Strategy-Proof Probabilistic Mechanisms in Economies with Pure Public Goods[1]

Bhaskar Dutta

Indian Statistical Institute, 7 S. J. S. Sansanwal Marg, New Dehli 110 016, India
E-mail: dutta@isid.ac.in

Hans Peters[2]

Department of Quantitative Economics, University of Maastricht,
P.O. Box 616, 6200 MD Maastricht, The Netherlands
E-mail: h.peters@ke.unimaas.nl

and

Arunava Sen

Indian Statistical Institute, 7 S. J. S. Sansanwal Marg, New Dehli 110 016, India
E-mail: asen@isid.ac.in

Received October 24, 2000

Public good economies where agents are endowed with strictly convex continuous single-peaked preferences on a convex subset of Euclidean space are considered. Such an economy arises for instance in the classical problem of allocating a given budget to finance the provision of several public goods where the agents have monotonically increasing strictly convex continuous preferences. A probabilistic mechanism assigns a probability distribution over the feasible alternatives to any profile of reported preferences. The main result of the paper establishes that any strategy-proof (in the sense of A. Gibbard, *Econometrica* **45** (1977), 665–681) and unanimous mechanism must be a random dictatorship. *Journal of Economic Literature* Classification Numbers: D70, D71, H40, C60.

Key Words: Public goods; probabilistic mechanism; strategy-proofness; random dictatorship.

[1] The authors thank the Indian Statistical Institute at New Delhi and the Faculty of Economics and Business Administration of the University of Maastricht for their hospitality. Thanks are also due to Lars Ehlers, Ton Storcken, and the anonymous referees for helpful suggestions.

[2] To whom correspondence should be addressed.

1. INTRODUCTION

The classic result of Gibbard [6] and Satterthwaite [17] demonstrated that when preferences are unrestricted, essentially the only decentralized decision procedure or mechanism which will always induce individual agents to truthfully report their private information is the dictatorial one. The impossibility of constructing *strategy-proof* mechanisms has resulted in a huge literature which has explored many different variants of the basic result. One variant which is the main focus of this paper is the extension of the Gibbard–Satterthwaite result to mechanisms which assign a *probability distribution* over the set of outcomes for each profile of preferences. Gibbard [7] characterized the class of such strategy-proof *probabilistic* mechanisms or *decision schemes*. He showed that a strategy-proof probabilistic mechanism is a convex combination of *duples* and *unilaterals*, where a duple is a mechanism which assigns positive probability to at most two alternatives, and a unilateral is one in which a single individual is a dictator.[3]

Such a mechanism need not satisfy *ex post Pareto optimality*, that is, it may assign a positive probability to an alternative a which is worse than another alternative b for all individuals. Hylland [9] showed that even if probabilistic mechanisms are allowed to utilize information about individual *cardinal* utilities, the only strategy-proof probabilistic mechanisms satisfying ex post Pareto optimality (or even unanimity) are *random dictatorships*, in which each individual is assigned a fixed probability of being a dictator. Hylland also showed that if the stronger requirement of *ex ante Pareto optimality* is imposed, then only dictatorial mechanisms can satisfy strategy-proofness. See also Nandeibam [15] and Duggan [4], who provide proofs of the random dictatorship result when the mechanism is allowed to use only ordinal information. By way of an illustration, if there are two individuals 1 and 2 and two alternatives a and b such that 1 prefers a and 2 prefers b, then the random dictatorship where a and b result with equal probability seems a good solution. If, however, a third alternative c is available which is ranked second by both individuals, then it is not so obvious that the random dictatorship is an attractive way to solve the problem. Thus, although in some situations random dictatorship may be an improvement compared to deterministic dictatorship, it is generally speaking not considered to be an appealing solution.

It is clear from the literature on deterministic mechanisms that the possibility of constructing non-dictatorial strategy-proof mechanisms depends crucially on the domain of individual preferences as well as on the structure of the set of feasible outcomes. Perhaps the best illustration of this is the

[3] See also Barbera [1, 2] for related results on probabilistic mechanisms.

case where individual preferences are *single-peaked* and the range of possible outcomes is *one-dimensional*. Moulin [14] shows that (under anonymity) a strategy-proof mechanism selects the peak of the median voter after addition of *phantom voters*, i.e., fixed ballots. This result has been extended to probabilistic mechanisms by Ehlers *et al.* [5], who show that in a strategy-proof probabilistic mechanism the fixed ballots are replaced by fixed probability distributions.

While single-peakedness makes sense in a variety of economic and political models, the assumption that the set of outcomes is one-dimensional is unduly restrictive. In many cases, it makes sense to assume that individual rankings of outcomes depend on *several* characteristics. For instance, each outcome may represent levels of public expenditure on different public goods such as defence, education. Alternatively, outcomes may represent locations of different public facilities, alternative tax proposals, and so on. Given a fixed available budget and monotonically increasing convex preferences such situations can be modeled by assuming single-peaked preferences on a higher dimensional set of alternatives.

