No.384

An Elementary Proof of the Gibbard-Satterthwaite Theorem: The Three Alternative and Two Participant Case

by

Tatsuyoshi Saijo*
Institute of Socio-Economic Planning
University of Tsukuba
Tsukuba, Ibaraki 305
Japan

September 1988

* The author wishes to thank Mark Bates and Steve LeRoy for their comments.

Abstract

This pedagogic paper gives a simple proof of the Gibbard-Satterthwaite theorem for the three alternative and two participant case appealing not only to Pareto efficiency but also to monotonicity of a social choice function.

1. Introduction

Suppose that there are two participants and three alternatives. Suppose further that each participant can choose preferences out of all six possible preferences that do not allow indifference between two alternatives, and any alternative is socially attainable.[1] Under these assumptions, Feldman (1979) ingeniously provides an elementary proof of the Gibbard-Satterthwaite theorem, which shows the strategy-proofness of a social choice function implies dictatorship.[2] In his proof, he extensively uses the property that strategy-proofness implies Pareto efficiency. On the other hand, Muller-Satterthwaite (1977) show that strategy-proofness is equivalent to monotonicity [see also Maskin (1977)] of the social choice function providing that no indifference is allowed.[3] Since Pareto efficient social choice functions are not necessarily monotonic, there is an apparent gap between monotonicity and Pareto efficiency. Using monotonicity together with Pareto efficiency, this paper presents a further simplification and clarification of Feldman's proof.

2. Notation and Definitions

Suppose that there are two participants denoted 1 and 2, and that there are three alternatives, say x, y, and z. The alternative may be thought of as candidates in an election or public projects for a community. Assuming that no participant is ever indifferent between any two distinct alternatives, we have the following six possible preferences on the alternative set:

1 $x > y > z$

2 $x > z > y$

3 $y > x > z$

4 $y > z > x$

5 $z > x > y$

6 $z > y > x$,

where ">" represents "is strictly preferred to." To simplify the notation, for example, preferences $y > x > z$ is denoted *3*. Each participant reports his/her preferences out of six possibilities, whether they are true or not, to the center, and the center computes a socially optimal alternative based upon the preference reports. The socially optimal alternative is computed from a social choice function f. The input of this function is a pair of preferences, for example, (*3*,*5*), where the first element is participant 1's preferences and the second is participant 2's. The output is an alternative. For example, $f(3,5) = x$ (see also Tables 1 and 2). We shall write participant i's preference p_i so that in general, we write $a = f(p_1,p_2)$, where a is x, y, or z. We shall write $a(>,p_1)b$ if a is strictly preferred to b by p_1, and similarly write $a(\geq,p_2)b$ if $a(>,p_2)b$ or a is indifferent to b by p_2. Of course, if they are indifferent, $a = b$ by assumption. We assume that participants 1 and 2 have already agreed upon the shape of the social choice function and that it is public knowledge.

A social choice function f satisfies *strategy-proofness* if for a true preference pair $(\hat{p}_1,\hat{p}_2)$, the following conditions are satisfied:

$$(1)\quad \begin{cases} f(\hat{p}_1,\hat{p}_2)(\geq,\hat{p}_1)f(p_1,\hat{p}_2) \text{ for all six possible } p_1 \text{ ; and} \\ f(\hat{p}_1,\hat{p}_2)(\geq,\hat{p}_2)f(\hat{p}_1,p_2) \text{ for all six possible } p_2. \end{cases}$$[4]

Allowing no indifference, Muller-Satterthwaite prove that strategy-proofness is equivalent to monotonicity. The social choice function satisfies *monotonicity* if for any pairs of preferences $(\bar{p}_1,\bar{p}_2)$ and $(\tilde{p}_1,\tilde{p}_2)$ with $a = f(\bar{p}_1,\bar{p}_2)$,

if for all b, $a(\geq,\bar{p}_i)b$ implies $a(\geq,\tilde{p}_i)b$ $(i=1,2)$, then $a = f(\tilde{p}_1,\tilde{p}_2)$.

Suppose that an alternative a is socially optimal under $(\bar{p}_1,\bar{p}_2)$. Consider now a new preference profile $(\tilde{p}_1,\tilde{p}_2)$. Assume that the alternatives which are not preferred or indifferent to a by $\bar{p}_i$ $(i=1,2)$ are again not preferred or indifferent to a by $\tilde{p}_i$ $(i=1,2)$. Hence, every participant's ranking of a goes up or stays the same by the preference change. Since every participant says that the ranking of a is the same as before or better than before, and a is socially optimal under the previous preference profile, a should be socially optimal under a new preference profile, i.e., $a = f(\tilde{p}_1,\tilde{p}_2)$.

