Strategy-Proofness and Pivotal Voters: A Direct Proof of the Gibbard-Satterthwaite Theorem
Author(s): Salvador Barberá
Source: *International Economic Review*, Vol. 24, No. 2 (Jun., 1983), pp. 413-417
Published by: Blackwell Publishing for the Economics Department of the University of Pennsylvania and Institute of Social and Economic Research -- Osaka University
Stable URL: http://www.jstor.org/stable/2648754
Accessed: 28/08/2011 15:38

INTERNATIONAL ECONOMIC REVIEW
Vol. 24, No. 2, June, 1983

STRATEGY-PROOFNESS AND PIVOTAL VOTERS: A DIRECT PROOF OF THE GIBBARD-SATTERTHWAITE THEOREM*

BY SALVADOR BARBERÁ

1. INTRODUCTION

This paper presents a proof of the Gibbard-Satterthwaite theorem. The importance of the theorem itself has been widely recognized, and there already exist several alternative proofs. The present one is offered because it is technically simple, *direct*, and emphasizes the role of *pivotal individuals* in the manipulation of voting schemes.

Some indirect proofs of the theorem rely on the connections between voting schemes and social welfare functions, and one interesting corollary of such proofs is to provide new arguments in behalf of Arrow's condition of independence of irrelevant alternatives. Yet, a *direct* proof has the advantage of showing the complete independence of the manipulation phenomena from any requirement of consistency in choice under varying sets of alternatives.

Also, our proof emphasizes the role of *pivotal individuals* in manipulation. *A classical approach in social choice theory is to concentrate on the global distribution of power among groups of agents*, i.e., on the ability of coalitions to impose outcomes regardless of the preferences of agents not in the coalition. For example, standard proofs of Arrow's Theorem concentrate on decisive coalitions, to end up proving that there must exist one such coalition containing a single individual, the dictator. *Our approach concentrates on the local distribution of power among single agents*, i.e., on the ability of each single individual to determine the social outcome, given the preferences of all other members of society.[1] This approach may be more intuitive for some readers; moreover, it closely parallels the work on incentives in economic environments with public goods and suggests a possible way to unify results which have been obtained from different initial specifications. We do not attempt such a rejoinder here, but let us observe the following. For certain environments with public goods, where transfer payments make perfect sense, nontrivial strategy-proof mechanisms can be obtained by appropriately taxing pivotal agents so that they take into account the costs that their actions impose upon others.[2] One explanation for the nonexistence of equivalent mechanisms in our case is that the environment, as specified, does not allow for a corresponding procedure whereby different costs to individuals could be attached to different acts of preference revelation.

* Manuscript received June 12, 1981; revised May 6, 1982.

[1] See Barberá [1980] for a proof of Arrow's Theorem along these lines.

[2] See Green and Laffont [1978, Part 2]

The paper deals for expository reasons with the case where individual preferences are strong, i.e., no individual is ever indifferent between two alternatives. The extension to the case where individual preferences are weak orderings is immediate and involves no additional conceptual problems. A short argument can be found in Schmeidler and Sonnenschein [1978].

2. STATEMENT AND PROOF OF THE GIBBARD-SATTERTHWAITE THEOREM

Individuals. Let $I=\{1, 2,\ldots, n\}$, an initial segment of the integers. Elements of I are called the *individuals*.

Alternatives. Let A be a finite set. Elements of A are called the *alternatives*, and they are denoted by $x, y, z, \ldots$.

Preference Relations. Let $\mathscr{P}$ be the set of complete, reflexive, transitive and antisymmetric binary relations on A. Elements of $\mathscr{P}$ are called (*strong*) *preference orderings* and are denoted by $P, P', P_i, \ldots$.

Given $P \in \mathscr{P}$ and $x \in A$, we say that P' *is a reshuffling of* P *around* x iff $(\forall z)(zP'x \leftrightarrow zPx)$. In words, a reshuffling of a preference ordering around an alternative x is another ordering under which x preserves the same positions relative to all other alternatives.

Let $r(P, x)$ denote the set of all reshufflings of P around x.

