The Proof of the Gibbard-Satterthwaite Theorem Revisited

Lars-Gunnar Svensson*†

March 8, 1999

Abstract

This paper provides three short and very simple proofs of the classical Gibbard-Satterthwaite theorem. The theorem is first proved in the case with only two individuals in the economy. The many individual case follows then from an induction argument (over the number of individuals). The proof of the theorem is further simplified when the voting rule is assumed to be neutral.

JEL classification: D70

Key words: voting, strategy-proofness

*Financial support from The Bank of Sweden Tercentenary Foundation and from The Swedish Council for Research in the Humanities and Social Sciences is gratefully acknowledged.

†address: Lars-Gunnar Svensson, Deparment of Economics, Lund university, P.O. Box 7082, S-22007 LUND, Sweden; e-mail: Lars-Gunnar.Svensson@nek.lu.se; telephone number: +46 46 222 86 79; fax number +46 46 222 46 13

1 Introduction

The objective of this paper is to present short and very simple proofs of the classical *Gibbard - Satterthwaite theorem* (Gibbard (1973), Satterthwaite (1975)). Among the large number of existing proofs, one can distinguish two different lines of proofs. The first one, and the common one, is based on Arrow's impossibility theorem and uses the correspondence between strategy-proofness and the independence of irrelevant alternatives requirement. The other approach to prove the theorem is more direct and can be found in e.g. Barberá (1983) and in Barberá and Peleg (1990). In particular, the paper of Barberá and Peleg provides a short and elegant proof of the theorem in the case with two individuals in the economy.

Here we also employ the direct approach and first prove the theorem when there are two individuals. The proof and the arguments in this paper differ from the proof in Barberá and Peleg in the two individual case, and in addition, we follow up with a simple induction proof in the many (but finite) individual case.[1] The proof is further simplified when the voting rule is assumed to be neutral.

The paper is organized as follows. The voting model is presented in section 2, some useful lemmas are given in section 3, while the two individual case and the many individual case were neutrality is assumed, are to be find in section 4. In an appendix an induction proof for the many individual case is given. In addition, the proof in the two individual case along the lines in Barberá and Peleg (1990) are given there.

2 The voting model

The voting model is well-known so we will only give a short description of the basic elements. Let $N = \{1, 2, \ldots, n\}$ be a finite set of *individuals* and let $A = \{a_1, a_2, \ldots a_m\}$ be a finite set of *alternatives*. *Preferences* over A are rankings of the various elements (i.e. complete, transitive, asymmetric binary relations[2]). The set of all possible rankings of A represented by utility functions is denoted U, so for $a, b \in A$ with $a \neq b$ and $u \in U$, $u(a) > u(b)$ or $u(b) > u(a)$, but not both. *Preference profiles* are elements in $\mathcal{U} = U^n$. A preference profile $u = (u_1, u_2, \ldots u_n)$ can also be denoted (u_i, u_{-i}) for $i \in N$.

[1] In the many individual case Barberá and Peleg have more complicated arguments than those we use in our induction proof, but on the other hand they also prove the theorem under more general assumptions, e.g. the number of alternatives may be infinite.

[2] An alternative here is to assume that the individual preferences are complete and transitive binary relations, and hence allow for indifferences. Our more narrow class of preferences is, however, motivated by the desire to get short and simple proofs.

A *voting rule* is a mapping f from in $\mathcal{U}$ to A. A voting rule f is *manipulable* precisely when there is an individual $i \in N$, preferences $u_i' \in U$, and a preference profile $u \in \mathcal{U}$, such that $u_i(f(u_i', u_{-i})) > u_i(f(u))$. If a voting rule is not manipulable, it is *strategy-proof*. A voting rule f is *dictatorial* if there is an individual $i \in N$ (the dictator) such that $u_i(f(u)) \geq u_i(a)$ for all $a \in A$ and for all $u \in \mathcal{U}$.

