Manipulation of Voting Schemes: A General Result

Author(s): Allan Gibbard

Source: *Econometrica*, Jul., 1973, Vol. 41, No. 4 (Jul., 1973), pp. 587-601

Published by: The Econometric Society

Stable URL: https://www.jstor.org/stable/1914083

ECONOMETRICA

VOLUME 41 July, 1973 NUMBER 4

MANIPULATION OF VOTING SCHEMES: A GENERAL RESULT

BY ALLAN GIBBARD

It has been conjectured that no system of voting can preclude *strategic voting*—the securing by a voter of an outcome he prefers through misrepresentation of his preferences. In this paper, for all significant systems of voting in which chance plays no role, the conjecture is verified. To prove the conjecture, a more general theorem in game theory is proved: a *game form* is a game without utilities attached to outcomes; only a trivial game form, it is shown, can guarantee that whatever the utilities of the players may be, each player will have a dominant pure strategy.

1. INTRODUCTION

I SHALL PROVE in this paper that any non-dictatorial voting scheme with at least three possible outcomes is subject to individual manipulation. By a "voting scheme," I mean any scheme which makes a community's choice depend entirely on individuals' professed preferences among the alternatives. An individual "manipulates" the voting scheme if, by misrepresenting his preferences, he secures an outcome he prefers to the "honest" outcome—the choice the community would make if he expressed his true preferences.

The result on voting schemes follows from a theorem I shall prove which covers schemes of a more general kind. Let a *game form* be any scheme which makes an outcome depend on individual actions of some specified sort, which I shall call *strategies*. A voting scheme, then, is a game form in which a strategy is a profession of preferences, but many game forms are not voting schemes. Call a strategy *dominant* for someone if, whatever anyone else does, it achieves his goals at least as well as would any alternative strategy. Only trivial game forms, I shall show, ensure that each individual, no matter what his preferences are, will have available a dominant strategy. Hence in particular, no non-trivial voting scheme guarantees that honest expression of preferences is a dominant strategy. These results are spelled out and proved in Section 3.

The theorems in this paper should come as no surprise. It is well-known that many voting schemes in common use are subject to individual manipulation. Consider a "rank-order" voting scheme: each voter reports his preferences among the alternatives by ranking them on a ballot; first place on a ballot gives an alternative four votes, second place three, third place two, and fourth place one. The alternative with the greatest total number of votes wins. Here is a case in which an individual can manipulate the scheme. There are three voters and four alternatives; voter a ranks the alternatives in order $xyzw$ on his ballot; voter b in order $wxyz$; and voter c's true preference ordering is $wxyz$. If c votes honestly, then, the winner is his second choice, x, with ten points. If c pretends that x is his last choice by giving his preference ordering as $wyzx$, then x gets only eight points, and c's first choice, w, wins with nine points. Thus c does best to misrepresent his preferences.

Since many voting schemes in common use are known to be subject to manipulation, writers on the subject have conjectured, in effect, that all voting schemes are manipulable. Dummett and Farquharson define "voting procedure" roughly as "voting scheme" is defined here, and remark, "It seems unlikely that there is any voting procedure in which it can never be advantageous for any voter to vote 'strategically,' i.e., non-sincerely" [**3**, p. 34]. The definition of manipulability used here is roughly that originated by Dummett and Farquharson. The result they prove, however, applies only to a special class of voting schemes which they call "majority games," not to voting schemes in general.

Vickrey [**5**, p. 518] makes a related conjecture on manipulability. He conjectures that immunity to manipulation is equivalent to the conjunction of two of the conditions[1] that figure in the Arrow impossibility theorem [**1**]. Arrow's conditions are jointly inconsistent, and hence from Vickrey's conjecture, it would follow that a scheme satisfying the remaining Arrow conditions is manipulable—almost the result in this paper. Indeed the proof in this paper proceeds roughly by confirming Vickrey's conjecture.

A result such as the one given here, then, was to be expected. It does not, however, turn out to be easy to prove from known results: the proof given here relies on the Arrow impossibility theorem, but not in a simple way. I leave the statement and proof of the results in this paper until later; first, informal elucidation.

2. MANIPULABILITY IN THE WORLD

A way of making decisions can be represented by a variety of mathematical structures. In the next part, theorems are proved about structures of two kinds, called "game forms" and "voting schemes." In this part, I shall argue that the game form and the voting scheme that represent a decision-making system pick out the aspects of the system pertinent to its manipulability.

First, then, let us provide definitions of "voting scheme" and "game form." A voting scheme is a formula by which individual preferences among alternatives completely determine a community choice, or "outcome." A voting scheme, then, is a function of the following sort. Let there be n voters, and let Z be the set of alternatives open to society. Call an ordering of Z a *preference ordering*, and an n-tuple of preference orderings a *preference n-tuple*. ("Orderings" in this paper allow ties.) A preference ordering is thus an individual's account of his preferences among available alternatives, and a preference n-tuple consists of a profession of preferences from each individual. A *voting scheme* is a function which assigns a member of Z to each possible preference n-tuple for a given number n and set Z.

A voting scheme is a special case of what I shall call a "game form," and the theorem on voting schemes is a special case of a general result about game forms which I shall give. A game form, as I shall explain, is a system which allows each individual his choice among a set of *strategies*, and makes an *outcome* depend, in a determinate way, on the strategy each individual chooses. A "strategy" here is the same as a pure strategy in game theory, and indeed a game form is a game with

[1] *Independence of irrelevant alternatives* and *positive association*, which Vickrey calls *non-perversity*.

no individual utilities yet attached to the possible outcomes. Formally, then, a *game form* is a function g with a domain of the following sort. To each player 1 to n is assigned a non-empty set, $S_1, \ldots, S_n$ respectively, of *strategies.* It does not matter, for purposes of the definition, what a strategy is. The domain of the function g consists of all n-tuples $\langle s_1, \ldots, s_n \rangle$, where $s_1 \in S_1$, $s_2 \in S_2, \ldots, s_n \in S_n$. The values of the function g are called *outcomes.* A voting scheme, it follows, is a game form such that, for each player, his set of strategies is the set of all orderings of a set Z of available alternatives, where Z includes the set X of outcomes.

