ELSEVIER

Economics Letters 70 (2001) 381–385

economics letters

www.elsevier.com/locate/econbase

Another direct proof of the Gibbard–Satterthwaite Theorem

Arunava Sen*

Indian Statistical Institute, Delhi Center, 7 S.J.S. Sansanwal Marg, New Delhi 110016, India

Received 25 February 2000; accepted 15 June 2000

Abstract

This paper provides a new and direct proof of the Gibbard–Satterthwaite Theorem based on induction on the number of individuals.

Keywords: Strategy-proof; Gibbard–Satterthwaite Theorem

JEL classification: D71; D82

1. Introduction

The Gibbard–Satterthwaite Theorem (henceforth, the G–S Theorem) is a fundamental result in the theory of incentives. It considers a situation where a collective decision has to be made by a group of individuals regarding the selection of an outcome. The choice of this outcome depends on the preferences that each agent has over the various feasible outcomes. However, these preferences are known only to the agents themselves. The G–S Theorem states that (under mild assumptions) the only procedures which will provide incentives for each individual to report his private information truthfully is one where the responsibility of choosing the outcome is left solely to a single individual (referred to as the dictator).

There are several existing proofs of the G–S Theorem. A natural line of reasoning is to exploit the connection with Arrow's celebrated Impossibility Theorem which also demonstrates the existence of a dictator in a different choice problem. Instances of this approach are Gibbard's original proof (Gibbard (1973)) and the first proof in Schmeidler and Sonnenschein (1978). The main idea here is to construct a social welfare function from a social choice function and then demonstrate that if the latter satisfies the property of strategy-proofness, then the former satisfies the independence of irrelevant alternative axioms. There are also several direct proofs of the G–S Theorem including Satterthwaite's

*Tel.: +91-11-651-4594; fax: +91-11-685-6779.
E-mail address: asen@isid.ac.in (A. Sen).

0165-1765/01/$ – see front matter
PII: S0165-1765(00)00362-1

original proof (Satterthwaite (1975)). A salient approach uses 'option sets', a technique pioneered by Barberà (1983) and Barbera and Peleg (1990). This approach has been shown to be very powerful in characterizing strategy-proof social choice problems in restricted domain environments. Recently, two papers, Benoit (1999) and Reny (1999) have also provided relatively simple proofs of the G–S Theorem. Both these proofs are directly inspired by ideas developed in Geanokoplos (1996) to prove Arrow's Impossibility Theorem.

In this paper, we give another simple and direct proof of the G–S Theorem. The proof relies on induction on the number of individuals. Some of the special features of this argument and comparisons with the other proofs of the same result are discussed in Section 4.

2. Notation

The set $I=\{1,\ldots,N\}$ is the set of individuals. The set of social states or outcomes is a finite set $A=\{a,b,c,d,x,\ldots\}$. Let $\mathscr{P}$ denote the set of all strict orderings (indifference is not permitted) of the elements of A. Each individual i is assumed to have a preference ordering P_i which is an element of the set $\mathscr{P}$. A profile P is an N-tuple $(P_1,P_2,\ldots,P_N)$.

A *Social Choice Function* (SCF) f is a mapping $f:\mathscr{P}^N \to A$.

A SCF f satisfies *unanimity* if for all outcomes a and profiles P such that all individuals i rank a first according to P_i, then $f(P)=a$.

We assume throughout that SCFs under consideration satisfies unanimity.

The SCF f is *manipulable* at profile P by individual i via $\bar{P}_i$ if $f(\bar{P}_i,P_{-i})P_i f(P)$. It is *strategy-proof* if it is not manipulable by any individual at any profile.

The SCF f is *dictatorial* if there exists an individual i (referred to as the *dictator*) such that for all preference profiles P, if $f(P)=a$, then a is first-ranked according to P_i.

The Gibbard–Satterthwaite Theorem. Assume $|A|\geq 3$. Then a SCF $f:\mathscr{P}^N \to A$ is strategy-proof if and only if it is dictatorial.

3. The proof

This proof proceeds by induction on the number of individuals.

