Rev. Econ. Design 8, 319–328 (2003) DOI: 10.1007/s10058-003-0102-2



An alternative direct proof of Gibbard's random dictatorship theorem

Yasuhito Tanaka

Faculty of Economics, Doshisha University, Kamigyo-ku, Kyoto, 602-8580, Japan (e-mail: yatanaka@mail.doshisha.ac.jp)

Received: 15 October 2001 / Accepted: 23 May 2003

Abstract. We present an alternative proof of the Gibbard's random dictatorship theorem with ex post Pareto optimality. Gibbard(1977) showed that when the number of alternatives is finite and larger than two, and individual preferences are linear (strict), a strategy-proof decision scheme (a probabilistic analogue of a social choice function or a voting rule) is a convex combination of decision schemes which are, in his terms, either unilateral or duple. As a corollary of this theorem (credited to H. Sonnenschein) he showed that a decision scheme which is strategy-proof and satisfies ex post Pareto optimality is randomly dictatorial. We call this corollary the *Gibbard's random dictatorship theorem*. We present a proof of this theorem which is direct and follows closely the original Gibbard's approach. Focusing attention to the case with ex post Pareto optimality our proof is more simple and intuitive than the original Gibbard's proof.

JEL classification: D71, D72

Key words: Decision scheme, ex post Pareto optimality, Gibbard's random dictatorship theorem

1 Introduction

We present an alternative proof of the Gibbard's random dictatorship theorem with ex post Pareto optimality. Gibbard (1977) showed that when the number of alternatives is finite and larger than two, and individual preferences are linear (strict), a strategy-proof decision scheme (a probabilistic analogue of a social choice function or a voting rule) is a convex combination of decision schemes which are, in

The author is grateful to an anonymous referee and the Associate editor of this journal for very helpful comments and suggestions. And this research has been supported by a grant from the Zengin Foundation for Studies on Economics and Finance in Japan.

his terms, either unilateral or duple. As a corollary of this theorem (credited to H. Sonnenschein) he showed that a decision scheme which is strategy-proof and satisfies ex post Pareto optimality¹ is randomly dictatorial. We call this corollary the *Gibbard's random dictatorship theorem*. Nandeibam (1998) presented an alternative proof of this theorem. His proof is indirect. He proved the theorem using a probabilistic analogue of the Arrow impossibility theorem by Pattanaik and Peleg (1986). Duggan (1996) presented a direct and geometric proof of the theorem².

We present a proof of this theorem which is direct and follows more closely the original Gibbard's approach. Focusing attention to the case with ex post Pareto optimality our proof is more simple and intuitive than the original Gibbard's proof and other proofs. Terminologies in this paper also follow those of Gibbard (1977).

2 Decision scheme and its strategy-proofness

There is a society with n individuals and m alternatives. n and m are finite positive integers such that $n \geq 2$ and $m \geq 3$. The set of individuals is denoted by N, and the set of alternatives is denoted by A. The individuals are represented by i, j and so on, and the alternatives are represented by x, y, z and so on. The preference of individual i about the alternatives is represented by a linear order P_i , which is complete, antisymmetric and transitive. An n-tuple of individual preferences is called a profile. Profiles are denoted by a, b and so on. The preference of individual i at a profile a is denoted by a, b and so on.

We introduce some key terms.

Decision scheme. It is a social choice rule which assigns a probability distribution to alternatives depending on a profile. Denote the probability assigned to x at a profile a by p(x,a), and the vector of probabilities (or the probability distribution) at a is denoted by d(a). p(x,a) is called the probability of x at a, and d(a) is called a value of the decision scheme.

Utility function. To treat individual evaluation about random realization of alternatives we must consider utility functions for individuals. Denote individual i's utility function over alternatives at a profile a by $u_i^a(x)$. It satisfies von Neumann-Morgenstern expected utility axioms. If, when xP_i^ay (or yP_i^ax) we have $u_i^a(x)>u_i^a(y)$ (or $u_i^a(x)<u_i^a(y)$), then $u_i^a(x)$ fits to individual i's preference at a. His expected utility when the value of the decision scheme is d(a) is written as follows,

$$U_i^a(d(a)) = \sum_{j=1}^m u_i^a(x_j) p(x_j, a)$$

Strategy-proofness. Let a and b two profiles between which only the preference of individual i differs. Denote values of a decision scheme at a and b by d(a) and

¹ The condition of ex post Pareto optimality is called *Pareto optimificity ex post* in Gibbard (1977).