Voting schemes and game forms are mathematical structures used to represent flesh and blood systems of decision making which might be instituted. They represent decision-making systems which leave nothing to chance, but let the choice a community makes depend solely on what its members do. Other structures could be used for the same purpose, but here, voting schemes and game forms are especially apt: each, I shall argue, applies to a wide range of non-chance decision-making systems, and each picks out certain aspects of a system which pertain to manipulability.

Game forms apply to the widest range of decision-making systems, and apply to each in a clear-cut way. Every non-chance procedure by which individual choices of contingency plans for action determine an outcome is characterized by a game form. Game forms, then, characterize any non-chance procedure we would consider voting. In representing a system as a game form, each possible way of voting counts as a strategy. In single-ballot voting of any sort, for instance, a strategy would consist of a way of marking the ballot. In sequential voting, a strategy is a way of marking each ballot on the basis of what has gone before. For any such system, it is clear what constitutes a strategy, and what each combination of strategies has as its outcome. It is clear, then, what game form characterizes the system.

For game forms alone, however, there is no such thing as manipulation. To manipulate a system, a voter must misrepresent his preferences. Nothing in the structure of a game form tells us what strategy "honestly" represents any given preference ordering, and hence which strategies would misrepresent it. To talk of manipulation, then, we must specify not only a game form, but for each voter and preference ordering P, we must specify the strategy which "honestly represents" P. Only then can we apply the definition of manipulation as securing an outcome one prefers by selecting a strategy other than the one that honestly represents one's preferences.

Manipulability, then, is a property of a game form $g(s_1, \ldots, s_n)$ plus n functions $\sigma_1, \ldots, \sigma_n$, where for each individual k and preference ordering P, $\sigma_k(P)$ is the strategy for k which honestly represents P. Formally, then, where Z is the set of all alternatives open to the community, each σ_k is a function whose arguments are all orderings of Z and whose values are strategies open to k. Manipulability is a property of a game form in conjunction with an n-tuple $\langle \sigma_1, \ldots, \sigma_n \rangle$ of such functions.

Where a decision-making system is characterized by functions $g, \sigma_1, \ldots, \sigma_n$ as I have indicated, it is characterized by a voting scheme,

$$v(P_1, \ldots, P_n) = g(\sigma_1(P_1), \ldots, \sigma_n(P_n)).$$

For each n-tuple $\langle P_1, \ldots, P_n \rangle$, $v(P_1, \ldots, P_n)$ is the outcome if individuals $1, \ldots, n$ honestly profess preference orderings $P_1, \ldots, P_n$ respectively. Whereas manipulability is not a property of a game form alone, it is a property of a voting scheme alone. Voting scheme v is manipulable if for some k and preference n-tuples $\langle P_1, \ldots, P_n \rangle$ and $\langle P'_1, \ldots, P'_n \rangle$, $P_i = P'_i$ except when $i = k$, and

$$v(P'_1, \ldots, P'_n)\, P_k\, v(P_1, \ldots, P_n).$$

For, then, if P_k is k's real preference ordering, given the way the others vote, k prefers the result of expressing preference ordering P'_k to that of expressing P_k.

Note that to call a voting scheme manipulable is not to say that, given the actual circumstances, someone really is in a position to manipulate it. It is merely to say that, given some possible circumstances, someone could manipulate it. A voting scheme is manipulable, then, unless its structure guarantees that no matter how each person votes, no one will ever be in a position to manipulate the scheme.

Manipulability pertains to voting schemes, and in that sense, then, a voting scheme picks out the aspects of a decision-making system which pertain to manipulation. In some cases, however, it will not be clear what voting scheme characterizes a given decision-making system. It will be clear enough what game form characterizes it, but the voting scheme which characterizes it is derived from the game form by means of the functions $\sigma_1, \ldots, \sigma_n$. What are we to make of these functions? They characterize "honest" voting, I have said. That makes sense as long as for each individual k and preference ordering P, it is clear what strategy for k "honestly expresses" P. For many systems of voting, however, it is not always clear what constitutes honesty. A system may give no single clear way to express certain preference orderings.

Suppose, for instance, a club is to vote first on whether to have a party, and then, if the motion to have a party carries, on whether to make it alcoholic. What strategy would count as expressing the following preference ordering: a non-alcoholic party first, no party at all second, and an alcoholic party last? It is not at all clear. Hence, although it is clear what game form characterizes the system, it is not at all clear what voting scheme, if any, characterizes it. Manipulability is most clearly a property of voting schemes, but many real systems of voting are not clearly characterized by any one voting scheme.

In short, then, game forms are more versatile but manipulability pertains more directly to voting schemes. Any non-chance system of decision making is characterized by a game form in a clear-cut way, but manipulability is not a property of game forms alone. It is rather a property of a game form plus the functions $\sigma_1, \ldots, \sigma_n$ which characterize honest voting. Equivalently, it is a property of the voting scheme defined from the game form g and $\sigma_1, \ldots, \sigma_n$. Unless, however, the system prescribes for each preference ordering a way to express it, the choice of functions $\sigma_1, \ldots, \sigma_n$ to characterize honest expression of preferences will be to some degree arbitrary. Hence the choice of a voting scheme to characterize the system will be to some degree arbitrary. Game forms most clearly characterize decision-making systems, but manipulation pertains to voting schemes.