Let π be a permutation of A - a change of names of the alternatives[3]. If $u \in \mathcal{U}$ is a preference profile, a preference profile πu is defined as $(\pi u)_i(a) = u_i(\pi^{-1}(a))$ for $i \in N$ and $a \in A$. A voting rule f is *neutral* if, for all preference profiles $u \in \mathcal{U}$, $f(\pi u) = \pi f(u)$. This means that the outcome of a neutral voting rule is independent of the names of the alternatives.

Finally, a voting rule f is *onto* if for each $a \in A$ there is a preference profile $u \in \mathcal{U}$ such that $f(u) = a$. Obviously, a neutral voting rule is always onto.

The theorem of this paper is:

The Gibbard-Satterthwaite theorem: *A strategy-proof voting rule that is onto is dictatorial if the number of alternatives is at least three.*

3 Some useful lemmas

For the proof of the main theorem two useful and simple lemmas will be employed. The first one is a monotonicity lemma. It is a well-known result and has been used in various forms frequently before, but the proof is very simple and we repeat it for completeness. The lemma says that a strategy-proof voting rule is constant for all changes of reported preferences such that alternatives worse than the outcome alternative (a) before the change are also worse than a after the change. More exactly:

Lemma 1 *(monotonicity) Let f be a strategy-proof voting rule, $u \in \mathcal{U}$ a preference profile, and $f(u) = a$. Then $f(v) = a$ for all preference profiles $v \in \mathcal{U}$ such that for all $x \in A$ and $i \in N$,*

$$v_i(a) \geq v_i(x) \text{ if } u_i(a) \geq u_i(x).$$

Proof. Suppose first that $v_i = u_i$ if $i > 1$. Let $f(v_1, u_{-1}) = b$. From strategy-proofness follows that $u_1(b) \leq u_1(a)$, and hence from the assumption of the lemma, $v_1(b) \leq v_1(a)$. Strategy-proofness also implies that $v_1(b) \geq v_1(a)$ and then, because preferences are strict, $a = b$. The lemma now follows after

[3]Strictly speaking, π is a permutation of the set $\{1, 2, \ldots m\}$ of indices of the elements in A. However, for $a_j \in A$ we will write $\pi(a_j)$ instead of $a_{\pi(j)}$.

repeating the arguments above while changing the preferences for only $i = 2$, and then for only $i = 3$, and so forth. ■

The second lemma says that the outcome of a strategy-proof and onto voting rule must be (weakly) Pareto optimal.

Lemma 2 *(Pareto optimality) Let f be a strategy-proof voting rule that is onto. If $u \in \mathcal{U}$ is a preference profile and $a, b \in A$, $a \neq b$, two alternatives such that $u_i(a) > u_i(b)$ for all $i \in N$, then $f(u) \neq b$.*

Proof. Suppose that $f(u) = b$. Since f is onto there is a preference profile $v \in \mathcal{U}$ such that $f(v) = a$. Let also $u' \in \mathcal{U}$ be a preference profile such that for all $i \in N$,

$$u'_i(a) > u'_i(b) > u'_i(x) \text{ and } u'_i(x) = u_i(x) \text{ for } x \in A - \{a, b\}.$$

By monotonicity (Lemma 1) it now follows that $b = f(u) = f(u')$ and $a = f(v) = f(u')$, which is a contradiction. Hence $f(u) \neq b$. ■

4 The proof of the theorem under simplifying assumptions

4.1 $n = 2$

First assume that the number of individuals is two. The following example illustrates the idea in the proof of the theorem in that case.

Example. Consider a situation with two individuals and three alternatives, a, b, c. To prove the theorem we first want to identify a dictator. Therefore consider a preference profile where both individuals consider the alternatives a and b to be better than c, but they have different highest ranked alternative. The preferences are illustrated in the matrix u below where individual i's ranking is given in column i, $i = 1, 2$.

A Pareto consistent voting rule cannot select the alternative c (Lemma 2). Assume that the outcome (bold) of the voting rule is a when the preference profile is u. Then consider the preference profile v below.

$$u = \begin{pmatrix} \mathbf{a} & b \\ b & \mathbf{a} \\ c & c \end{pmatrix} \quad v = \begin{pmatrix} \mathbf{a} & b \\ b & c \\ c & \mathbf{a} \end{pmatrix}$$

The outcome in this case must also be a. The reason is that the outcome must be a or b by Pareto consistency but it cannot be b because of strategy-proofness. But given a preference profile where the outcome (a) is the best

alternative for one individual and the worst alternative for the other individual, it follows from monotonicity (Lemma 1) that the outcome of the voting rule is always a when the first individual reports a to be his best alternative. He becomes the dictator for this alternative.