² Recently Dutta et al. (2002) have extended the result of Gibbard (1977) to the case where individuals have strictly convex continuous single-peaked preferences on a convex subset of Euclidean space.

d(b). If for a utility function which fits to individual i's preference we have

$$U_i^a(d(b)) > U_i^a(d(a))$$

then the decision scheme is manipulable for him at a. If a decision scheme is not manipulable for any individual at any profile for any utility function which fits to each individual preference, it is strategy-proof.

Further we use some terms and notations which are due to Gibbard (1977).

Adjacent alternatives. $(xP_i!y)$ If for a pair of alternatives $\{x,y\}$ we have xP_iy for individual i and there is no alternative which satisfies xP_izP_iy , we denote $xP_i!y$. And it is said that x and y are adjacent in individual i's preference.

Pairwise responsive. Suppose that at a profile a x and y satisfy $xP_i!y$. Let b be a profile at which $yP_i!x$ and all other preferences are the same as those at a, and z be an arbitrary alternative other than x and y. Denote the probabilities assigned by a decision scheme to z at a and b by p(z,a) and p(z,b). If p(z,b)=p(z,a), the decision scheme is pairwise responsive. Then, the sum of the probabilities assigned to x an y does not change when $xP_i!y$ changes to $yP_i!x$ keeping other preferences constant.

Localized. Let X be a set of some alternatives and Y be a set of all other alternatives (Y = A - X). Suppose that, at a profile a, for any pair of alternatives $\{x,y\}$ such that $x \in X$ and $y \in Y$ we have xP_iy for some individual i. Denote the sum of the probabilities of all alternatives in X by p(X,a). If p(X,a) is not affected by a change in individual i's preference about alternatives in X and a change in his preference about alternatives in Y (his preference between x and y does not change), then the decision scheme is called localized.

Next we assume the following form of Pareto principle.

Ex post Pareto optimality. For any pair of alternatives $\{x, y\}$, if all individuals prefer x to y, the probability of y is zero.

If the probability assigned to the top-ranked alternative of individual i is always one, he is the dictator. If there is the dictator for a decision scheme, it is dictatorial. Further we define a $randomly\ dictatorial$ decision scheme.

Randomly dictatorial decision scheme. If the probability distribution assigned to the alternatives by a decision scheme is a convex combination of probability distributions assigned by some dictatorial decision schemes, then it is randomly dictatorial.

A randomly dictatorial decision scheme has the following properties.

- (1) It assigns positive probabilities to only top-ranked alternatives for individuals. The value of the probability of an alternative does not depend on what alternative it is, but depends on for whom it is the top-ranked alternative. And they do not depend on the preference of each individual about alternatives other than his top-ranked alternative.
- (2) The probability of an alternative is regarded as the sum of the weights of individuals for whom it is the top-ranked alternative.

(3) The weight of each individual is independent of the preferences of individuals, and is not larger than one.

3 The random dictatorship theorem and an outline of its proof

We will show the following theorem.

Theorem 1. A strategy-proof decision scheme which satisfies ex post Pareto optimality is randomly dictatorial.

We will prove this theorem by several lemmata. The proof will proceed along the following processes.

- (1) (Lemma 1 and 2) If a decision scheme is strategy-proof, it is localized and pairwise responsive.
- (2) (Lemma 3) Let $\{x,y\}$ be an arbitrary pair of two alternatives which are adjacent and are not top-ranked alternative for an individual (denoted by individual 1). If a decision scheme is strategy-proof and satisfies ex post Pareto optimality, then, when the preference of individual 1 changes from $xP_1!y$ to $yP_1!x$ and all other preferences do not change, the probabilities of x and y do not change.

This result will be used in the proofs of Lemma 4 and 5.

