset{\sim}{U}$ that give multiple equilibria, the choice set $c(M,\underset{\sim}{U})$ will contain the outcomes of all equilibria.

Because SCF's will be used here to represent the results of strategic interactions, we can draw on game theory to place conditions on the SCF's we consider. In game theory, cardinal utility scales are significant only up to positive linear transformations. If for fixed $a > 0$ and b, $U'(x) = aU(x) + b$ for every alternative x, then which of the scales U or U' is ascribed to a player makes no

difference to the behavior to be expected of him. We shall consider, then, only SCF's that satisfy the following condition.

> Scale Invariance. Let $a_1, \ldots, a_n$ be positive real numbers, and let $b_1, \ldots, b_n$ be any real numbers. Suppose $\underset{\sim}{U}$ and $\underset{\sim}{U}'$ are such that for every person i and every alternative $x \in L$, $U_i'(x) = a_i U_i(x) + b_i$. Then for any M, $c(M,\underset{\sim}{U}') = c(M,\underset{\sim}{U})$.

In the second place, I shall assume here that which alternatives are feasible is common knowledge, in the sense that everyone knows it, everyone knows that everyone knows, and so forth. The utilities players ascribe to non-feasible alternatives, then, will have no bearing on their behavior. Even if the system permits voters to express preferences involving non-feasible alternatives, they will decide what preferences to express not on the basis of how much they like the various non-feasible alternatives, but on the basis of how much they like the various feasible alternatives and how they expect their expressions of preferences which involve non-feasible alternatives to affect the social choice among feasible alternatives.[3] The SCF's we consider, then, should satisfy this condition.

> Independence of Preferences Involving Non-Feasible Alternatives (IPINFA). Let M be a feasible set, and suppose $\underset{\sim}{U}$ and $\underset{\sim}{U}'$ agree on M, in the sense that for every person i and every $x \in M$, $U_i'(x) = U_i(x)$. Then $c(M,\underset{\sim}{U}') = c(M,\underset{\sim}{U})$.

This is Arrow's Independence of Irrelevant Alternatives in something close to its original form. (Arrow, 1963, p. 27)

Scale Invariance and IPINFA together have a special significance for choice from pairs of alternatives: they entail the following condition.

> Determination of Pairwise Choice by Pairwise Preferences (DPCPP). Let $\underset{\sim}{U}$, $\underset{\sim}{U}'$, x, and y be such that for all people i, $U_i'(x) > U_i'(y)$ iff $U_i(x) > U_i(y)$, and $U_i'(x) < U_i'(y)$ iff $U_i(x) < U_i(y)$. Then $c(\{x,y\},\underset{\sim}{U}') = c(\{x,y\},\underset{\sim}{U})$.

Lemma 1. Suppose SCF c satisfies Scale Invariance and IPINFA. Then c satisfies DPCPP.

Proof: For each i, transform scales U_i linearly to scales V_i so that if $U_i(x) > U_i(y)$ then $V_i(x) = 1$ and $V_i(y) = 0$; if $U_i(x) = U_i(y)$, then $V_i(x) = V_i(y) = 0$; and if $U_i(x) < U_i(y)$, then $V_i(x) = 0$ and $V_i(y) = 1$. Transform scales U_i' to V_i' in the same manner. Then for all i, $V_i'(x) = V_i(x)$ and $V_i'(y) = V_i(y)$, and so by IPINFA, $c(\{x,y\},\underset{\sim}{V}') = c(\{x,y\},\underset{\sim}{V})$. By Scale Invariance, $c(\{x,y\},\underset{\sim}{V}) = c(\{x,y\},\underset{\sim}{U})$ and $c(\{x,y\},\underset{\sim}{V}') = c(\{x,y\},\underset{\sim}{U}')$. Therefore $c(\{x,y\},\underset{\sim}{U}') = c(\{x,y\},\underset{\sim}{U})$, and the Lemma is proved.[4]

DPCPP says that in the case of a pairwise social choice, not only are the utilities of the feasible alternatives all that matter, but that strength of preference does not matter: all that matters is who prefers the one and who prefers the other.

This is the only condition on SCF's that will be exploited in what follows. The condition might have been justified directly on the basis of what a strategic player will do to influence a social choice between a pair of alternatives x and y. If he prefers x to y, no matter how weakly, he will do whatever holds out the best prospect of securing x as opposed to y. The same holds, mutatis mutandis, if he prefers y to x. In no case will his strength of preference make any difference to what he does.