Given $P \in \mathscr{P}$ and $x \in A$, ${}_xP$ denotes the ordering obtained from P by placing x down to last place. Formally, ${}_xP$ is defined so that $(\forall z)z\,{}_xPx$, and $(\forall y, z)[y \neq x \;\&\; z \neq x] \rightarrow [yPz \leftrightarrow y\,{}_xPz]$.

Similarly, ${}^xP$ denotes the ordering obtained from P by lifting x up to first place. Formally, ${}^xP$ is defined so that $(\forall z)x\,{}^xPz$ and $(\forall y, z)[y \neq x \;\&\; z \neq x] \rightarrow [yPz \leftrightarrow y\,{}^xPz]$.

Given $P \in \mathscr{P}$ and $Y \subseteq A$, the *choice of* P *in* Y is the element $C(P, Y) \equiv \{x \in Y \mid xPz (\forall z \in Y)\}$.

Preference Profiles. Let $\mathscr{P}^n$ stand for the n-fold cartesian product of $\mathscr{P}$. Elements of $\mathscr{P}^n$ are denoted by $\boldsymbol{P}, \boldsymbol{P}', \ldots$, and are called *preference profiles*. When this does not lead to confusion, the i-th element of a profile $\boldsymbol{P}$ will be denoted by P_i, that of a profile $\boldsymbol{P}'$ by $P'_i, \ldots$.

Given $\boldsymbol{P} \in \mathscr{P}^n$ and $P'_i \in \mathscr{P}$, $\boldsymbol{P}/P'_i$ denotes the profile $\hat{\boldsymbol{P}}$ where $\hat{P}_i = P'_i$ and $(\forall j \neq i)\hat{P}_j = P_j$. That is, $\boldsymbol{P}/P'_i$ is obtained from $\boldsymbol{P}$ by just changing to P'_i the preferences of individual i, and keeping those of all other individuals the same.

Voting Schemes. A *voting scheme* is a function $f: \mathscr{P}^n \rightarrow A$. We denote the range of f by t_f, and let $\# t_f$ be the number of elements it contains.

A voting scheme is *dictatorial* iff $\exists i \in I$ such that $(\forall \boldsymbol{P} \in \mathscr{P}^n) f(\boldsymbol{P}) = C(P_i, t_f)$. That is, if the social outcome is always the alternative that one fixed individual considers best among those which ever obtain.

A voting scheme is *manipulable* iff $\exists \boldsymbol{P} \in \mathscr{P}^n$, $P' \in \mathscr{P}$, $i \in I$ such that

$f(\boldsymbol{P}/P_i)P_i f(\boldsymbol{P})$. We then say that i can manipulate f at $\boldsymbol{P}$ via P_i'.

If f is not manipulable, it is *strategy-proof.*

Pivotal voters. Given a voting scheme f, the options of an individual i at profile $\boldsymbol{P}$ are those alternatives that may become the social outcome for different choices in $\mathscr{P}$ of i's stated preferences, when individuals other than i declare their preferences to be those in $\boldsymbol{P}$. Formally, *the set $o(i, \boldsymbol{P})$ of individual i's options at profile $\boldsymbol{P}$* is defined by

$$o(i, \boldsymbol{P}) = \{x \in A / \exists\, P_i' \in \mathscr{P} \ni f(\boldsymbol{P}/P_i') = x\}.$$

Clearly, $o(i, \boldsymbol{P})$ is never empty, since $(\forall P)(\forall i) f(P) \in o(i, \boldsymbol{P})$. When $o(i, \boldsymbol{P})$ contains more than one element, we say that *i is pivotal at $\boldsymbol{P}$.* Thus, an individual is pivotal at a profile f if he has at least one way of changing the social outcome by changing his vote.

THE GIBBARD-SATTERTHWAITE THEOREM. *Let f be a voting scheme whose range contains more than two alternatives. Then f is either dictatorial or manipulable.*

PROOF. It will suffice to start from any specific voting scheme f, (with $\# t_f > 2$) which is strategy proof, and show that it must be dictatorial.