- (3) (Lemma 4) Let a be a profile such that x is the top-ranked alternative for individuals in a group of individuals V, and it is bottom-ranked for other individuals. Denote the probability of x in this case by $p_V(x,a)$. Then, if a decision scheme is strategy-proof and satisfies ex post Pareto optimality, we obtain the following results.
 - (i) $p_N(x,a) = 1$.
 - (ii) $p_V(x,a)$ does not depend on the preferences of individuals about alternatives other than x.
 - (iii) Let b be a profile such that the top-ranked alternative for individuals in V is z, and it is the bottom-ranked alternative for other individuals. Denote the probability of z at b by $p_V(z,b)$. Then, we have $p_V(z,b) = p_V(x,a)$.

The value of $p_V(x,a)$ defined here does not depend on individual preferences except for the fact that the top-ranked alternative is common for individuals in V, and this alternative is bottom-ranked for other individuals. The value of $p_V(x,a)$ is determined by what members constitute the group V, and so it is regarded as the weight of V. It is denoted by $\mu(V) = p_V(x,a)$. Denote the set of individuals whose top-ranked alternative is x at a profile a by V(x,a), and the weight of this group by $\mu(V(x,a))$.

(4) (Lemma 5) If a decision scheme is strategy-proof and satisfies ex post Pareto optimality, then the probability assigned to x by the decision scheme at a profile a is represented as follows.

$$p(x, a) = \mu(V(x, a))$$

- (5) (Lemma 6) Let V_1 and V_2 be groups of individuals which are disjoint ($V_1 \cap V_2 = \emptyset$). If a decision scheme is strategy-proof and satisfies ex post Pareto optimality, we have $\mu(V) = \mu(V_1) + \mu(V_2)$ for $V = V_1 \cup V_2$.
- (6) (Lemma 7) If a decision scheme is strategy-proof and satisfies ex post Pareto optimality, the probability assigned to x by the decision scheme are equal to the sum of the weights of individuals for whom x is the top-ranked alternative.

Then, we obtain a random dictatorship.

In the next section we prove these lemmata. And in the last section we will prove the converse of Theorem 1.

4 Lemmata and their proofs

First we show the following lemma.

Lemma 1. If a decision scheme is localized, it is pairwise responsive.

Proof. Suppose that x and y are adjacent in the preference of individual i $(xP_i!y)$. Let z be an alternative other than x and y. Denote the set of alternatives which individual i prefers to z by W, and denote the set which is the union of W and $\{z\}$ by W'. Since x and y are adjacent, they belong to W' if and only if they belong to W. Even when individual i's preference changes from $xP_i!y$ to $yP_i!x$, the sum of the probabilities assigned by a localized decision scheme to alternatives in W and that for W' do not change. Therefore, the probability of z does not change, and the decision scheme is pairwise responsive. \square

From strategy-proofness we obtain the following result.

Lemma 2. If a decision scheme is strategy-proof, it is localized and pairwise responsive.

Proof.

(1) Assume that it is not localized. Let X be the set of alternatives which individual i prefers to other alternatives. Denote the sum of the probabilities of alternatives in X at profiles a and b by p(X,a) and p(X,b). If the decision scheme is not localized, there is a case where when individual i's preference changes from P_i^a to P_i^b , p(X,b) and p(X,a) are different. Let $p(X,b)-p(X,a)=\varepsilon>0$ ($0<\varepsilon<1$), and let u_i^a be a utility function which fits to individual i's preference at a and satisfies $1\leq u_i^a(x)<1+\varepsilon$ for $x\in X$ and $0\leq u_i^a(y)<\varepsilon$ for $y\in Y$. Then, expected utilities of individual i at a and b satisfy the following inequalities.

$$U_i^a(d(a)) < (1+\varepsilon)p(X,a) + \varepsilon[1-p(X,a)] = p(X,a) + \varepsilon$$

$$U_i^a(d(b)) \ge 1 \cdot p(X,b) + 0 \cdot [1 - p(X,b)] = p(X,b) = p(X,a) + \varepsilon$$

From these inequalities we obtain $U_i^a(d(b)) > U_i^a(d(a))$. Note that $U_i^a(d(b))$ and $U_i^a(d(a))$ are expected utilities for the utility function which fits to the

individual i's preference at a. $U_i^a(d(b)) > U_i^a(d(a))$ means that at the profile a individual i can gain higher expected utility by revealing his preference at b, and so the decision scheme is manipulable. In the case where p(X,b)-p(X,a)<0, we can show the similar result by considering a utility function which fits to the individual i's preference at b. Therefore, we have p(X,b)-p(X,a)=0, and the decision scheme is localized.