The moral, of course, is that unless a decision-making system prescribes clearly how each voter is honestly to express each possible preference ordering, manipulation of the system is an unclear notion. A voter manipulates the system if, by misrepresenting his preferences, he secures an outcome he prefers. Unless we have clear standards of honest representation and hence of misrepresentation, manipulation makes no clear sense.

Even so, we can prove a general result on the manipulability of decision-making systems, and do so either in terms of game forms or of voting schemes. What we can show is this: however we characterize honest voting in a system, the system as we characterize it will be manipulable. All non-trivial voting schemes are manipulable, so that no matter what voting scheme we choose to characterize a system, the system, as characterized, will be manipulable. Whatever functions $\sigma_1, \ldots, \sigma_n$ we choose to characterize honest voting, honesty will not always be the best policy. The only exceptions are trivial systems—dictatorial systems and systems with no more than two outcomes.

Here is the result put in terms of game forms. A strategy s^* is *dominant* for player k and preference ordering P of the set of outcomes if, for each fixed assignment of strategies to players other than k, strategy s^* for k produces an outcome at least as high in preference ordering P as does any other strategy open to k. For player 1, for instance, s^* is dominant for P if there is no strategy n-tuple $\langle s_1, \ldots, s_n \rangle$ such that

$$g(s_1, s_2, \ldots, s_n)\, P\, g(s^*, s_2, \ldots, s_n).$$

A game form is *straightforward* if for every player k and preference ordering P of the outcomes, some strategy is dominant for k and P. The theorem on game forms says that no non-trivial game form is straightforward.

From that the result put in terms of voting schemes follows. The argument is given in Section 3 in the proof of the corollary. A voting scheme, as I have said, is a game form of a special kind, in which the strategies are preference orderings of the alternatives. Now take a voting scheme, and take a voter k and preference ordering P for which no strategy is dominant. Then in particular, honest voting is not a dominant strategy for k and P. Thus if P is k's real preference ordering, then given some possible way the others might vote, k does best to misrepresent his preferences. The voting scheme is therefore manipulable. The result on manipulability, then, can be put in terms either of game forms or of voting schemes, and the result put in terms of voting schemes follows from the result put in terms of game forms.

Some further comments on voting schemes. They characterize a large variety of systems. A voting scheme need in no way be democratic, and it need not guarantee that all individuals count alike. Some voting schemes represent dictatorships, some oligarchies, and some democracies. Nor must a voting scheme treat all alternatives in the same way. A voting scheme might, for instance, allow Jones a special say on what groceries get delivered to him. It might also rule out duels even if everyone wanted one to be fought. Some voting schemes treat all alternatives alike; others do not.

A voting scheme must assign an outcome to every preference n-tuple, not just to some. Murakami [**4**, pp. 75–77] discusses group manipulability of structures which do not meet this condition, but as far as I can see, doing so makes no sense. However people vote, something will happen. If some preference n-tuples lead to stalemate and inaction, then inaction is a possible outcome. If someone prefers inaction to the outcome he would secure with honest voting, and he can secure inaction by misrepresenting his preferences, the system is manipulable. Stalemate must be counted as an outcome, and so in discussing manipulability, we should consider a function which assigns a value to every preference n-tuple and not just to some.

Finally, neither voting schemes nor game forms allow ties. Both take single outcomes as values, and for good reason. In questions of manipulability, the final outcome is what matters; manipulation, after all, is a way of securing a final outcome one prefers. Here we are considering decision-making systems in which chance plays no part, and to display manipulation of such a system, we need functions whose values are definite final outcomes.

In this respect, a voting scheme differs from an Arrow "constitution" [**2**], which it resembles in most other respects. Both a constitution and a voting scheme take preference n-tuples as arguments, but whereas to each preference n-tuple a voting scheme assigns a single alternative, a constitution assigns a choice function—a function which, for each non-empty set of alternatives, chooses a non-empty subset. Now this subset may have more than one member. Some constitutions, then, allow ties.

Voting schemes rule out ties, for in systems which leave nothing to chance, ties make no sense. A voter misrepresents his preferences in order to secure a decision he prefers, and in the end only one alternative is chosen. In a non-chance decision-making system, it does a voter no good to have an alternative he likes tie for winning place if some other alternative tied with it is actually chosen.[2] To investigate manipulability, we must consider the entire system by which the choice is made, including the system for breaking any ties which may develop. That means considering a system which results in a single choice. Voting schemes and game forms, then, suit the present purpose; some Arrow constitutions do not.

Suppose, though, a system breaks ties by chance. Game forms and voting schemes, I have said, characterize only non-chance decision-making systems. What can we say about the manipulability of systems which make an outcome depend partly on preferences but also partly on chance?

A system which broke ties by chance would not be a voting scheme. It would assign to each preference n-tuple not an outcome, but a lottery among possible outcomes. We might call it a "mixed decision scheme." Let a *prospect* be an assignment to alternatives of probabilities which total one. Then a *mixed decision scheme* is a function which assigns a prospect to each preference n-tuple.

Just as we can talk about the manipulability of a voting scheme, we can talk about the manipulability of a mixed decision scheme. Call a mixed decision scheme

[2] Vickrey [**5**, p. 508] makes roughly the same point.

manipulable unless it is the case that, whenever everyone expresses his preferences honestly in an election, the scheme assigns to that election a prospect each voter likes as well as any prospect he could have secured by misrepresenting his preferences, given the actual votes of everyone else. Whereas, with trivial exceptions, all voting schemes are manipulable, it is easy to find a mixed decision scheme which is not manipulable. Take the scheme which assigns to each alternative the fraction of voters for whom it is a first choice. In other words, each voter writes his first choice on a ballot; a single ballot is drawn at random; and the choice on that ballot is selected. A voter then has every incentive to give his true first choice. If his ballot is not drawn, it makes no difference how he votes, whereas if his ballot is drawn, the voter gets his true first choice if and only if he puts it as first choice on his ballot. His second and lower choices do not matter, for only the choice on the ballot can affect the outcome. Hence the system is not manipulable, and we have established the existence of a mixed decision scheme which is not manipulable, not dictatorial, and can allow a large number of possible outcomes.

