Discussion Paper No. 89

THE POSSIBILITY OF A CHEAT PROOF SOCIAL CHOICE FUNCTION: A THEOREM OF A. GIBBARD AND M. SATTERTHWAITE

by

David Schmeidler *

and

Hugo Sonnenschein **

Revised: May 1974 ***

* Tel Aviv University and the Foerder Institute

** Northwestern University

*** We are indebted to Mark Satterthwaite, Leonid Hurwicz, and Allan Gibbard for criticism and suggestions. David Schmeidler's research was partially supported by a grant from the Foerder Institute and Hugo Sonnenschein's by a grant from the National Science Foundation.

THE POSSIBILITY OF A CHEAT PROOF SOCIAL CHOICE FUNCTION:

A THEOREM OF A. GIBBARD AND M. SATTERTHWAITE

by

David Schmeidler, Tel Aviv University and the
Foerder Institute

Hugo Sonnenschein, Northwestern University

Revised: May 1974

We prove a result which was formulated independently by A. Gibbard [2] and M. Satterthwaite [3]. Their theorem is of importance because it provides an attractive new way of viewing Arrow's classic result on Social Welfare Functions [1]. By requiring that strategic considerations cannot be beneficially employed, it frees the statement of the "General Possibility Theorem" from the assumption of independence of irrelevant alternatives.

The purpose of this paper is to present a simple proof of the Gibbard-Satterthwaite theorem, and to design the presentation so that the transition from Arrow's books to these pages can be made with minimum effort. We hope that this will increase the rate at which their theorem is absorbed into the literature, for it richly deserves substantial recognition.

The symbol $\mathcal{A}$ denotes a set of basic alternatives, and $\Sigma(\mathcal{A})$ denotes the class of orderings (connected (xPy,yPx, or x = y), asymmetric, and transitive relations) on $\mathcal{A}$. A Social Welfare Function (SWF) F^* is a function $F^*:\Sigma^n(\mathcal{A}) \to \Sigma(\mathcal{A})$. Note that we have "built in" the assumption of universal domain and that individuals are not permitted "indifference"; the latter requirement will be dropped at the conclusion of the paper. A generic element of the domain of F^* is called a preference profile and is denoted by $(P_1,P_2,\ldots,P_n)$. A generic element of the range is called a social preference relation and is denoted by P.

A SWF F^* satisfies

Independence of Irrelevant Alternatives (IIA) if: for all x,y and all pairs of preference profiles $(P_1,P_2,\ldots,P_n)$ and $(P_1',P_2',\ldots,P_n')$, $x\ P_i\ y$ if and only if $xP_i'y$ for all i, implies $F^*(P_1,P_2,\ldots,P_n) = F^*(P_1',P_2',\ldots,P_n')$.

Pareto (P) if:

for profiles $(P_1,P_2,\ldots,P_n)$, xP_iy for all i implies xPy, where $P = F^*(P_1,P_2,\ldots,P_n)$.

Arrow's Theorem [1] states:

If a SWF F^* satisfied (IIA), (P), and $\mathcal{A}$ has at least three elements, then there exist an "i" such that for all $(P_1,P_2,\ldots,P_n)$, $F^*(P_1,P_2,\ldots,P_n) = P_i$. (Individual "i" is called a SWF dictator.)

A Social Choice Function (SCF) is a function $F:\Sigma^n(\mathcal{A}) \to \mathcal{A}$. A SCF is manipulable at $(P_1,P_2,\ldots,P_n) \in \Sigma^n(\mathcal{A})$ if there exists $P_i' \in \Sigma(\mathcal{A})$ such that $F(P_1,P_2,\ldots,P_i',\ldots,P_n)\ P_i\ F(P_1,P_2,\ldots,P_i,\ldots,P_n)$. The SCF F is cheat proof [2] if there is no preference profile at which it is manipulable. A Social Choice Function is dictatorial if there exists an i such that for all $(P_1,P_2,\ldots,P_n)$ and for all $x \neq F(P_1,P_2,\ldots,P_n)$ in the range of $F, F(P_1,P_2,\ldots,P_n)P_i\ x$. (Individual "i" is called a SCF dictator.) We will now state the theorem which is the subject of this paper.

---

[1] The Theorem is usually stated for preferences represented by weak orderings ("indifference" allowed); however, most proofs of Arrow's Theorem apply equally well to the case we consider here.

[2] We observe the relation to Nash equilibrium.

Theorem (Gibbard - Sattherthwaite): If a Social Choice Function F is cheat proof and the range of $\mathcal{A}'$ of F contains at least three alternatives, then it is dictatorial.

Proof:

(i) If for some i, $F(P_1,P_2,\ldots,P_i,\ldots,P_n)=x$, $F(P_1,P_2,\ldots,P'_i,\ldots,P_n)=y$, with $x \neq y$ and xP_iy if and only if xP'_iy, then F is manipulable (at either $(P_1,P_2,\ldots,P_i,\ldots,P_n)$ or $(P_1,P_2,\ldots,P'_i,\ldots,P_n)$.)

(ii) Assume that F is cheat proof with range $\mathcal{A}' \subset \mathcal{A}$. It follows that: if $B \subset \mathcal{A}'$ and $(P_1,P_2,\ldots,P_n)$ satisfies the condition that for each i, $x \in B$ and $y \in \mathcal{A} \setminus B$ implies xP_iy, then $F(P_1,P_2,\ldots,P_n) \in B$.