(2) From Lemma 1 if a decision scheme is localized, it is pairwise responsive. \Box

Next we show

Lemma 3. Let $\{x,y\}$ be an arbitrary pair of two alternatives which are adjacent and are not top-ranked alternative for an individual (denoted by individual 1). If a decision scheme is strategy-proof and satisfies ex post Pareto optimality, then, when the preference of individual 1 changes from $xP_1!y$ to $yP_1!x$ and all other preferences do not change, the probabilities of x and y do not change.

Proof. Denote the profile assumed in this Lemma by a. Let b be a profile such that individual 1's preference changes from $xP_1^a!y$ to $yP_1^b!x$ and all other preferences do not change. Denote the probabilities of x and y at a and b by p(x,a), p(y,b) and so on. Then, by pairwise responsiveness we obtain

$$p(y,b) - p(y,a) = p(x,a) - p(x,b) = \varepsilon \tag{1}$$

In the rest of the proof we show $\varepsilon = 0$.

Let z be the top-ranked alternative for individual 1 at a. From the assumption of this lemma z is different from x and y. If z is the top-ranked alternative for all individuals, by ex post Pareto optimality the probabilities of all alternatives other than z are zero, and that of z is one. Therefore, a change in the order of x and y in each individual's preference does not affect the probability of any alternative, and we have $\varepsilon=0$.

Suppose that there are some individuals whose top-ranked alternative is not z, and let individual 2 be one of them. Let a_2 be a profile such that the order of z and w_1 , which satisfies $w_1P_2!z$ in individual 2's preference, has changed at a, and b_2 be a profile such that the order of z and w_1 in individual 2's preference has changed at b. Then, if w_1 is different from x and y, pairwise responsiveness implies $p(x,a)=p(x,a_2), p(y,a)=p(y,a_2), p(x,b)=p(x,b_2)$ and $p(y,b)=p(y,b_2)$. From (1) we obtain

$$p(y,b_2) - p(y,a_2) = p(y,b) - p(y,a) = p(x,a) - p(x,b)$$

= $p(x,a_2) - p(x,b_2) = \varepsilon$ (2)

$$U_{i}^{b}(d(a)) \ge 1 \cdot p(X, a) + 0 \cdot [1 - p(X, a)] = p(X, a) = p(X, b) + \varepsilon$$
$$U_{i}^{b}(d(b)) < (1 + \varepsilon)p(X, b) + \varepsilon[1 - p(X, b)] = p(X, b) + \varepsilon$$

and

$$U_i^b(d(a)) > U_i^b(d(b)).$$

Next, consider the case where w_1 and y are the same. By pairwise responsiveness we have p(z,a)=p(z,b) and $p(z,a_2)=p(z,b_2)$. Comparing a with a_2 , and b with b_2 , using these relations, we obtain

$$p(z, a_2) - p(z, a) = p(y, a) - p(y, a_2)$$

and

$$p(z, a_2) - p(z, a) = p(y, b) - p(y, b_2)$$

From them

$$p(y, a) - p(y, a_2) = p(y, b) - p(y, b_2)$$

It means

$$p(y,b) - p(y,a) = p(y,b_2) - p(y,a_2)$$
(3)

Comparing a_2 with b_2 , pairwise responsiveness implies

$$p(y, b_2) - p(y, a_2) = p(x, a_2) - p(x, b_2)$$

From (3) it is equal to ε .

Similarly in the case where w_1 and x are the same, we can show⁴

$$p(y, b_2) - p(y, a_2) = p(x, a_2) - p(x, b_2) = \varepsilon$$

These results mean that the magnitudes of changes in the probabilities of x and y due to a change in the order of x and y in individual 1's preference, which is the value of ε , is not affected by a change in the order of z and w_1 in individual 2's preference.