(a) Notice that, since f is strategy-proof, $(\forall P \in \mathscr{P}^n)(\forall i \in I) f(\boldsymbol{P}) \in C(o(i, \boldsymbol{P}), P_i)$. That is: *for every profile, the outcome must be the best option at that profile for each one of the individuals.*

(b) *No individual can change the outcome of f at a profile by changing his preferences to a reshuffling around this outcome.* Formally, $(\forall \boldsymbol{P} \in \mathscr{P}^n)$ $(\forall i \in I)[P' \in r(P_i, f(\boldsymbol{P})] \rightarrow [f(\boldsymbol{P}/P_i') = f(\boldsymbol{P})]$. This is because, otherwise, i could manipulate f either at $\boldsymbol{P}$ via P' or at $\boldsymbol{P}/P'$ via P_i.

(c) *An individual who changes his preferences to a reshuffling around the outcome of a profile cannot change the options of any other individual.* Formally, where $\boldsymbol{P}' = \boldsymbol{P}/P_i'$, $(\forall \boldsymbol{P} \in \mathscr{P}^n)(\forall i, j \in I)[P_i' \in r(P_i, f(\boldsymbol{P}))] \rightarrow [o(j, P) = o(j, P')]$. To prove it, it will suffice to show that $o(j, \boldsymbol{P}') \subseteq o(j, \boldsymbol{P})$, since $f(\boldsymbol{P}) = f(\boldsymbol{P}')$ by (b), $P_i \in r(P_i', f(\boldsymbol{P}'))$ by definition, and all the argument could thus be reversed to show conversely that $o(j, \boldsymbol{P}) \subseteq o(j, \boldsymbol{P}')$.

Suppose, then, that $o(j, \boldsymbol{P}') \not\subseteq o(j, \boldsymbol{P})$, and let $y \in o(j, \boldsymbol{P}')$, $y \notin o(j, \boldsymbol{P})$. Let $f(\boldsymbol{P}) = f(\boldsymbol{P}') = x \neq y$. Since $x = C(o(j, \boldsymbol{P}), P_j)$, and $y \notin o(j, \boldsymbol{P})$, we have that $x = C(o(j, \boldsymbol{P}), {}^yP_j) = C(o(j, \boldsymbol{P}/{}^yP_j)) = f(\boldsymbol{P}/{}^yP_j)$. On the other hand, since $y \in o(j, \boldsymbol{P}')$, it must be that $f(\boldsymbol{P}'/{}^yP_j) = y$. Since $\boldsymbol{P}'/{}^yP_j = (\boldsymbol{P}/{}^yP_j)/P_i'$, either i can manipulate from $\boldsymbol{P}'/{}^yP_j$ via P_i, or else i can manipulate from $\boldsymbol{P}/{}^yP_j$ via P_i', a contradiction to f's strategy-proofness.

(d) *Under no profile can two individuals be pivotal unless each one of them has only two options, and these options are the same for both.*

Suppose otherwise, i.e., $\exists i, j \in I$, distinct $x, y, z \in A$ and $\boldsymbol{P} \in \mathscr{P}^n$ such that $f(\boldsymbol{P}) = x$; $x, y \in o(i, \boldsymbol{P})$; $x, z \in o(j, \boldsymbol{P})$. Consider the profile $\boldsymbol{P}'$ where (1) for all $k \notin \{i, j\}$; $P_k = P_k'$; (2) $P_i' = {}_zP_i$ if xP_iz, or $P_i' = {}^zP_i$ if zP_ix; and (3) $P_j' = {}_yP_j$

if xP_jy, or $P'_j = {}^yP_j$ if yP_jx. In words: P'_i is obtained from P_i by placing z last while preserving the ranking of all other alternatives, when xP_iz; if, on the contrary, zP_ix, then P'_i places z first. Notice that, when zP_ix, z cannot be an option for i, since $f(\boldsymbol{P})=x$ and therefore x is the best option for each individual at $\boldsymbol{P}$. Similar remarks apply to P'_j. Since $P'_i \in r(P_i, x)$ and $P'_j \in r(P_j, x)$, we must have that $f(\boldsymbol{P}')=x$, $o(i, \boldsymbol{P}')=o(i, \boldsymbol{P})$ and $o(j, \boldsymbol{P}')=o(j, \boldsymbol{P})$.