In the formal proof below we show that each alternative has a dictator and then it must, of course, be the same individual for all alternatives.

Theorem 1 *A strategy-proof voting rule f that is onto is dictatorial if the number of alternatives is at least three and the number of individuals is two.*

Proof. Let $u \in U^2$ be a preference profile and $a, b \in A$ two alternatives with $a \neq b$ such that

$$u_1(a) > u_1(b) > u_1(x) \text{ and } u_2(b) > u_2(a) > u_2(x)$$

for all $x \in A - \{a, b\}$. Then $f(u) = a$ or b by Pareto optimality (Lemma 2). Suppose that $f(u) = a$.

Now let preferences $v_2 \in U$ satisfy

$$v_2(b) > v_2(x) > v_2(a)$$

for all $x \in A - \{a, b\}$. Then $f(u_1, v_2) = a$ or b by Pareto optimality (Lemma 2) and $f(u_1, v_2) \neq b$ by strategy-proofness. Hence $f(u_1, v_2) = a$.

Monotonicity (Lemma 1) now implies that $f(u') = a$ for all $u' \in U^2$ such that a is the best alternative according to preferences u'_1.

By repeating the analysis above for *all* pairs of alternatives from A we receive two sets $A_i \subset A$, $i = 1, 2$, where A_i contains those alternatives x such that the outcome of f is always x if individual i reports x to be his best alternative.

Let $A_3 = A - (A_1 \cup A_2)$. Obviously $\#A_3 \leq 1$. Since the voting rule is a function and $m \geq 3$, one of the sets A_1 and A_2 must be empty. We have assumed that $a \in A_1$ so $A_2 = \emptyset$. Finally $A_3 = \emptyset$, because if $c \in A_3$, let $u \in U^2$ be a preference profile such that

$$u_1(c) > u_1(a) > u_1(x) \text{ and } u_2(a) > u_2(c) > u_2(x)$$

for all $x \in A - \{a, c\}$. By the arguments above, $c \in A_1$ or $a \in A_2$, which is a contradiction. Hence $A_1 = A$ and $i = 1$ is a dictator.

If we had assumed that $a \in A_2$ at the beginning of the proof then individual 2 had become a dictator. ■

4.2 A neutral voting rule and $m \geq n$.

In the appendix we provide a short induction proof of the theorem in the general case with an arbitrary but finite number of individuals. To get a simple and direct proof of the theorem without induction based on the two individual case, we will here make two additional and simplifying assumptions. The first one is to assume that there are at least as many alternatives as there are individuals, i.e. $m \geq n$. Second we will assume neutrality, i.e. that the outcome of the voting rule is independent of the names of the alternatives. We may note that a neutral voting rule trivially is also onto, so Lemma 2 is still valid.[4]

To illustrate the simple idea in the proof - which is a generalization of the idea in the proof of Theorem 1 - consider the following example.

Example. Consider four individuals and five alternatives, a, b, c, d, e. The main step in the proof is to identify one individual and one alternative such that if the individual reports that particular alternative to be his best alternative the alternative will also be the outcome of the voting rule. Then, by neutrality, that individual will be a dictator.

Therefore consider a preference profile where the various individuals have permuted rankings of four alternatives a, b, c, d, but all consider e to be the worst alternative. The preferences are illustrated in the matrix u below. Here the ranking of individual i is given in column i, $i = 1, 2, 3, 4$.

$$u = \begin{pmatrix} \mathbf{a} & b & c & d \\ b & c & d & \mathbf{a} \\ c & d & \mathbf{a} & b \\ d & \mathbf{a} & b & c \\ e & e & e & e \end{pmatrix}$$

A Pareto consistent voting rule cannot select the alternative e (Lemma 2). Assume that the outcome (bold) of the voting rule is a when the preference profile is u. Then consider the preference profile u^1 below.

$$u^1 = \begin{pmatrix} \mathbf{a} & d & d & d \\ d & \mathbf{a} & \mathbf{a} & \mathbf{a} \\ b & b & b & b \\ c & c & c & c \\ e & e & e & e \end{pmatrix}$$

[4]We may also note that the preceding proof of the theorem presupposes that there are at least three alternatives ($m \geq 3$). However, by assuming neutrality that particular requirement is superfluous.