By similar arguments we can show that a change in the order of z and w_2 , which is the next upper alternative in individual 2's preference, does not affect the value of ε , and until z shifts up to the top of individual 2's preference, the value of ε does not change. These arguments are applied to all individuals other then individual 2 whose top-ranked alternative is not z. After all until z shifts up to the top of all individuals' preferences the value of ε does not change. When z is the top-ranked alternative for all individuals, ex post Pareto optimality means that the probability of z is one. Therefore, always we have $\varepsilon=0$, and the lemma has been proved. \Box

Next we show

Lemma 4. Let a be a profile such that x is the top-ranked alternative for individuals in a group of individuals V, and it is bottom-ranked for other individuals. Denote the probability of x in this case by $p_V(x,a)$. Then, if a decision scheme is strategy-proof and satisfies ex post Pareto optimality, we obtain the following results.

$$p(x,b) - p(x,a) = p(x,b_2) - p(x,a_2)$$

⁴ Instead of (3) we obtain

- (1) $p_N(x,a) = 1$.
- (2) $p_V(x, a)$ does not depend on the preferences of individuals about alternatives other than x.

(3) Let b be a profile such that the top-ranked alternative for individuals in V is z, and it is the bottom-ranked alternative for other individuals. Denote the probability of z at b by $p_V(z,b)$. Then, we have $p_V(z,b) = p_V(x,a)$.

Proof.

- (1) By ex post Pareto optimality the probabilities of alternatives other than x are all zero. Thus, the probability of x is one.
- (2) By Lemma 2 the decision scheme is localized, and so the probability of x is not affected by changes in the preferences of individuals in V about alternatives other than x. Since x is bottom-ranked for individuals other than those in V, localizedness also implies that the probability of x is not affected by changes in the preferences of individuals other than those in V about alternatives other than x.
- (3) Consider the following profiles:
 - (i) a: The top-ranked alternative for individuals in V is x, and they have a preference $xP_i!zP_i!y$; and the bottom-ranked alternative for other individuals is x, and they have a preference $yP_i!zP_i!x$.
 - (ii) a': The preferences of individuals in V change from $xP_i!zP_i!y$ to $xP_i!yP_i!z$.
 - (iii) b: The top-ranked alternative for individuals in V is z, and they have a preference $zP_i!xP_i!y$; and the bottom-ranked alternative for other individuals is z, and they have a preference $yP_i!xP_i!z$.
 - (iv) b': The preferences of individuals in V change from $zP_i!xP_i!y$ to $zP_i!yP_i!x$.

Individual preferences about alternatives other than x,y and z do not change. The probability of x at a is equal to $p_V(x,a)$. By pairwise responsiveness the probability of x at a' is also equal to $p_V(x,a)$. By Lemma 3 the probability of z at a and that at a' are equal. By ex post Pareto optimality the probability of z at a' is zero. Therefore, the probability of z at a' is also zero. The probability of z at a' is also equal to a' in a pairwise responsiveness the probability of a' at a' is zero. Therefore, the probability of a' at a' is also zero. Between a' and a' in each individual preference differs. Therefore, pairwise responsiveness implies that the probabilities of alternatives other than a' and a' are not changed between a' and a' in the sum of the probabilities of a' and a' is also not changed. Therefore, we obtain a' in a' in a' in a part a' is also not changed. Therefore, we obtain a' in a' in

The value of $p_V(x,a)$ defined here does not depend on individual preferences except for the fact that the top-ranked alternative is common for individuals in V, and this alternative is bottom-ranked for other individuals. The value of $p_V(x,a)$ is determined by what members constitute the group V, and so it is regarded as the

weight of V. It is denoted by

$$\mu(V) = p_V(x, a) \tag{4}$$

If V includes only one individual, $\mu(V)$ is the weight of that individual.

Denote the set of individuals whose top-ranked alternative is x at a profile a by V(x,a), and the weight of this group by $\mu(V(x,a))$. We will show the following result.