That leaves the question of whether any non-manipulable mixed decision schemes are attractive. Exactly what is required for a scheme to be "attractive," I cannot specify. Clearly, though, the scheme I have presented is unattractive; it leaves too much to chance. On the other hand, a system which allowed only occasional ties to be broken by chance might be quite attractive. Work needs to be done on mixed decision schemes. It would be good to identify properties which would make a mixed decision scheme attractive, and to have theorems on the manipulability of classes of mixed decision schemes which, by various criteria, are attractive. No such work is attempted in this paper.

I have argued, then, that in discussing manipulability of systems which leave nothing to chance, we must consider functions whose values are single outcomes—voting schemes or game forms. For systems with an element of chance, we must consider functions whose values are prospects—mixed decision schemes. There exists a non-manipulable mixed decision scheme, but whether any non-manipulable mixed decision schemes are attractive in any way remains to be seen.

Back, then, to the topic of this paper: voting schemes which leave nothing to chance. Every voting scheme is dictatorial, limited to one or two possible outcomes, or subject to manipulation. Why should that matter? It means that no system of decision making but a trivial one can depend on informed self-interest to make outcomes a function of true preferences. If a system does make outcomes a function of preferences, it is in virtue of individual integrity, ignorance, or stupidity, or because preferences are sufficiently predictable that the system does not have to accommodate all possible patterns of preferences. For suppose a system accommodates all possible preference patterns, and makes outcomes a function of preferences. Then where v is that function, v is a voting scheme, and unless trivial, is manipulable. Hence for some k and $P_1, \ldots, P_n$, let $x = v(P_1, \ldots, P_n)$. Then for some y, $y P_k x$, but given the preferences of everyone else, for some P'_k, if k's preference ordering were P'_k, the outcome would be y. Thus if k acted as if his preference ordering were P'_k, the outcome would be y, which would be more to his liking than the outcome he actually secures. The way k actually acts, given his

preferences, is not the way which best promotes k's interests. The way k acts, then, must depend on something other than informed self-interest—perhaps ignorance, integrity, or stupidity. No straightforward appeal to informed self-interest can make the outcome a non-trivial function of preferences regardless of what those preferences are.

I have argued, then, that game forms and voting schemes are the best subjects for manipulability theorems on non-chance systems of decision-making, and that the theorems proved here have regrettable consequences.

3. MANIPULABILITY THEOREMS AND PROOFS

The results and proofs which follow are self-contained.

A game form is characterized by:

(i) A set X, whose members are called *possible outcomes*, or simply *outcomes*. Unless otherwise stated, variables x, y, and z will range over outcomes.

(ii) A positive integer n, called the *number of players*. The n players will be denoted by the integers 1 to n, and variables i, j, and k will range over these integers.

(iii) n sets S_i, one for each i. For each i the members of S_i are called *strategies* for i. The word "strategy," then, refers here to what in game theory is usually called a "pure strategy." An n-tuple $\langle s_1, \ldots, s_n \rangle$, with $s_1 \in S_1, \ldots, s_n \in S_n$ will be called a *strategy n-tuple*. Strategy n-tuples will be indicated by bold-face small letters on the pattern $\boldsymbol{s} = \langle s_1, \ldots, s_n \rangle$, $\boldsymbol{s}' = \langle s_1', \ldots, s_n' \rangle$, and so forth.

(iv) A function g, defined for every strategy n-tuple, whose range is X.

Strictly speaking, a game form is simply a function g which can be characterized as above. We can define a *game form*, then, as a function whose domain is the Cartesian product $S_1 \times \ldots \times S_n$ of a finite number of finite non-empty sets. Its values are called *outcomes*, its arguments are called *strategy n-tuples*, and a member of a set S_i is called a *strategy* for i.

We now define what it is for a game form g to be straightforward. An *ordering* of a set Z is a two-place relation P between members of Z, such that for all x, y, and z in Z,

(1a) $\quad \sim(x\,P\,y\ \&\ y\,P\,x),$

(1b) $\quad x\,P\,z \rightarrow (x\,P\,y \vee y\,P\,z).$

A *preference ordering* is an ordering of X, the set of outcomes. The variable P then means "is preferred to." Distinct x and y may be indifferent under ordering P; if so, neither $x\,P\,y$ nor $y\,P\,x$.

Now we will explain some matters of notation. In the first place, for any two-place relation P between members of X, $x\,R\,y$ will mean $\sim y\,P\,x$, and $x\,I\,y$ will mean $\sim x\,P\,y\ \&\ \sim y\,P\,x$. Thus if P is a preference ordering, P indicates strict preference, I indifference, and R preference or indifference. On the same pattern, $x\,R'\,y$ will mean $\sim y\,P'\,x$, $x\,R_k\,y$ will mean $\sim y\,P_k\,x$, and so forth. Likewise $x\,I'\,y$ will mean $\sim x\,P'\,y\ \&\ \sim y\,P'\,x$, $x\,I_i\,y$ will mean $\sim x\,P_i\,y\ \&\ \sim y\,P_i\,x$, and so forth. In the second place, for any n-tuple indicated by boldface type, the result of altering

its kth place will be indicated by a symbol on the following pattern. Where

$$\boldsymbol{s} = \langle s_1, \ldots, s_n \rangle,$$

we have

$$\boldsymbol{s}k/t = \langle s_1, \ldots, s_{k-1}, t, s_{k+1}, \ldots, s_n \rangle.$$

In other words, $\boldsymbol{s}' = \boldsymbol{s}k/t$ iff $s'_k = t$ and

$$(\forall i)[i \neq k \rightarrow s'_i = s_i].$$

Where P is a preference ordering, a strategy t is *P-dominant* for k if for every strategy n-tuple $\boldsymbol{s}$, $g(\boldsymbol{s}k/t) R g(\boldsymbol{s})$. In other words, t is P-dominant for k iff no matter what strategies are fixed for everyone else, strategy t for k produces an outcome at least as high in preference ordering P as does any other. A game form is *straightforward* if, for every preference ordering P and player k, there is a strategy which is P-dominant for k.