All alternatives in the profile u^1 ranked higher than a are also ranked higher than a in the profile u. Then by monotonicity (Lemma 1), the outcome of the voting rule in this case must also be a. Next consider the preference profiles u^2, u^3 and u^4 below.

$$u^2 = \begin{pmatrix} \mathbf{a} & d & d & d \\ d & b & \mathbf{a} & \mathbf{a} \\ b & c & b & b \\ c & e & c & c \\ e & \mathbf{a} & e & e \end{pmatrix}, u^3 = \begin{pmatrix} \mathbf{a} & d & d & d \\ d & b & b & \mathbf{a} \\ b & c & c & b \\ c & e & e & c \\ e & \mathbf{a} & \mathbf{a} & e \end{pmatrix}, u^4 = \begin{pmatrix} \mathbf{a} & d & d & d \\ d & b & b & b \\ b & c & c & c \\ c & e & e & e \\ e & \mathbf{a} & \mathbf{a} & \mathbf{a} \end{pmatrix}.$$

By Pareto consistency the outcome of the voting rule given the preference profile u^2 cannot be b, c or e because they are dominated by the alternative d. Moreover, strategy-proofness excludes d to be the outcome and hence a is still the outcome. The same arguments show that a must be the outcome given the profiles u^3 and u^4. But now follows immediately from monotonicity (Lemma 1) that the outcome of the voting rule is a as soon as individual 1 reports a to be his best alternative. Finally, neutrality shows that individual 1 is a dictator.

In the theorem below these ideas are formalized.

Theorem 2 *A strategy-proof voting rule f that is neutral is dictatorial if the number of alternatives is at least three and at least as many as there are individuals, $m \geq n$.*

Proof. Let $u \in \mathcal{U}$ be defined as

$$\begin{aligned} u_i(a_j) &= n + i - j \text{ if } i \leq j \leq n, \\ u_i(a_j) &= i - j \text{ if } j < i, \\ u_i(a_j) &= n - j \text{ if } j > n. \end{aligned}$$

This means that the various individual rankings of the n first alternatives are permuted. Those alternatives are also ranked before alternatives with label $j > n$. By Pareto consistency (Lemma 2), $f(u) = a_j$ for some $j \leq n$. Assume that $j = 1$. Let $u' \in \mathcal{U}$ be defined as

$$\begin{aligned} u'_1(a_1) &= n + 2 \text{ and } u'_1(a_n) = n + 1, \\ u'_i(a_n) &= n + 2 \text{ and } u'_i(a_1) = n + 1 \text{ for } i > 1, \\ u'_i(a_j) &= u_i(a_j) \text{ for } j \neq 1 \text{ and } j \neq n. \end{aligned}$$

Hence all individuals consider the alternatives a_1 and a_n to be better than the other alternatives. Also note that the ranking of a_1 and a_n is the same

in the profiles u and u', and in u',the alternatives a_1 and a_n are both ranked before the other alternatives. Hence by monotonicity (Lemma 1), $f(u') = f(u) = a_1$. Finally define profiles $u^k \in \mathcal{U}$, $k = 1, 2, 3, \ldots n-1$ recursively according to

$$\begin{aligned} u^1 &= u', \\ u_i^{k+1} &= u_i^k \text{ for } i \neq k+1, \\ u_{k+1}^{k+1}(x) &= u_{k+1}^k(x) \text{ for } x \in A - \{a_1\}, \\ u_{k+1}^{k+1}(a_1) &= -m. \end{aligned}$$