Proof: If not let $F(P_1,P_2,\ldots,P_n) = y \notin B$ and $F(P'_1,\ldots,P'_n) = x \in B$. Consider $\{z_i\}_{i=0}^{i=n}$ defined by $z_i = F(P'_1,P'_2,\ldots,P'_i,P_{i+1},\ldots,P_n)$ and let j be the least integer such that $z_i \in B$. Then F is manipulable (at $(P'_1,P'_2,\ldots,P'_{j-1},P_j,\ldots,P_n)$), which contradicts the assumption that F is cheat proof.

(iii) If F is cheat proof with range $\mathcal{A}' \subset \mathcal{A}$, then it generates a SWF $F^*: \Sigma^n(\mathcal{A}') \to \Sigma(\mathcal{A})$ which satisfies (P) and (IIA).

Given a profile $(P_1,P_2,\ldots,P_n)$ and any pair $\{x,y\}$, write xPy if $x = F(P'_1,P'_2,\ldots,P'_n)$, where P'_i is derived from P_i by moving x and y to the top of i's list and preserving the P_i ordering within $\{x,y\}$ and $\mathcal{A}\setminus\{x,y\}$. (More formally, for $(P_1,P_2,\ldots,P_n) \in \Sigma^n(\mathcal{A})$ and any pair $x,y \in \mathcal{A}$, $P'_i (i = 1,2,\ldots,n)$ is defined as follows: (a) xP'_iy if and only if xP_iy and (b) for all w,z such that $\{w,z\} \cap \{x,y\} = \emptyset$, xP'_iz, yP'_iz, and wP'_iz if and only if wP_iz.) The relation P is connected by (ii) and asymmetric by definition. The function F^* defined by $(P_1,P_2,\ldots,P_n) \to P$ satisfies (IIA) by (i) and a chain argument similar to that given in (ii). Also by (ii) it satisfies (P) on $\mathcal{A}'$. It remains to prove that for each $(P_1,P_2,\ldots,P_n)$, $F^*(P'_1,P'_2,\ldots,P'_n)$ is transitive.

Proof of Transitivity of P: If P is not transitive there exists $(P_1,P_2,\ldots,P_n)$ such that with x and y taken to the top (as in the definition of P) x is the social choice, with y and z taken to the top y is the social choice, and with x and z taken to the top z is the social choice. Let $(P'_1,P'_2,\ldots,P'_n)$ be defined by taking x,y and z to the top of i's list and preserving the P_i ordering within the sets $\{x,y,z\}$ and $\mathcal{O}\setminus\{x,y,z\}$, $i = 1,2,\ldots,n$. Assume without loss of generality that $x = F(P'_1,P'_2,\ldots,P'_n)$. Let $(P''_1,\ldots,P''_n)$ be defined by moving y to third position in each of the primed preferences. This puts x and z on top in each of the double primed preferences, and since xP_iz if and only if xP''_iz for all i, by assumption and (IIA), $F(P''_1,\ldots,P''_n) = z$. Proceeding as in (ii), let $w_i = F(P''_1,P''_2,\ldots,P''_{i-1},P''_i,P'_{i+1},\ldots,P'_n)$, $0 \leqq i \leqq n$, and let j be the first integer such that $w_i \neq x$. If $w_j = z$ then (i) establishes that F is manipulable if $w_j = y$ then F is manipulable (by j at $(P''_1,P''_2,\ldots,P''_j,P'_{j+1},\ldots,P'_n)$.) This contradicts the assumption that F is cheat proof, and thus P must be transitive.

(iv) Now, assume that F satisfies the hypothesis of the theorem. By (iii), the fact that the range $\mathcal{O}'$ of F has at least three elements, and Arrow's Theorem, there exists an individual "i" who is a SWF dictator for F^*. But this imples that "i" is also a SCF dictator for F, and the proof is completed.

Finally, we observe that the proof can be extended to the situation in which preferences are represented by weak orderings ("indifference" allowed).[3/] To accomplish this observe first that if F is cheat proof on weak order

---

[3/] Both Gibbard and Satterthwaite consider this case.

profiles (with three alternatives in the range), then of course it is cheat proof when restricted to those profiles which are orderings in every coordinate ("indifference" not allowed). Also, since the proof in (ii) applies to weak order profiles, the range of F cannot shrink, and thus by the theorem we have just demonstrated, there is an individual "i" who is a SCF dictator on the restricted domain. We will now prove that for all weak order profiles $(R_1,R_2,\ldots,R_n)$, $F(R_1,R_2,\ldots,R_n)$ must belong to the set of R_i maximal elements in the range of F, i.e., "i" is a SCF dictator on the entire domain of F. Let B denote the R_i maximal elements in the range of F and let $(P_1,P_2,\ldots,P_n) \in \Sigma^n(\mathcal{A})$ such that for all $y \in B$ and $z \in \mathcal{A} \setminus B$, both yP_1z and $zP_iy (i \neq 1)$. Define $w_i = F(P_1,P_2,\ldots,P_i,R_{i+1},\ldots,R_n)$, $0 \leq i \leq n$, and let j be the least i such that $w_i \in B$. If $j = 1$, then F is manipulable by 1; if $j > 1$, then F is again maniuplable (by j at $(P_1,P_2,\ldots,P_j,R_{j+1},\ldots,R_n)$.)

REFERENCES

[1] Arrow, K. J., Social Choice and Individual Values (2nd ed.), New York, John Wiley and Sons, Inc., 1963.

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

[3] Satterthwaite, M., "The Existence of a Strategy Proof Voting Procedure: A Topic in Social Choice Theory," Ph.D. Dissertation, University of Wisconsin, Madison, 1973.