Step 1. We show that the theorem holds in the case of two individuals. Let $N=\{1,2\}$ and let f be a strategy-proof SCF. We will show that f either picks 1's first-ranked outcome for all profiles or 2's first-ranked outcome for all profiles.

Claim A. Fix a profile P. Then $f(P)$ must either be the first-ranked outcome according to P_1 or the first-ranked outcome according to P_2.

Proof. Suppose not. Suppose that a and b are the first-ranked outcomes according to P_1 and P_2, respectively, but that $f(P)=c$ where c is distinct from a and b. Note that a and b must also be distinct from each other, otherwise we immediately contradict unanimity. Let $\bar{P}_2$ be an ordering where b is ranked first and a second. Observe that $f(P_1,\bar{P}_2)$ cannot be b because then 2 would manipulate at P via $\bar{P}_2$. Nor can this outcome be any outcome x distinct from a and b. If it were, then a would be strictly better than x according to P_2 and 2 would manipulate via an ordering where a is ranked first. The outcome would then be a by virtue of unanimity. Therefore $f(P_1,\bar{P}_2)=a$.

Let $\bar{P}_1$ be an ordering where a and b are ranked first and second, respectively. We must have $f(\bar{P}_1,\bar{P}_2)=a$ or else 1 will manipulate at $(\bar{P}_1,\bar{P}_2)$ via P_1. Let $f(\bar{P}_1,P_2)=x$. If $x=b$, then 2 manipulates at $(\bar{P}_1,P_2)$ via $\bar{P}_2$. If x is distinct from both a and b, then x is worse than b for 1 according to $\bar{P}_1$. Therefore 1 will manipulate at $\bar{P}_1,P_2$ via an ordering where b is first ranked (this will yield b by virtue of unanimity). Therefore $x=a$. But then 1 manipulates at P via $\bar{P}_1$. □

Claim B. If f picks 1's first-ranked outcome at a profile where 1 and 2's first-ranked outcomes are distinct, then it picks 1's first-ranked outcome at all profiles.

Proof. Let P be a profile where 1 and 2's first-ranked outcomes are a and b, respectively, where a and b are distinct. Holding 2's preferences fixed, observe that the outcome for all profiles where a is first-ranked for 1 must be a, otherwise 1 will manipulate via P_1. Similarly, holding 1's preferences fixed at P_1, observe that 2 can never obtain the outcome b by varying his preference ordering. Now consider any profile where a and b are first-ranked by 1 and 2, respectively. From Claim A, it follows that the outcome must either be a or b. Applying the earlier arguments we conclude that the outcome must, in fact, be a.

Consider an arbitrary outcome c distinct from both a and b. In view of the argument in the previous paragraph, we can assume without loss of generality that c is ranked second in P_1. Let $\bar{P}_1$ be a preference ordering where c is ranked first and a, second. Applying Claim A, it follows that $f(\bar{P}_1,P_2)$ is either b or c. However, if it were b, 1 would manipulate at $(\bar{P}_1,P_2)$ via P_1. Therefore $f(\bar{P}_1,P_2)=c$. Now applying the arguments in the previous paragraph once again, it follows that for any profile, the outcome is 1's first-ranked outcome provided that 2's first-ranked outcome is b.

We now complete the proof of the Claim by showing that the outcome is 1's first-ranked outcome irrespective of 2's first-ranked outcome. Pick an arbitrary outcome x distinct from b and c. We can assume without loss of generality that b and x are ranked first and second, respectively, in P_2. Let $\bar{P}_2$ be an ordering where x is first and b, second. We know that $f(\bar{P}_1,\bar{P}_2)$ must either be c or x. But if it is x then 2 will manipulate at $f(\bar{P}_1,P_2)$ via $\bar{P}_2$. Since x and c were picked arbitrarily, the Claim is established. □

Claim B in conjunction with unanimity are sufficient to complete Step 1.

Step 2. Pick an integer N with $N \geq 3$. We show that statement (a) implies statement (b).