By Pareto consistency $f(u^k) = a_1$ or $f(u^k) = a_n$. But strategy-proofness implies that $f(u^{k+1}) = f(u^k)$ and hence $f(u^n) = a_1$. In the utility profile u^n, a_1 is the best alternative for individual 1, while it is the worst alternative for all other individuals. Monotonicity (Lemma 1) then implies that $f(u) = a_1$ as soon as individual 1 reports a_1 to be his best alternative. Then by neutrality, individual 1 becomes a dictator. ■

5 Appendix

The following is an *alternative* proof of Theorem 1. The proof is almost the same as the one in Barberá and Peleg (1990). It is a simple proof but a rather different type of arguments are used here compared to the first proof of Theorem 1.

For given preferences $u_1 \in U$, let

$$\alpha(u_1) = \{a \in A;\ a = f(u_1, u_2) \text{ for some preferences } u_2 \in U\},$$

i.e. $\alpha(u_1)$ is the range of f for fixed preferences u_1. Obviously the set $\alpha(u_1)$ is the *choice set* of individual 2 when the preferences of individual 1 are u_1. When individual 2 has reported his preferences u_2, a strategy-proof voting rule f requires that the outcome is the best alternative in $\alpha(u_1)$ according to u_2.

The set $\alpha(u_1)$ will play a central role in the proof of the theorem. The basic idea is to prove that $\alpha(u_1)$ contains exactly one alternative for all preferences u_1 implying that individual 1 is a dictator, or that $\alpha(u_1) = A$ for all preferences u_2 implying that individual 2 is a dictator. The properties of $\alpha(u_1)$ are revealed by a number of small lemmas.

Lemma 3 *If a is the best alternative in A according to $u_1 \in U$, then $a \in \alpha(u_1)$.*

Proof. Follows directly from Pareto optimality (Lemma 2). ■

Lemma 4 *If a is the best alternative in A according to $u_1 \in U$ as well as to $u_1' \in U$, then $\alpha(u_1) = \alpha(u_1')$.*

Proof. Suppose that $b \in \alpha(u_1)$ but $b \notin \alpha(u_1')$. Let $u_2 \in U$ be such that $u_2(b) > u_2(a) > u_2(x)$ for all $x \in A - \{a, b\}$. By Lemma 3, $a \in \alpha(u_1')$ so $f(u_1', u_2) = a$. But $f(u_1, u_2) = b$, so f is manipulable - a contradiction. Hence $\alpha(u_1) \subset \alpha(u_1')$ and hence by a symmetry argument, $\alpha(u_1) = \alpha(u_1')$. ■

Lemma 5 *If a is the best and w the worst alternative in $\alpha(u_1)$ according to preferences $u_1 \in U$, then $b \in \alpha(u_1)$ if $u_1(a) > u_1(b) > u_1(w)$.*

Proof. Suppose that $b \notin \alpha(u_1)$. Let $u_1' \in U$ be such that $b \in \alpha(u_1')$. Such preferences exists since f is onto. Now let $u_2 \in U$ be preferences such that $u_2(b) > u_2(w) > u_2(x)$ for all $x \in A - \{w, b\}$. But then by strategy-proofness, $f(u_1, u_2) = w$ and $f(u_1', u_2) = b$, so f is manipulable - a contradiction. ■

Lemma 6 *If $\#\alpha(u_1) > 1$ for some preferences $u_1 \in U$, then $\alpha(u_1) = A$.*

Proof. Let a be the best and w the worst alternative in $\alpha(u_1)$ according to u_1. Suppose that there is an alternative $b \in A - \alpha(u_1)$. Let $u_1' \in U$ be preferences such that $u_1'(a) > u_1'(b) > u_1'(w)$ and such that a is the best alternative in A according to u_1'. Then by Lemma 4, $\alpha(u_1) = \alpha(u_1')$. But by Lemma 5, $b \in \alpha(u_1')$ and hence $b \in \alpha(u_1)$. This is a contradiction, so $\alpha(u_1) = A$. ■

Lemma 7 *If $\#\alpha(u_1') > 1$ for some preferences $u_1' \in U$, then $\alpha(u_1) = A$ for all preferences $u_1 \in U$.*