6. Best Outcomes

Return now to the main question of this paper: whether there is any possible system of human interaction which will ensure that rational agents, acting strategically through the system, will always produce a best feasible outcome. I have argued that if it is common knowledge which alternatives are feasible and that all agents involved are rational, then strategic interaction through a system can be represented by a SCF which satisfies Scale Invariance and IPINFA, and thus satisfies DPCPP. For these conditions of common knowledge, then, the question of this paper can be put as a question about SCF's. Is there, we can ask, a SCF satisfying Scale Invariance and IPINFA, such that for any feasible set M and utility profile $\underset{\sim}{U}$, the choice set $c(M,\underset{\sim}{U})$ consists only of best feasible alternatives?

A modification of the Arrow theorem shows that the answer is negative. On the basis of a few weak assumptions about the relation is better than (all but one of which have already been made), I shall show that if there are at least four alternatives, then there is no SCF satisfying Scale Invariance and IPINFA which always confines the choice set to best feasible alternatives.

Consider again the relation is better than. Whether one alternative is better than another depends at least in part, I have supposed, on the preferences of the people involved. I shall represent this dependence by a social welfare function f, where for any utility profile $\underset{\sim}{U}$, $f(\underset{\sim}{U})$ is a binary relation between alternatives. To say that $\langle x,y\rangle \in f(\underset{\sim}{U})$ is to say that if peoples' preferences were as given by utility profile $\underset{\sim}{U}$, then x would be better than y.

Here I do not mean to suppose that what is better than what depends exclusively on peoples' utilities. Other factors may be relevant. If they are, assume them to be fixed in some way. f will then represent the way what is better than what depends on individual utilities when those other factors are held fixed in that way.

A variety of positions on the significance of individual utility scales will be compatible with what I shall be saying. For all I shall say, levels of utility may be interpersonally comparable or not, strength of preference may be interpersonally comparable or not, and the strengths of different

preferences of the same person may be comparable or not. Thus the ethical significance of a utility scale may differ from its behavioral significance. Interpersonal comparisons of utilities will not, I supposed earlier, bear on the choices people make, but they may, for all I am supposing, have a bearing on such ethical questions as which of two alternatives is better.

The formal definition of a social welfare function is this. As with a SCF, we begin with a number n of *people* and a non-empty set L of *alternatives*. An n-person utility profile over L is defined as before. An *n-person social welfare function* (or *SWF*) over L is a function f whose domain consists of all n-person utility profiles $\underset{\sim}{U}$ over L, and whose value $f(\underset{\sim}{U})$ for any such $\underset{\sim}{U}$ is a two-place relation on L.

On a SWF f, we can impose conditions that we take to characterize the dependence of the relation *is better than* on individual utilities. One such condition is built into the mechanism of a SWF: that whether one alternative is better than another does not depend on which alternatives are feasible. Two other conditions were discussed earlier, and need now only be formulated.

Ordering. For any $\underset{\sim}{U}$, where $P = f(\underset{\sim}{U})$, we have

$(\forall x,y)\sim[x\,P\,y\ \&\ y\,P\,x]$ (*full asymmetry*)

$(\forall x,y,z)[(\sim x\,P\,y\ \&\ \sim y\,P\,z) \longrightarrow \sim x\,P\,z]$ (*negative transitivity*).

Unanimity. For any $\underset{\sim}{U}$, x, and y, if for all people i, $U_i(x) > U_i(y)$, then where $P = f(\underset{\sim}{U})$, we have $x P y$.

The final condition to be imposed is that no one is so much more significant than everyone else that the social ordering could never go against his preferences.

No Weak Dictator. For every person i, there are a $\underset{\sim}{U}$, x, and y such that where $P = f(\underset{\sim}{U})$, we have both $U_i(x) > U_i(y)$ and $y P x$.

It remains only to impose a condition on the relation between a SCF and a SWF. The condition should say that the choice sets of the SCF consist only of best feasible alternatives, where what makes one alternative better than another is indicated by the SWF.[5]

Optimality. For any feasible set M and utility profile $\underset{\sim}{U}$,

$$c(M,\underset{\sim}{U}) \subseteq \{x \mid x \in M \;\&\; \sim(\exists y \in M)\langle y,x\rangle \in f(\underset{\sim}{U})\}.$$

Note that this condition requires more than Pareto optimality: it requires that the alternatives that might be put into effect be fully best, as determined by the standard given by f.