(a) for all K with $K \leq N$, if $f:\mathscr{P}^K \to A$ is strategy-proof, then f is dictatorial.
(b) if $f:\mathscr{P}^N \to A$ is strategy-proof, then f is dictatorial.

Assume that statement (a) holds. Let f be a strategy-proof SCF $f:\mathscr{P}^N \to A$. Define a SCF $g:\mathscr{P}^{N-1} \to A$ as follows. For all $(P_1,P_3,\ldots,P_N) \in \mathscr{P}^{N-1}$, $g(P_1,P_3,P_4,\ldots,P_N)=f(P_1,P_1,P_2,\ldots,P_N)$. In other words, individuals 1 and 2 are 'coalesced'. This 'individual' will be referred to as individual 1 in the SCF g. It follows trivially from the assumption that f satisfies unanimity that g satisfies unanimity as well. We claim that g is strategy-proof. It is clear that individuals 3 through N cannot manipulate in

g because then they can manipulate in f as well. Pick an arbitrary $N-1$ person profile $(P_1,P_3,\dots,P_N)$ and let $g(P_1,P_3,\dots,P_N)=f(P_1,P_1,P_3,\dots,P_N)=a$. Let $\bar{P}_1$ be an arbitrary ordering. Let $f(\bar{P}_1,P_1,P_3,\dots,P_N)=b$ and let $f(\bar{P}_1,\bar{P}_1,P_3,\dots,P_N)=g(\bar{P}_1P_3,\dots,P_N)=c$. Since f is strategy-proof $a \neq b$ implies aP_1b and $b \neq c$ implies bP_1c. Since P_1 is transitive, $a \neq c$ implies aP_1c. Therefore g cannot be manipulated by 1.

Since g satisfies unanimity and is strategy-proof, we can apply statement (a) and conclude that g is dictatorial. There are two cases to consider. Suppose that the dictator, individual j is one of individuals from 3 through N. We claim that j is a dictator in f. Pick an arbitrary profile $(P_1,P_2,P_3,\dots,P_N)$. Let a be first-ranked according to P_j and let $f(P_1,P_2,P_3,\dots,P_N)=b$. Since j dictates in g, 1 can change the outcome from b in the profile $(P_1,P_2,P_3,\dots,P_N)$ to a by announcing P_2. Since f is strategy-proof, we must have aP_1b. Similarly, since $f(P_1,P_1,P_3,\dots,P_N)=a$, we must have aP_1b, or else 2 will manipulate at $(P_1,P_1,P_3,\dots,P_N)$ via P_1. Since P_1 is a strict ordering, we have $a=b$. Therefore j dictates in f.

Finally we need to consider the case where j is individual 1 in g. Pick an arbitrary $N-2$ person profile $(P_3,P_4,\dots,P_N)$. Now define a two-person SCF h as follows: for all pairs of orderings P_1,P_2, $h(P_1,P_2)=f(P_1,P_2,P_3,\dots,P_N)$. Since individual 1 is a dictator in g, it follows that h satisfies unanimity. Moreover, since f is strategy-proof, it follows immediately that h is strategy-proof too. From Step 1, we know that h is strategy-proof, i.e., h is dictatorial. In order to complete the proof, we need only to show that the identity of the dictator does not depend on the $N-2$ profile $(P_3,P_4,\dots,P_N)$. Suppose that it does depend on this profile. Assume without loss of generality that 1 is dictator for $(P_3,P_4,\dots,P_N)$ while 2 is dictator for $(\bar{P}_3,\bar{P}_4,\dots,\bar{P}_N)$. Now progressively change preferences for each individual from 3 through N from the first profile to the second. There must be an individual j with $3 \leq j \leq N$ such that 1 is the dictator in $(\bar{P}_3,\dots,\bar{P}_{j-1},P_j,\dots,P_N)$ while 2 dictates in $(\bar{P}_3,\dots,\bar{P}_{j-1},\bar{P}_j,P_{j+1},\dots,P_N)$. Let a and b be such that aP_jb. Pick P_1 and P_2 such that b and a are first-ranked in P_1 and P_2, respectively. Then, $f(P_1,P_2,\bar{P}_3,\dots,\bar{P}_{j-1},P_j,\dots,P_N)=b$ while $f(P_1,P_2,\bar{P}_3,\dots,\bar{P}_{j-1},\bar{P}_j,P_{j+1},\dots,P_N)=a$. Clearly j will manipulate at $(P_1,P_2,\bar{P}_3,\dots,\bar{P}_{j-1},P_j,\dots,P_N)$ via $\bar{P}_j$. This completes the proof of Step 2. Since the result is trivially true in the case where $N=1$, Steps 1 and 2 complete the proof of the Theorem. □.