A player k is a *dictator* for game form g if, for every outcome x, there is a strategy $s(x)$ for k such that $g(\boldsymbol{s}) = x$ whenever $s_k = s(x)$. A game form g is *dictatorial* if there is a dictator for g.

THEOREM: *Every straightforward game form with at least three possible outcomes is dictatorial.*

Before proceeding with the proof, we present a corollary.

A *voting scheme* is a game form v such that for some set Z with $X \subseteq Z$, the set S_i of strategies open to each player i is the set of orderings of Z. A voting scheme is *manipulable* if for some k, for some n-tuple $\boldsymbol{P}$ of orderings of Z, and for some ordering P^* of Z, $v(\boldsymbol{P})\ P^*\ v(\boldsymbol{P}k/P^*)$.

COROLLARY: *Every voting scheme with at least three outcomes is either dictatorial or manipulable.*

PROOF (from the principal theorem): Suppose v is non-dictatorial and has at least three possible outcomes. Then, since v is a game form, v is not straightforward, and thus for some k and P, no strategy is P-dominant for k. P here is an ordering of X, the set of outcomes, and a strategy is an ordering P^* of the set Z of alternatives, with $X \subseteq Z$. Let P^* extend P to Z, so that for all x and y in X,

$$x\ P^*\ y \leftrightarrow x\ P\ y. \tag{2}$$

Then, in particular, P^* is not P-dominant for k. Hence for some strategy n-tuple $\boldsymbol{P}$ of orderings of Z, it is not the case that $v(\boldsymbol{P}k/P^*)\ R v(\boldsymbol{P})$, and so $v(\boldsymbol{P})\ P\ v(\boldsymbol{P}k/P^*)$. But since $v(\boldsymbol{P}) \in X$ and $v(\boldsymbol{P}k/P^*) \in X$, from (2), $v(\boldsymbol{P})\ P^*\ v(\boldsymbol{P}k/P^*)$, and v is manipulable. Assuming the main theorem, we have proved the corollary.

Back, now, to the main theorem. We prove it by means of the Arrow impossibility theorem, which I shall now state. A *preference n-tuple* over a set X is an n-tuple

$\langle P_1, \ldots, P_n \rangle$ whose terms are preference orderings of X. Preference n-tuples will be designated in bold-face type on the pattern $\boldsymbol{P} = \langle P_1, \ldots, P_n \rangle$, $\boldsymbol{P}' = \langle P'_1, \ldots, P'_n \rangle$, and so forth. A *social welfare function* is a function whose arguments, for some fixed n and X, are all preference n-tuples $\boldsymbol{P}$ over X, and whose values are preference orderings of X. Arrow [1] showed that every social welfare function violates at least one of the following *Arrow conditions.*

Scope: X has at least three members.

Unanimity: If $P = f(\boldsymbol{P})$ and $(\forall i)\, x\, P_i\, y$, then $x\, P\, y$.

Pairwise Determination:[3] If

$$(\forall i)[x\, P_i\, y \leftrightarrow x\, P'_i\, y],$$

$$(\forall i)[y\, P_i\, x \leftrightarrow y\, P'_i\, x],$$

$$P = f(\boldsymbol{P}),$$

$$P' = f(\boldsymbol{P}'),$$

then

$$x\, P\, y \leftrightarrow x\, P'\, y.$$

Non-dictatorship: There is no dictator for f, where a *dictator* for f is a k such that for every $\boldsymbol{P}$, x, and y, if $x\, P_k\, y$ and $P = f(\boldsymbol{P})$, then $x\, P\, y$.

The proof of the theorem on straightforwardness takes up the remainder of the paper. Let g be a straightforward game form, fixed for the entire proof, with at least three outcomes. We must prove g dictatorial.

Since g is straightforward, for every i and P, there is a strategy s which is P-dominant for i. For each i, let σ_i be a function such that for every P, strategy $\sigma_i(P)$ is P-dominant for i. For each preference n-tuple $\boldsymbol{P}$, let

$$\sigma(\boldsymbol{P}) = \langle \sigma_1(P_1), \ldots, \sigma_n(P_n) \rangle.$$

The functions σ and $\sigma_1, \ldots, \sigma_n$ will be fixed throughout the proof; v will be the composition of g and σ, so that for all $\boldsymbol{P}$, $v(\boldsymbol{P}) = g(\sigma(\boldsymbol{P}))$.

We now use v to generate from each preference n-tuple $\boldsymbol{P}$ a two-place relation $f(\boldsymbol{P})$ which turns out to be an ordering. The function f so defined is thus an Arrow social welfare function. We shall show that f satisfies all the Arrow conditions except non-dictatorship, and is therefore dictatorial. From this it will follow that g is dictatorial. Hence any straightforward game form with at least three outcomes is dictatorial.

A *chain ordering* is an ordering in which no distinct items are indifferent: P is a chain ordering iff

$$(\forall x)(\forall y)[x\, I\, y \rightarrow x = y].$$

Let Q be a chain ordering of X, fixed for the entire proof. We let preference n-tuple $\boldsymbol{P}$ determine a two-place relation between members of X in the following way.