3. Gibbard-Satterthwaite Theorem

We shall prove a simplified version of the Gibbard-Satterthwaite theorem with three steps.

Gibbard-Satterthwaite Theorem [Gibbard (1973), Satterthwaite (1975)]: Assume that there are two participants and three alternatives, each alternative is attainable by the social choice function, and each preference relation is one of the six possibilities. If the social choice function satisfies strategy-proofness, it must be dictatorial.

Step 1: Strategy-proofness implies monotonicity.[5]

Choose any two preference profiles $(\bar{p}_1,\bar{p}_2)$ and $(\breve{p}_1,\breve{p}_2)$ such that a = $f(\bar{p}_1,\bar{p}_2)$, and for all c, $a(\geq,\bar{p}_i)c$ implies $a(\geq,\breve{p}_i)c$ (i=1,2). We shall show that $f(\breve{p}_1,\bar{p}_2)$ = a, and then show that $f(\breve{p}_1,\breve{p}_2)$ = a. Since (1) is satisfied when $(\hat{p}_1,\hat{p}_2) = (\bar{p}_1,\bar{p}_2)$, i.e., $(\bar{p}_1,\bar{p}_2)$ is a true preference profile, $f(\breve{p}_1,\bar{p}_2)$ is available and hence (2) $a(\geq,\bar{p}_1)f(\breve{p}_1,\bar{p}_2)$. Similarly since (1) is satisfied when $(\hat{p}_1,\hat{p}_2) = (\breve{p}_1,\bar{p}_2)$, a is available and hence

$$f(\breve{p}_1,\bar{p}_2)(\geq,\breve{p}_1)a.$$

By the presumption of monotonicity, (2) implies

$$a(\geq,\breve{p}_1)f(\breve{p}_1,\bar{p}_2).$$

Since no indifference is allowed, $f(\breve{p}_1,\bar{p}_2)$ = a. A similar argument will be applied to between $f(\breve{p}_1,\bar{p}_2)$ and $f(\breve{p}_1,\breve{p}_2)$.[6] Therefore, we have $f(\breve{p}_1,\breve{p}_2)$ = a. I.e., f is monotonic.

Step 2: Monotonicity implies Pareto efficiency.

We must show that for each (p_1,p_2), there is no other alternative b such that $b(>,p_i)f(p_1,p_2)$ for all i. Suppose, toward a contradiction, that there is a pair of preferences (p_1,p_2) whose output $f(p_1,p_2)$ is not Pareto efficient. That is, there is b such that $b(>,p_i)f(p_1,p_2) = a$ for all i. Then a is *at most* ranked the second for i = 1,2. Now construct a new pair of preferences (p'_1,p'_2) such that b is top-ranked and a second. Since f is monotonic, $f(p_1,p_2) = a$, and for all c, $a(\geq,p_i)c$ implies $a(\geq,p'_i)c$ for all i, we have $f(p'_1,p'_2) = a$. On the other hand, since we assume that any alternative is attainable, there is a pair of preferences (p''_1,p''_2) with $f(p''_1,p''_2) = b$. Again, by monotonicity, $f(p'_1,p'_2) = b$, which contradicts $f(p'_1,p'_2) = a \neq b$. Therefore, f is Pareto efficient.

Step 3: the social choice function is dictatorial.

See Table 1. Each cell contains some alternatives which are the candidates of the socially optimal alternative with the fact that the social choice function is Pareto efficient. We shall consider two cases. Since $f(1,4)$ is either x or y, first suppose $f(1,4) = x$. Since f satisfies monotonicity, the cells of the first and second rows should be x.

Consider the sixth column. We shall show $f(3,6) = y$. Suppose not: $f(3,6) = z$. Participant 1 could be better off by changing his preferences from 3 to 2, i.e., $x(>,3)z$, which contradicts the strategy-proofness. Since $f(3,6) = y$, by monotonicity, $f(4,6) = y$. We shall now show $f(4,5) =$

y. Suppose not: $f(4,5) = z$. When the true preference pair is $(4,6)$, participant 2 could be better off by changing his preferences from 6 to 5, i.e., $z(>,6)y$, which contradicts (1). Since $f(4,5) = y$ and f is monotonic, the cells of the third and fourth rows should be y.