4. Concluding remarks

It may be tempting to conclude that Step 1 in the proof is unnecessary. One could reason that since the result is trivial in the case where $N=1$, the induction step, Step 2, is sufficient to establish the Theorem. Unfortunately, this is not true. The induction step is only valid for $N \geq 3$ because its proof explicitly required the result to hold in the case where $N=2$.

Various forms of the induction step have appeared earlier in the literature. One of the earliest references in this regard is Maskin (1978) where a similar argument is employed to characterize domains which admit non-dictatorial Arrovian social welfare functions. A general version of Step 2 can be found in Aswal et al. (1999) which also contains a brief survey of related literature. The arguments in Step 1 can be extended to provide a simple proof of Gibbard's random dictatorship result (Gibbard, 1977)). Details appear in Dutta et al. (2000).

An important feature of the proof of the G–S Theorem presented in this paper is that it does not require all preference orderings to exist in the admissible domain. In fact, all that is required is that for all outcome pairs *a* and *b* there exists an admissible ordering where *a* is first-ranked and *b* second. No restrictions are imposed on the way in which outcomes are ranked 'lower down'.The equivalence between strategy-proofness and dictatorship is thus far more pervasive than is suggested by the G–S Theorem. Further investigation on the domains where this equivalence holds has been undertaken in Aswal et al. (1999). We note that proofs which rely on the link between strategy-proof SCFs and social welfare functions which satisfy Arrovian axioms necessarily require the full strength of the complete domain assumption. The proofs of Reny and Benoit, too, depend critically on the full domain assumption for its simplicity (restrictions are placed on the way outcomes are ranked at the 'bottom' of admissible preference orderings). The Barberà–Peleg proof, on the other hand requires only the 'free pair at the top' assumption.

References

Aswal, N., Chatterji, S., Sen, A., 1999. Dictatorial Domains, CORE Discussion Paper no. 9940.

Barberà, S., 1983. Strategy-proofness and pivotal voters: a direct proof of the Gibbard–Satterthwaite Theorem. International Economic Review 24, 413–417.

Barberà, S., Peleg, B., 1990. Strategy-Proof Voting Schemes with Continuous Preferences. Social Choice and Welfare 7, 31–38.

Benoit, J.-P., 1999. The Gibbard–Satterthwaite Theorem: A Simple Proof, mimeo, New York University.

Dutta, B., Peters, H., Sen, A., 2000. Random dictators in ordinal and cardinal models, in preparation.

Geanokoplos, J., 1996. Three Brief Proofs of Arrow's Impossibility Theorem, mimeo, Cowles Foundation, Yale University.

Gibbard, A., 1973. Manipulation of voting schemes: a general result. Econometrica 41, 587–601.

Gibbard, A., 1977. The manipulation of schemes that mix voting with chance. Econometrica 45, 665–681.

Maskin, E., 1978. Social Welfare Functions on Restricted Domains, mimeo, MIT 1978.

Reny, P., 1999. Arrow's Theorem and the Gibbard–Satterthwaite Theorem: A Unified Approach, mimeo, University of Chicago.

Satterthwaite, M., 1975. Strategyproofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions. Journal of Economic Theory 10, 187–217.

Schmeidler, D., Sonnenschein, H., 1978. Two proofs of the Gibbard–Satterthwaite Theorem on the possibility of a strategy-proof social choice function. In: Gottinger, H.W., Leinfellner, W. (Eds.), Decision Theory and Social Ethics. D. Reidel, Dordrecht.