[3] Arrow gives an equivalent condition, the *independence of irrelevant alternatives*, but in effect uses the condition given here in the proof of his theorem.

Let $Z \subseteq X$. For each i, we derive a chain ordering $P_i * Z$ of X from the ordering P_i by moving the members of Z to the top, preserving their ordering with respect to each other except in case of ties, and otherwise ordering everything according to Q. In other words, for each pair of alternatives x and y,

(3a) If $x \in Z$ and $y \in Z$, then $x\,(P_i * Z)\,y$ iff either $x\,P_i\,y$ or both $x\,I_i\,y$ and $x\,Q\,y$.

(3b) If $x \in Z$ and $y \notin Z$, then $x\,(P_i * Z)\,y$.

(3c) If $x \notin Z$ and $y \notin Z$, then $x\,(P_i * Z)\,y$ iff $x\,Q\,y$.

For each i we have defined a two-place relation $P_i * Z$ between members of X. Let

$$\boldsymbol{P} * Z = \langle P_1 * Z, \ldots, P_n * Z \rangle.$$

The following features of the $*$ operator follow easily from (3a)–(3c). (i) For each i, $P_i * Z$ is a chain ordering. (ii) If $Y \subseteq Z$, then $(\boldsymbol{P} * Z) * Y = \boldsymbol{P} * Y$. (iii) Suppose $\boldsymbol{P}$ and $\boldsymbol{P}'$ agree on Z; that is, suppose

$$(\forall i)(\forall x)(\forall y)\,[(x \in Z\ \&\ y \in Z) \rightarrow (x\,P_i\,y \leftrightarrow x\,P'_i\,y)].$$

Then $\boldsymbol{P} * Z = \boldsymbol{P}' * Z$. Features (i)–(iii) will be cited at later points in the proof.

Now let xPy be the relation

$$x \neq y\ \&\ x = v(\boldsymbol{P} * \{x, y\}).$$

We have defined a two-place relation P as a function of $\boldsymbol{P}$, and we shall call that function f, so that $P = f(\boldsymbol{P})$. We shall show in the following three assertions that f is an Arrow social welfare function which satisfies the Arrow conditions other than non-dictatorship.

In what follows, P means $f(\boldsymbol{P})$, P' means $f(\boldsymbol{P}')$, and so forth. Also, xRy means $\sim y\,P\,x$, and so

$$x\,R\,y \leftrightarrow [x = y \vee y \neq v(\boldsymbol{P} * \{x, y\})].$$

Hence, (iv) if $x\,P\,y$, then $x\,R\,y$.

As a first step in showing that f satisfies the Arrow conditions, note that f satisfies the Arrow condition of *pairwise determination.*

(v) Suppose

$$(\forall i)\,[x\,P_i\,y \leftrightarrow x\,P'_i\,y],$$

$$(\forall i)\,[y\,P_i\,x \leftrightarrow y\,P'_i\,x].$$

Then $x\,P\,y \leftrightarrow x\,P'\,y$.

For then from (iii), $\boldsymbol{P} * \{x, y\} = \boldsymbol{P}' * \{x, y\}$, and hence

$$x \in v(\boldsymbol{P} * \{x, y\}) \leftrightarrow x \in v(\boldsymbol{P}' * \{x, y\}),$$

which is to say $x\,P\,y \leftrightarrow x\,P'\,y$.

Assertion 1, which follows, is used in a number of ways throughout the rest of the proof of the theorem. It says, in effect, that if $y\,R\,x$, and nobody is indifferent

between x and y, then given the strategies of those who prefer y to x, those who prefer x to y could not, whatever their strategies, get x chosen.

ASSERTION 1: *Let* $\boldsymbol{s} = \sigma(\boldsymbol{P})$. *Suppose for strategy n-tuple* $\boldsymbol{s}'$ *and alternatives* x *and* y, $x \neq y$, *and*

(4a) $(\forall i)[y\, P_i\, x \rightarrow s_i' = s_i]$,

(4b) $(\forall i) \sim x\, I_i\, y$,

(4c) $y\, R\, x$.

Then $x \neq g(\boldsymbol{s}')$.

PROOF: Suppose on the contrary that $x = g(\boldsymbol{s}')$. We shall show that for some k, $\sigma_k(P_k)$ is not P_k-dominant for k, contrary to what has been stipulated for σ_k. Let $\boldsymbol{P}^* = \boldsymbol{P} * \{x, y\}$, and let strategy n-tuple $\boldsymbol{t} = \sigma(\boldsymbol{P}^*)$. Then $y\, R\, x$ means

$$x = y \vee x \neq v(\boldsymbol{P}^*),$$

and since $x \neq y$ and $v(\boldsymbol{P}^*) = g(\sigma(\boldsymbol{P}^*)) = g(\boldsymbol{t})$, we have $x \neq g(\boldsymbol{t})$. Now let $\boldsymbol{s}^0, \ldots, \boldsymbol{s}^n$ be the sequence obtained by starting with $\boldsymbol{s}'$ and at each step k replacing s_k' with t_k. Thus

$$\boldsymbol{s}^0 = \boldsymbol{s}',$$
$$\boldsymbol{s}^k = \boldsymbol{s}^{k-1} k / t_k,$$

and in reverse order,

$$\boldsymbol{s}^n = \boldsymbol{t},$$
$$\boldsymbol{s}^{k-1} = \boldsymbol{s}^k k / s_k'.$$

In other words, for each k, $\boldsymbol{s}^k$ is the strategy n-tuple $\langle s_1^k, \ldots, s_n^k \rangle$ such that for each i,