The conditions imposed here bear a close resemblance to the Arrow conditions. Indeed, if we take the conditions needed for a cardinal version of the Arrow theorem (Sen, 1970, p. 129), there are only two differences. One is that Arrow's non-dictatorship condition is weaker than the

condition of No Weak Dictator given here. The other, more crucial difference is that Arrow has a strengthened version of the condition of Optimality. Optimality here requires that all members of the choice set of the SCF be best feasible alternatives; Arrow requires in addition that all best feasible alternatives be included in the choice set. In other words, Arrow strengthens the Optimality condition by requiring equality rather than subsethood.

What distinguishes the approach here from the Arrow approach is this. Here constraints on a theory of what feasible alternatives are best are distinguished from constraints, either ethical or practical, on a system of group decision.[6] The relation <u>is better than</u>, it seems to me, should be an ordering determined at least in part by individual utilities and independent of considerations of feasibility. I can see no reason, though, for requiring the way it is determined to satisfy both Scale Invariance and IPINFA. These latter constraints, on the other hand, apply inevitably to most systems of human interaction. I can see no reason, though, for requiring the results of group choices between pairs of alternatives to yield an ordering. I agree that group decision ought to be constrained by considerations of which feasible alternatives are best, and that which feasible alternatives are best is determined by an ordering, but I see no reason for group decisions to be fully determined by the ethical consideration of which feasible alternatives are best. Hence I accept that all chosen alternatives should be optimal, but not that all optimal alternatives should stand a chance of being chosen.

7. The Impossibility Theorem

If there are at least four alternatives, it will now be shown, the conditions I have stated cannot be jointly met. I take the force of this theorem to be as follows: If it is common knowledge that people are rational, know which alternatives are feasible, and are disposed to act strategically, and if any utility scale whatsoever is possible for each of them, then no system of group decision-making (or interaction of any kind) will be ethically perfect by the standards I have proposed.

Theorem. Let a set L of alternatives have at least four members, let f be an n-person SWF over L, and let c be an n-person SCF over L. Then not all of the following hold: f satisfies Ordering, Unanimity, and No Weak Dictator, c satisfies IPINFA and Scale Invariance, and f and c are related by the condition of Optimality.

Proof:[7] Let a set I of people be weakly decisive for x over y if

$$(\exists \underset{\sim}{U})[((\forall i \in I)\ U_i(x) > U_i(y)) \ \&\ ((\forall j \notin I)\ U_j(y) > U_j(x)) \ \&\ x \in c(\{x,y\}, \underset{\sim}{U})].$$

This will be written $x\ D_I\ y$. I is strongly decisive for x over y if $x \neq y$ and

$$(\forall U)[((\forall i \in I)\ U_i(x) > U_i(y)) \longrightarrow x = c(\{x,y\}, \underset{\sim}{U})].$$

This will be written $x\ \bar{D}_I\ y$.

Lemma 2. If for some x and y, $x\ D_{\{i\}}\ y$, then for all x and y, $x\ \bar{D}_{\{i\}}\ y$.

Proof: Suppose $x\ D_{\{i\}}\ y$. Then by DPCPP, for any $\underset{\sim}{U}$ such that $U_i(x) > U_i(y)$ and $(\forall j \neq i)\ U_j(x) < U_j(y)$, we have $x \in c(\{x,y\},\underset{\sim}{U})$. Let $z \notin \{x,y\}$, and let $\underset{\sim}{U}$ be such that i orders alternatives x, y, and z in order $x\,y\,z$, and everyone else prefers y to both z and x. Then $x \in c(\{x,y\},\underset{\sim}{U})$. Let $P = f(\underset{\sim}{U})$, and let R be the corresponding loose preference relation (c.l.p.r.): the relation such that for all v and w, $v\,R\,w$ iff $\sim w\,P\,v$. By Optimality, we then have $x\,R\,y$. Everyone prefers y to z, and so by Unanimity, $y\,P\,z$. Therefore by RP-transitivity, $x\,P\,z$. Hence by Optimality, $x = c(\{x,z\},\underset{\sim}{U})$. Since the only assumption about the ordering of x with respect to z was that $U_i(x) > U_i(z)$, by DPCPP, $x = c(\{x,z\},\underset{\sim}{U})$ whenever i prefers x to z: in other words, $x\ \bar{D}_{\{i\}}\ z$.