Let now P''_i be defined as follows. If xP_iz (and xP'_iz, by construction), then $P''_i = {}^yP'_i$. If zP_ix (and zP'_ix), then $P''_i = {}^z({}^yP_i)$. That is: when z is the last alternative in P'_i, y is first in P''_i; when z is first in P'_i, then z is first and y is second in P''_i; in both cases, the relative positions of all other alternatives among themselves are the same as in P_i. Similarly, define P''_j so that $P''_j = {}^zP'_j$ if xP_jy, and $P''_j = {}^y({}^zP_j)$ if yP_jx. Observe that $y=C(o(\boldsymbol{P}', i), P''_i)$ since y is ranked either first or second in P''_i, and z cannot be an option for i in $\boldsymbol{P}'$ when P'_i (and thus P''_i), rank z in first place.

Similarly, $z=C(o(\boldsymbol{P}', j), P''_j)$. Therefore, $f(\boldsymbol{P}'/P''_j)=y$, and $f(\boldsymbol{P}'/P''_i)=z$. By construction, $P''_j \in r(y, P'_j)$, and $P''_i \in r(z, P'_i)$. Thus, $f(\boldsymbol{P}'/P''_i)/P''_j = y$, while $f(\boldsymbol{P}'/P''_j)/P''_i = z$, a contradiction, since the arguments of f in both expressions are identical and $z \neq y$.

(e) *If, for some profile $\boldsymbol{P} \in \mathscr{P}^n$, there exists an individual i whose set of options at $\boldsymbol{P}$, $o(i, \boldsymbol{P})$, consists of more than two alternatives, then the range of f equals this set of options, and the individual is a dictator.* By definition, $o(i, \boldsymbol{P})=o(i, \boldsymbol{P}/P'_i)$ for all $P'_i \in \mathscr{P}$. Partition $\mathscr{P}$ into equivalence classes $\mathscr{P}_w (w \in A)$, defined in such a way that $P'_i \in \mathscr{P}_w$ iff $C(o(i, \boldsymbol{P}), P'_i)=w$. By (a), $[P'_i \in \mathscr{P}_w] \to [f(\boldsymbol{P}/P'_i)=w]$. By (d), no $j \neq i$ can be pivotal at $\boldsymbol{P}/P'_i$. Thus, for any $j \in I$, P'_i, $P''_j \in \mathscr{P}$, we have that $o(i, \boldsymbol{P}/P''_j)=o(i, \boldsymbol{P})$, since $f[(\boldsymbol{P}/P''_j)/P'_i]=f[(\boldsymbol{P}/P'_i)/P''_j]=C(o(i, \boldsymbol{P}), P'_i)$. An iterative argument, and the fact that the proposed partition is indeed a partition, completes the proof.

(f) *Let $\boldsymbol{P}$ be a profile where several individuals are pivotal and have the same two-element set of options $\langle x, y\rangle$. Let $\boldsymbol{P}'$ be the profile obtained from $\boldsymbol{P}$ by placing the pair $\langle x, y\rangle$ is the last two places of each individual's ranking while keeping the same order between x and y for each i, as well as preserving the relative positions of all other alternatives among themselves. Then, $f(\boldsymbol{P})=f(\boldsymbol{P}')$.*

Formally, $\boldsymbol{P}'$ is the profile where, for all $i \in I$, $[xP_iy \leftrightarrow xP'_iy] \wedge [(\forall z, w \notin \langle x, y\rangle)(zP_iw \leftrightarrow zP'_iw] \wedge [\forall z \in \langle x, y\rangle)(\forall w \notin \langle x, y\rangle)\, wP'_iz]$. Let $f(\boldsymbol{P})=x$. Let V be the set of individuals which are pivotal at $\boldsymbol{P}$. It suffices to show that, for all i, $f(\boldsymbol{P}/P'_i)=x$, and that V is still the set of individuals which are pivotal at $\boldsymbol{P}/P'_i$. A simple induction argument would then complete the proof. The first part, that $f(\boldsymbol{P}/P'_i)=x$, is clear from (a). The second part, that the set of pivotal individuals cannot change, is the result of the following facts: (1) if i and j are pivotal at $\boldsymbol{P}$, they are also both pivotal at $\boldsymbol{P}/P'_i$; (2) if i is pivotal at P and j is not, then j is not pivotal at $\boldsymbol{P}/P'_i$; (3) if i is not pivotal at $\boldsymbol{P}$ and j is, then j is pivotal at $\boldsymbol{P}/P'_i$; (4) if neither i nor j is pivotal at