Lemma 5. If a decision scheme is strategy-proof and satisfies ex post Pareto optimality, then, the probability assigned to x by the decision scheme at some profile a is represented as follows:

$$p(x,a) = \mu(V(x,a)) \tag{5}$$

Proof. First, when x is the top-ranked alternative for individuals in V and it is bottom-ranked for other individuals, by the definition of $\mu(V(x,a))$ we obtain $p(x,a)=\mu(V(x,a))$. From Lemma 4 its value does not depend on the preferences of individuals other than those in V about alternatives other than x. On the other hand, from Lemma 3 even if the order of x and any other alternative changes in the preferences of individuals other than those in V, the probability of x does not change so long as x is not top-ranked for them. Thus, generally we obtain $p(x,a)=\mu(V(x,a))$. \square

Next we show

Lemma 6. Let V_1 and V_2 be groups of individuals which are disjoint $(V_1 \cap V_2 = \emptyset)$. If a decision scheme is strategy-proof and satisfies ex post Pareto optimality, we have $\mu(V) = \mu(V_1) + \mu(V_2)$ for $V = V_1 \cup V_2$.

Proof. Suppose that x is the top-ranked alternative for individuals in V_1 , the top-ranked alternative for individuals in V_2 is $y(\neq x)$, and their secondly-ranked alternative is x. And suppose that the top-ranked alternative for all other individuals is different from x and y, and they prefer x to y. Denote this profile by a. Then, by Lemma 5 the probability of x and that of y at a are, respectively, equal to $\mu(V_1)$ and $\mu(V_2)$.

Now, let b be a profile at which the order of x and y in the preferences of individuals in V_2 have changed from $yP_i!x$ to $xP_i!y$. From Lemma 5 the probability of x at b is equal to $\mu(V_1 \cup V_2)$. By pair-wise responsiveness the sum of the probabilities of x and y does not change, and by ex-post Pareto optimality the probability of y at b is zero. Therefore, we obtain

$$p(x,b) = \mu(V_1 \cup V_2) = p(x,a) + p(y,a)$$

This means

$$\mu(V_1 \cup V_2) = \mu(V_1) + \mu(V_2)$$

If V_1 and V_2 consist of, respectively, only one individual, this lemma implies that the weight of a set which contains two individuals is equal to the sum of the weights of the two individuals, and the weight of a set which contains one more individual is equal to the sum of the weights of the three individuals, and so on. Therefore, the weight of a group is equal to the sum of the weights of members of the group. Since $p_N(x,a) = 1$, Lemma 5 and 6 imply the following result.

Lemma 7. If a decision scheme is strategy-proof and satisfies ex post Pareto optimality, the probability assigned to x by the decision scheme are equal to the sum of the weights of individuals for whom x is the top-ranked alternative.

This means that the decision scheme is randomly dictatorial, and the proof of Theorem 1 has been completed.

5 The converse of the random dictatorship theorem

Finally, we verify the converse of Theorem 1.

Theorem 2. A randomly dictatorial decision scheme is strategy-proof.

Proof. A randomly dictatorial decision scheme assigns positive probabilities to only alternatives which are top-ranked alternatives for some individuals, and the probabilities assigned to each alternative is equal to the sum of the weights of individuals for whom it is the top-ranked alternative. Let us consider a decision scheme $d(\cdot)$ and denote the weight of individual i by $\mu(i)$ for $i=1,2,\cdots,n$. Then, the expected utility for some individual (denoted by k) at a profile a when all individuals reveal their true preferences is

$$U_k^a(d(a)) = \sum_{i=1}^n u_k^a(\bar{x}_i)\mu(i)$$

 \bar{x}_i denotes the top ranked alternative for individual i. Suppose that individual k reveals a different preference and denote such a profile by b. If his top ranked alternative at a and that at b are the same, his expected utility does not change between a and b. On the other hand, if he reveals a different top ranked alternative, his expected utility clearly lowers. And hence he does not have an incentive to reveal a preference which is different from his true preference. This argument can be applied to all individuals. Therefore, any randomly dictatorial decision scheme is strategy-proof. \Box

References

Dutta, B., Peters, H., Sen, A. (2002) Strategy-proof probabilistic mechanisms in economies with pure public goods. *Journal of Economic Theory* 106: 392–416

Duggan, J. (1996) A geometric proof of Gibbard's random dictatorship theorem. *Economic Theory* 7: 365–369

Gibbard, A. (1977) Manipulation of schemes that mix voting with chance. *Econometrica* 45: 655–681Nandeibam, S. (1998) An alternative proof of Gibbard's random dictatorship result. *Social Choice and Welfare* 15: 509–519

Pattanaik, P.K., Peleg, B. (1986) Distribution of power under stochastic social choice rules. *Econometrica* 54: 909–921