(5a) $i \leqslant k \rightarrow s_i^k = t_i$,

(5b) $i > k \rightarrow s_i^k = s_i'$,

so that

$$\boldsymbol{s}^0 = \langle s_1', s_2', s_3', \ldots, s_n' \rangle,$$
$$\boldsymbol{s}^1 = \langle t_1, s_2', s_3', \ldots, s_n' \rangle,$$
$$\boldsymbol{s}^2 = \langle t_1, t_2, s_3', \ldots, s_n' \rangle,$$

and so forth. Since $x = g(\boldsymbol{s}')$ but $x \neq g(\boldsymbol{t})$, we have $x = g(\boldsymbol{s}^0)$ but $x \neq g(\boldsymbol{s}^n)$. Let k be the least such that $g(\boldsymbol{s}^k) \neq x$. We shall show that either t_k is not P_k^*-dominant for k or s_k is not P_k-dominant for k. Since $t_k = \sigma_k(P_k^*)$, and $s_k = \sigma_k(P_k)$, in either case the original characterization of σ is violated. The supposition that $x = g(\boldsymbol{s}')$ is therefore false. Consider two cases, jointly exhaustive.

CASE 1: $g(\boldsymbol{s}^k) = y$ and $y\ P_k\ x$.

Then since $g(\boldsymbol{s}^{k-1}) = x$, we have $g(\boldsymbol{s}^k)\ P_k\ g(\boldsymbol{s}^{k-1})$, and since $\boldsymbol{s}^{k-1} = \boldsymbol{s}^k k/s'_k$, it is not the case that $g(\boldsymbol{s}^k k/s'_k)\ R_k\ g(\boldsymbol{s}^k)$, and thus s'_k is not P_k-dominant for k. But since $y\ P_k\ x$, by (4a), $s'_k = s_k$, and s_k is not P_k-dominant for k. But since $s_k = \sigma_k(P_k)$, s_k is P_k-dominant for k, and we have a contradiction.

CASE 2: $g(\boldsymbol{s}^k) \neq y$ or $x\ P_k\ y$.

Then we always have $x\ P_k^*\ g(\boldsymbol{s}^k)$. For if $g(\boldsymbol{s}^k) = y$, then $x\ P_k\ y$, and by (3a) and the definition of $\boldsymbol{P}^*$, $x\ P_k^*\ y$. If, instead, $g(\boldsymbol{s}^k) \neq y$, then since $g(\boldsymbol{s}^k) \neq x$, we have $g(\boldsymbol{s}^k) \notin \{x, y\}$, and by (3b), again $x\ P_k^*\ g(\boldsymbol{s}^k)$. Now $x = g(\boldsymbol{s}^{k-1})$, and $\boldsymbol{s}^k = \boldsymbol{s}^{k-1}k/t_k$. Hence $g(\boldsymbol{s}^{k-1})\ P_k^*\ g(\boldsymbol{s}^{k-1}k/t_k)$, and t_k is not P_k^*-dominant for k. But since $t_k = \sigma_k(P_k^*)$, t_k is P_k^*-dominant for k, and we have a contradiction. In both Cases 1 and 2, then, the supposition that $x = g(\boldsymbol{s}')$ leads to a contradiction. Since by (4b), $\sim x\ I_k\ y$, the two cases exhaust the possibilities. Therefore $x \neq g(\boldsymbol{s}')$, and the assertion is proved.

From Assertion 1, a number of properties of v follow easily.

COROLLARY 1: *If* $(\forall i)\ x\ P_i\ y$, *then* $x\ P\ y$.

PROOF: Since x is an outcome, for some strategy n-tuple $\boldsymbol{s}'$, $x = g(\boldsymbol{s}')$. If $\boldsymbol{s} = \sigma(\boldsymbol{P})$, then all the hypotheses of Assertion 1 are satisfied except for (4c), and the conclusion of Assertion 1 is violated. Therefore (4c) is false, and $x\ P\ y$.

COROLLARY 2: *If* $(\forall i) \sim x\ I_i\ y$ *and* $y\ R\ x$, *then* $v(\boldsymbol{P}) \neq x$.

PROOF: (4b) and (4c) in Assertion 1 are satisfied. Let $\boldsymbol{s}' = \boldsymbol{s} = \sigma(\boldsymbol{P})$. Then in addition, (4a) is satisfied, and by Assertion 1, $g(\boldsymbol{s}') \neq x$. Thus since $g(\boldsymbol{s}') = g(\sigma(\boldsymbol{P})) = v(\boldsymbol{P})$, we have $v(\boldsymbol{P}) \neq x$.

COROLLARY 3: *If* $(\forall i) \sim x\ I_i\ y$ *and* $v(\boldsymbol{P}) = x$, *then* $x\ P\ y$.

This is the contrapositive of Corollary 2.

ASSERTION 2: *P is a preference ordering.*

P clearly satisfies the condition

$$(\forall x)(\forall y) \sim (x\ P\ y\ \&\ y\ P\ x),$$

for $x\ P\ y$ means

$$x \neq y\ \&\ x = v(\boldsymbol{P} * \{x, y\}),$$

and $y\ P\ x$ means

$$y \neq x\ \&\ y = v(\boldsymbol{P} * \{x, y\}).$$

It remains to show that P satisfies the condition that for all x, y, and z,

$$x\,P\,z \rightarrow (\forall y)(x\,P y \vee y\,P\,z).$$

Let $\boldsymbol{P}' = \boldsymbol{P} * \{x, y, z\}$. Then from (ii), $\boldsymbol{P}' * \{x, z\} = P * \{x, z\}$, so that since

$$x\,P\,z \leftrightarrow [x \neq z \,\&\, x = v(\boldsymbol{P} * \{x, z\})],$$

$$x\,P'\,z \leftrightarrow [x \neq z \,\&\, x = v(\boldsymbol{P}' * \{x, z\})],$$