$\boldsymbol{P}$, then neither is pivotal at $\boldsymbol{P}/P_i'$. We provide proofs of assertions (1) and (3). Proofs of (2) and (4) result from reversing those provided here. Consider assertion (1). Clearly, i is still pivotal at $\boldsymbol{P}/P_i'$, by definition, and $f(\boldsymbol{P}/P_i')=x$. Since j is pivotal at $\boldsymbol{P}$, $f(\boldsymbol{P}/_xP_j)=y$. If j is no longer pivotal at $(\boldsymbol{P}/P_i')$, then $f(\boldsymbol{P}/P_i')/_xP_j=x$. But then, since xP_iy and, $(\boldsymbol{P}/P_i')/_xP_j=(\boldsymbol{P}/_xP_j)/P_i'$, f is manipulable at $(\boldsymbol{P}/_xP_j)$ by i via P_i'. Now take assertion (3). Since j is pivotal at $\boldsymbol{P}$, $f(\boldsymbol{P}/_xP_j)=y$, and $f(\boldsymbol{P}/_xP_j)/P_i'=y$. Since i is not pivotal at $\boldsymbol{P}$, if j were no longer pivotal at $\boldsymbol{P}/P_i'$ we would have $f(\boldsymbol{P}/P_i')/_xP_j=x$. But this is a contradiction, since $(\boldsymbol{P}/_xP_j)/P_i'=(\boldsymbol{P}/P_i')/_xP_j$.

(g) *If no individual is ever pivotal under f for more than two alternatives, then the range of f consists of at most two alternatives.* Clearly, if no individual is ever pivotal, then f is constant. Suppose thus that one or more individuals are pivotal at some profile $\boldsymbol{P}\in\mathscr{P}^n$. If there are several pivots at $\boldsymbol{P}$, they all must have the same two-element set of options, by (d). Let these options be x and y, and assume $x=f(\boldsymbol{P})$. By (f), we can assume without loss of generality, that these two alternatives are ranked last and second to last by all individuals, in either order. Now suppose there was a third alternative z in the range of f, and let $\boldsymbol{P}'$ be a profile such that $f(\boldsymbol{P}')=z$. Consider the sequence of profiles $\boldsymbol{P}^{(k)}$, $k\in\{1, 2,\dots, n\}$, where $P_i^{(k)}=P_i'$ if $i\leqslant k$, and $P_i^{(k)}=P_i$ if $i>k$. There will exist some $h\in\{1,\dots, n\}$ for which $f(\boldsymbol{P}^{(k)})\notin\langle x, y\rangle$, while $f(\boldsymbol{P}^{(h)})\in\langle x, y\rangle$ for all $k<h$. But then, individual h can manipulate at $\boldsymbol{P}^{(h-1)}$ via P_h', a contradiction.

Let us now put the reasoning together to complete the proof. If f is strategy-proof and its range consists of more than three alternatives, there must be at least one profile where some individual is pivotal for more than two alternatives (by (g)). But then, by (e), this individual is a dictator.

Universidad del País Vasco, Spain

REFERENCES

BARBERÁ, S., "Pivotal Voters: A New Proof of Arrow's Theorem," *Economics Letters*, 6 (December, 1980), 13–16.

GIBBARD, A., "Manipulation of Voting Schemes: A General Result," *Econometrica*, 41 (July, 1973), 587–601.

GREEN, J. AND J. J. LAFFONT, *Incentives in Public Decision Making* (Amsterdam: North Holland, 1978).

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

SCHMEIDLER, D. AND H. SONNENSCHEIN, "Two Proofs of the Gibbard-Satterthwaite Theorem on the Possibility of a Strategy-Proof Social Choice Function," in, H. Gottinger and W. Leinfellner, eds., *Decision Theory and Social Ethics* (Dordrecht: Reidel, 1978).