The rest of the argument is the same as the above one. We shall show that $f(6,2) = z$. Suppose not: $f(6,2) = x$. Then participant 1 could be better off by changing his strategies from 6 to 4, which contradicts (1). Since $f(6,2) = z$ and f is monotonic, $f(5,2) = z$. We shall show that $f(5,1) = z$. If not, then when the true preference pair is $(5,2)$, participant 2 could be better off by changing from 2 to 1, which contradicts (1) again. Hence $f(5,1) = z$ and monotonicity together imply that the cells of the fifth and sixth rows should be z.

Now Table 1 will become Table 2. The socially optimal alternatives are exactly the same as participant 1's top-ranked alternatives. That is, participant 1 is a dictator. It should now be clear that if we assume $f(1,4) = y$, then participant 2 is a dictator.

Participant 1 \ Participant 2	1: x > y > z	2: x > z > y	3: y > x > z	4: y > z > x	5: z > x > y	6: z > y > x
1 x > y > z	x	x	x,y	x,y	x,z	x,y,z
2 x > z > y	x	x	x,y	x,y,z	x,z	x,z
3 y > x > z	x,y	x,y	y	y	x,y,z	y,z
4 y > z > x	x,y	x,y,z	y	y	y,z	y,z
5 z > x > y	x,z	x,z	x,y,z	y,z	z	z
6 z > y > x	x,y,z	x,z	y,z	y,z	z	z

Table 1: Possible alternatives of the Pareto efficient social choice function.

Participant 1 \ Participant 2	1: x > y > z	2: x > z > y	3: y > x > z	4: y > z > x	5: z > x > y	6: z > y > x
1 x > y > z	x	x	x	x	x	x
2 x > z > y	x	x	x	x	x	x
3 y > x > z	y	y	y	y	y	y
4 y > z > x	y	y	y	y	y	y
5 z > x > y	z	z	z	z	z	z
6 z > y > x	z	z	z	z	z	z

Table 2: If $f(1,4) = x$, then participant 1 is a dictator.

Footnotes

1. This condition, which says that the social choice function is onto, is called *citizen sovereignty.*

2. Kelly (1987) also follows Feldman's proof.

3. Monotonicity is equivalent to Muller-Satterthwaite's *strong positive association.* We follow Maskin's terminology (1977).

4. This definition says that reporting preference truthfully by every participant is a Nash equilibrium. This is equivalent to say that reporting preference truthfully is a dominant strategy for every participant [see Dasgupta-Hammond-Maskin's Theorem 7.1.1].

5. Step 1's proof is slightly different from Muller-Satterthwaite. Here, we directly prove the claim instead of using the method of contradiction.

6. Since (1) is satisfied when $(\hat{p}_1,\hat{p}_2) = (\tilde{p}_1,\bar{p}_2)$, i.e., $(\tilde{p}_1,\bar{p}_2)$ is a true preference profile, $f(\tilde{p}_1,\tilde{p}_2)$ is available and $a(\geq,\bar{p}_2)f(\tilde{p}_1,\tilde{p}_2)$. Similarly since (1) is satisfied when $(\hat{p}_1,\hat{p}_2) = (\tilde{p}_1,\tilde{p}_2)$, a is available and $f(\tilde{p}_1,\tilde{p}_2)(\geq,\tilde{p}_2)a$. By the presumption of monotonicity, $a(\geq,\bar{p}_2)f(\tilde{p}_1,\tilde{p}_2)$ implies $a(\geq,\tilde{p}_2)f(\tilde{p}_1,\tilde{p}_2)$. Since no indifference is allowed, $f(\tilde{p}_1,\tilde{p}_2) = a$.

References

Partha Dasgupta, Peter Hammond and Eric Maskin (1979), "The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility," *Review of Economic Studies* Vol.46, No.143, pp.181-216.

Allan Feldman (1979), "Manipulating Voting Procedures," *Economic Inquiry* Vol.17, pp.452-74.

Allan Gibbard (1973), "Manipulation of Voting Schemes: A General Result," *Econometrica* 41, pp.587-601.

Jerry Kelly (1987), *Social Choice Theory: An Introduction*, Springer-Verlag, Berlin.

Eric Maskin (1977), "Nash Equilibrium and Welfare Optimality," mimeo., M.I.T.

Eitan Muller and Mark Satterthwaite (1977), "The Equivalence of Strong Positive Association and Strategy-proofness," *Journal of Economic Theory* 14, pp.412-418.

Mark Satterthwaite (1975), "Strategy-proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions," *Journal of Economic Theory* 10, pp.187-217.