we have

$$x\,P'\,z \leftrightarrow x\,P\,z.$$

Similarly,

$$x\,P'\,y \leftrightarrow x\,P\,y,$$

$$y\,P'\,z \leftrightarrow y\,P\,z.$$

We need only to show, then, that for all x, y, and z,

$$x\,P'\,z \rightarrow (x\,P'\,y \vee y\,P'\,z).$$

Suppose $x\,P'\,z$. Then x and z are distinct, since $x\,P'\,z$ means

$$x \neq z \,\&\, x = v(\boldsymbol{P}' * \{x, z\}).$$

If $y = x$, we have $y\,P'\,z$, and if $y = z$, we have $x\,P'\,y$. There remains the case where $y \neq x$ and $y \neq z$. Then by (i) and the definition of $\boldsymbol{P}'$, each P_i' is a chain ordering, and

$$(\forall i)[\sim x\,I_i'\,y \,\&\, \sim x\,I_i'\,z \,\&\, \sim y\,I_i'\,z].$$

CASE 1: $x = v(\boldsymbol{P}')$.

Then by Corollary 3 to Assertion 1, $x\,P'\,y$.

CASE 2: $x \neq v(\boldsymbol{P}')$.

Since $x\,P'\,z$, by Corollary 2 to Assertion 1, $z \neq v(\boldsymbol{P}')$. If $w \notin \{x, y, z\}$, then by (3b) and the definition of $\boldsymbol{P}'$, $(\forall i)\,x\,P_i'\,w$. Hence by Corollary 1 to Assertion 1, $x\,P'\,w$, and by Corollary 2, $w \neq v(\boldsymbol{P}')$. We have, then, $x \neq v(\boldsymbol{P}')$, $z \neq v(\boldsymbol{P}')$, and if $w \notin \{x, y, z\}$, then $w \neq v(\boldsymbol{P}')$. Thus by elimination, $y = v(\boldsymbol{P}')$, and by Corollary 3 to Assertion 1, $y\,P'\,z$. From $x\,P'\,z$, we have shown that in Case 1, $x\,P'\,y$, and in Case 2, $y\,P'\,z$. This, as we said, suffices to show that P is an ordering, and the assertion is proved.

ASSERTION 3: *If g has at least three possible outcomes, then f violates the Arrow condition of non-dictatorship.*

That is, there is an individual k, called the dictator for f, such that for every preference n-tuple $\boldsymbol{P}$ and every x and y, $x\,P_k\,y \rightarrow x\,P\,y$.

PROOF: Since by Assertion 2, the values of f are preference orderings, f is an Arrow social welfare function. We have shown that f satisfies all of the Arrow conditions but *non-dictatorship*. In the first place, since every outcome of g is an outcome of v, *scope* holds. By (v), f satisfies *pairwise determination*, and by Corollary 1 to Assertion 1, f satisfies *unaminity*. Therefore, since the Arrow theorem says that no social welfare function satisfies all four Arrow conditions, f violates *non-dictatorship*.

It remains to show that since f violates the Arrow condition of *non-dictatorship*, g is dictatorial.

ASSERTION 4: *The dictator for f is dictator for g.*

PROOF: Let k be dictator for f. Then k is dictator for g if for every outcome y, there is a strategy $s(y)$ for k such that

$$(6) \qquad (\forall \boldsymbol{s}')[s'_k = s(y) \rightarrow g(\boldsymbol{s}') = y].$$

Let P^y be any ordering such that $(\forall x)[x \neq y \rightarrow y\, P^y\, x]$ and let $s(y) = \sigma_k(P^y)$. We appeal to Assertion 1 to show that this $s(y)$ satisfies (6).

Let $\boldsymbol{s}'$ be such that $s'_k = s(y)$, and suppose $x \neq y$. Then by the way P^y was characterized, $y\, P^y\, x$. We shall show that $g(\boldsymbol{s}') \neq x$. Let $\boldsymbol{P}$ be any preference n-tuple such that

$$(7a) \qquad P_k = P^y,$$

$$(7b) \qquad (\forall i)[i \neq k \rightarrow x\, P_i\, y],$$

and let $\boldsymbol{s} = \sigma(\boldsymbol{P})$. Then $s_k = \sigma_k(P_k) = \sigma_k(P^y) = s(y) = s'_k$. Thus since by (7b), only k prefers y to x, (4a) is satisfied, and since in addition, $y\, P^y\, x$ and hence by (7a) $y\, P_k\, x$, (4b) is satisfied also. Since $y\, P_k\, x$ and k is dictator for f, we have $y\, P\, x$, and (4c) is satisfied. Therefore by Assertion 1, $x \neq g(\boldsymbol{s}')$. We have shown that if $x \neq y$, then $x \neq g(\boldsymbol{s}')$. Hence $y = g(\boldsymbol{s}')$.

We have shown, then, that if $s(y) = \sigma_k(P^y)$, then (6) is satisfied. Thus k is dictator for g, and g is dictatorial. This completes the proof that any straightforward voting scheme with at least three outcomes is dictatorial.

University of Chicago

Manuscript received October, 1971.

REFERENCES

[1] ARROW, KENNETH J.: *Social Choice and Individual Values.* New York: Wiley, 1963.
[2] ———: "Values and Collective Decision Making," in P. Laslett and W. G. Runciman, eds., *Philosophy, Politics, and Society, Third Series.* Oxford: Blackwell, 1967.
[3] DUMMETT, M., AND R. FARQUHARSON: "Stability in Voting," *Econometrica*, 29 (1961), 33–43.
[4] MURAKAMI, YASUSUKE: *Logic and Social Choice.* New York: Dover, 1968.
[5] VICKREY, WILLIAM: "Utility, Strategy, and Social Decision Rules," *Quarterly Journal of Economics*, 74 (1960), 507–535.