Now let $w \neq z$. If $w = x$, then $w\ D_{\{i\}}\ z$. If $w \neq x$, let U be such that i ranks $w\,x\,z$ in that order, and everyone else prefers w to x. Let $P = f(\underset{\sim}{U})$ and R be the c.l.p.r. Then since $x\ \bar{D}_{\{i\}}\ z$, we have $x = c(\{x,z\},\underset{\sim}{U})$; by Optimality, $x\,R\,z$; by Unanimity, $w\,P\,x$; by PR-transitivity, $w\,P\,z$; by Optimality, $w = c(\{w,z\},\underset{\sim}{U})$; and hence by DPCPP, $w\ \bar{D}_{\{i\}}\ z$. This holds, then, for all $z \notin \{x,y\}$ and $w \neq z$.

The first argument now shows that for any $v \notin \{w,z\}$, $w\ \bar{D}_{\{i\}}\ v$. Since $w\ \bar{D}_{\{i\}}\ z$, we have $w\ \bar{D}_{\{i\}}\ v$ for all $v \neq w$. Here w is any alternative such that for some z, $w \neq z \notin \{x,y\}$, and since there are at least four alternatives, this is no restriction at all. That proves the Lemma.

Proof of Theorem: Let I be the minimal set such that for some x and y, $x\ D_I\ y$. Let w, x, y, and z be distinct, let $i \in I$, and let $\underset{\sim}{U}$ be such that i has the ranking $x\,y\,z\,w$, $I - \{i\}$ all have the ranking $z\,w\,x\,y$, and everyone else has the ranking $y\,z\,w\,x$. Let $P = f(\underset{\sim}{U})$, and let R be the c.l.p.r. Everyone in I prefers x to y, and so since $x\ D_I\ y$, by DPCPP, $x \in c(\{x,y\},\underset{\sim}{U})$. Hence by Optimality, $x\,R\,y$. Only those in $I - i$ prefer z to y, and so if $z \in c(\{y,z\},\underset{\sim}{U})$, then $I - \{i\}$ would be weakly decisive on a pair, contradicting the original characterization of I. Therefore $y = c(\{y,z\},\underset{\sim}{U})$, and by Optimality, $y\,R\,z$. Everyone prefers z to w, and so by Unanimity, $z\,P\,w$. Hence by RRP-transitivity, $x\,P\,w$. Since only i prefers x to w, we have $x\ D_{\{i\}}\ w$. Therefore by Lemma 2, $x\ \bar{D}_{\{i\}}\ y$ for all x and y.

It follows that i is a weak dictator for f. For for any $\underset{\sim}{U}$, x, and y, if $U_i(x) > U_i(y)$, then $x = c(\{x,y\},\underset{\sim}{U})$, and so by Optimality, $x\,R\,y$. That proves the Theorem.[8]

REFERENCES

Arrow, Kenneth, Social Choice and Individual Values (Second Edition), 1963.

Campbell, Donald E., "Democratic Preference Functions", Journal of Economic Theory 12, 259-72 (1976).

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

Gibbard, "Manipulation of Schemes That Mix Voting with Chance", forthcoming in Econometrica (1977).

Gibbard, "Straightforwardness of Game Forms with Lotteries as Outcomes", Discussion Paper No. 203, Center for Mathematical Studies in Economics and Management Science, Northwestern University, 1976.

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

Sen, Amartya K., Collective Choice and Social Welfare, 1970.

Zeckhauser, Richard, "Voting Systems, Honest Preferences, and Pareto Optimality", American Political Science Review 67, 934-46 (1973).

NOTES

1. This question has been suggested to me by Elaine Bennett, Mark Satterthwaite, Amartya Sen, and possibly other people.

2. The fundamental approach of this paper — studying whether strategic behavior might be benign by distinguishing a function that represents a system of group decision from a function that represents an ethical theory — was suggested to me independently by Elaine Bennett and Amartya Sen. It is the approach taken by Campbell (1976).

3. This argument is given by Campbell (1976, p. 264), and has been given on a number of occasions by Charles Plott.

4. Essentially the same argument is given by Sen (1970, pp. 129-30).

5. This condition is essentially Campbell's condition A_3 (1976, p. 264).

6. Campbell (1976, pp. 259, 264) makes this distinction.

7. The method of proof here follows, with appropriate modifications, that of the Arrow theorem (in Arrow, 1963, pp. 97-100).

8. I have been greatly helped in writing this paper by discussions with Elaine Bennett, Mark Satterthwaite, and Hugo Sonnenschein.