Proof. We note first that $\alpha(u_1') = A$ by Lemma 6. Suppose that there are preferences $u_1'' \in U$ such that $\alpha(u_1'') = \{a\}$, i.e. contains just one alternative. Let w be the worst alternative in $\alpha(u_1')$, i.e. in A, according to preferences u_1'. By Lemma 4 we can assume that $a \neq w$ without loss of generality. Also let $u_2 \in U$ be preferences such that $u_2(w) > u_2(a) > u_2(x)$ for all $x \in A - \{a, w\}$. Then $f(u_1', u_2) = w$ and $f(u_1'', u_2) = a$, and hence f is manipulable - a contradiction. ■

The alternative proof of Theorem 1: If $\#\alpha(u_1) = 1$ for all preferences $u_1 \in U$ then by Lemma 3 individual 1 is a dictator, and if $\#\alpha(u_1) > 1$ for some preferences $u_1 \in U$ then by Lemma 7, $\alpha(u_1) = A$ for all preferences, and hence individual 2 is a dictator. ■

Now let the number of individuals be any finite number. The main theorem then follows from an induction argument.

Theorem 3 *A strategy-proof voting rule f that is onto is dictatorial if the number of alternatives is at least three.*

Proof. Assume that the theorem is true for p individuals, $p < n$. We shall prove that it is also true for $p+1$ individuals. Since the theorem is true for $p = 2$ by Theorem 1, it then follows by induction that the theorem is true for n.

Let g be a voting rule with two individuals defined as

$$g(u_1, v) = f(u_1, v \dots v).$$

The voting rule g is onto by Pareto optimality (Lemma 2) since f is onto. The voting rule g is also strategy-proof. Because if it is not, there are profiles (u_1, v) and (u_1, v') in U^2, and alternatives $a, b \in A$ such that $g(u_1, v) = a$ and $g(u_1, v') = b$ and $v(b) > v(a)$. Let $u^k = (u_1, v', \dots v', v, \dots v) \in U^{p+1}$ contain k v':s and $p - k$ v:s, $0 \leq k \leq p$. Also let $a^k = f(u^k)$. The voting rule f is manipulable if $v(a^{k+1}) > v(a^k)$ for some $k < p$. But that must be the case because $v(a^p) = v(b) > v(a) = v(a^0)$. Hence we have a contradiction and the voting rule g must be strategy-proof. By Theorem 1**,** it is then dictatorial.

First, if $i = 1$ is the dictator for g then by Lemma 1 he is also dictator for f.

Second, if $i = 2$ is the dictator for g, let $u_1^* \in U$ be fixed preferences and consider the voting rule

$$h(u_2 \dots u_{p+1}) = f(u_1^*, u_2 \dots u_{p+1})$$

with p individuals. The voting rule h is strategy-proof, and it is onto because $i = 2$ is a dictator for g. Then, by the induction assumption, h is dictatorial. Suppose that $i = 2$ is the dictator for h and consider the voting rule

$$q(u_1, u_2) = f(u_1, u_2, u_3^* \dots u_{p+1}^*)$$

for arbitrarily fixed preferences $u_i^* \in U$, $i \geq 3$. The voting rule q is strategy-proof and onto, and hence dictatorial. But $i = 1$ cannot be the dictator so it must be $i = 2$. Since $u_i^* \in U$ was arbitrarily chosen $i = 2$ is the dictator for f for all preference profiles in U^{p+1}. ■

References

[1] Barberá, S., 1983, Strategy-proofness and pivotal voters: a direct proof of the Gibbard-Satterthwaite theorem, *International Economic Review* 24, 413 - 417.

[2] Barberá, S. and B. Peleg, 1990, Strategy-Proof Voting Schemes with Continuous Preferences, *Social Choice and Welfare* 7, 31 - 38.

[3] Gibbard, A., 1973, Manipulation of Voting Schemes: A General Result, *Econometrica* 41, 587 - 601.

[4] Satterthwaite, M., 1975, Strategy-proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions, *Journal of Economic Theory* 10, 187 - 216.