Several recent papers have focused on the characterization of strategy-proof (deterministic) rules when outcomes are multidimensional and preferences satisfy some generalized notion of single-peakedness. An early and influential paper was by Border and Jordan [3], who assumed that individual preferences are *star-shaped*. In this context the smaller domain of spherical preferences was considered in Kim and Roush [11] and Peters *et al.* [16]. Zhou [19] considers a larger domain by assuming that preferences are convex and continuous. Moreno [13] extends Zhou's result by restricting the domain of individual preferences to satisfy monotonicity. The main result of these papers is that the dictatorship result emerges once again if the range of outcomes is multidimensional.[4]

In the present paper, we move away from the setting of Gibbard [7, 8] and Hylland [9] by assuming that the set of alternatives is some convex set in $\mathbb{R}^k$ with $k > 1$, and that preferences are strictly convex and continuous with a unique peak. We then show that the only strategy-proof probabilistic mechanisms satisfying *unanimity* (if all individuals have the same preference, then the common best outcome is selected with probability one) are random dictatorships. Thus, we show that the extension to probabilistic mechanisms does not allow us to completely escape the negative consequences of the original Gibbard–Satterthwaite result even when the domain of the mechanism is restricted to preferences which are more appropriate for a large variety of political and economic models. As a byproduct of our

[4] Le Breton and Sen [12] derive decomposition results on strategy-proofness when the set of alternatives is multidimensional and preferences are separable. See Sprumont [18] for an insightful survey on strategy-proofness in multidimensional environments.

analysis, we also derive a result which is very close to the main result of Zhou [19].

The model and the main results are formulated in Section 2. In Section 3 we derive random dictatorship from unanimity and strategy-proofness for the case of two agents. The n-agents result is derived in Section 4, by mathematical induction based on the two-agent result. It is worthwhile to note that to perform the induction only strategy-proofness is needed. In Section 5 the deterministic case is considered, and Section 6 concludes.

2. MODEL AND MAIN RESULTS

The set of *alternatives* is denoted by A. Throughout, it is assumed that A is a non-empty convex subset of some Euclidean space $\mathbb{R}^k$. The dimension of A, denoted by $\dim(A)$, is the dimension of the smallest affine subset of $\mathbb{R}^k$ containing A.

The set of *agents* is denoted by N, where $N = \{1, ..., n\}$ for some natural number n.

A (single-peaked strictly convex continuous) *preference* r on A is a transitive and complete binary relation on A with the following properties:

(i) There is a unique $p(r) \in A$ with $p(r)\, ra$ for all $a \in A$. The point $p(r)$ is called the *peak* of r.

(ii) For all $a, b \in A$ with arb and $a \neq b$ and all $\lambda \in \mathbb{R}$ with $0 < \lambda < 1$ we have $[\lambda a + (1-\lambda)\, b]\, pb$, where p denotes the asymmetric part of r.

(iii) For every $a \in A$ the sets $B(a, r) := \{b \in A\colon bra\}$ and $\{b \in A\colon arb\}$ are closed.

The interpretation is as usual: arb means that an agent with preference r weakly prefers a to b, and apb means that a is strictly preferred to b. The set of all preferences is denoted by Ω. More frequently, the notation R_i will be used to denote a preference of agent $i \in N$, with corresponding asymmetric part P_i and symmetric part or indifference relation I_i. The conditions on R_i imply that the upper contour set $B(a, R_i)$, i.e., the set of alternatives weakly preferred to a by an agent with preference R_i, is not only closed but also strictly convex and that its boundary (indifference curve) consists exactly of those points $b \in A$ with bI_ia.

A (preference) *profile* is an element $R = (R_1, ..., R_n)$ of Ω^N.

Let the set A be endowed with the usual Borel σ-algebra, and let $M(A)$ denote the set of all probability measures on A.

A *probabilistic mechanism* $\mu\colon \Omega^N \to M(A)$ assigns to each profile of preferences $R = (R_1, ..., R_n)$ a probability measure $\mu(R) = \mu(R_1, ..., R_n)$ on A.

Henceforth, we will simply use the term *mechanism*. For a Borel set $B \subset A$, $\mu(R)(B)$ can be interpreted as the probability that the chosen alternative will lie in the set B if R is the profile of preferences and μ the mechanism that is used. It is assumed that any set $B \subset A$ occurring in the sequel is a Borel set. Instead of $\mu(R)(\{x\})$, where $x \in A$, the notation $\mu(R)(x)$ will be used.

Note that a mechanism as we have defined it uses only *ordinal* information about individual preferences. This is in contrast to Hylland [9] who considers the larger class of mechanisms which can in principle use *cardinal* information.

DEFINITION 2.1. A mechanism μ is *strategy-proof* if for all $R = (R_1, \ldots, R_n) \in \Omega^N$, all $i \in N$, all $R_i' \in \Omega$ and all $a \in A$

$$\mu(R)(B(a, R_i)) \geqslant \mu(R_{-i}, R_i')(B(a, R_i)),$$

where (R_{-i}, R_i') denotes the profile obtained from R be replacing agent i's preference R_i by R_i'.

In words, a mechanism is strategy-proof if no agent can increase the probability on one of his upper contour sets by not reporting his true preference.

At first sight, this may seem a somewhat unusual definition of strategy-proofness. Consider, however, the following alternative and perhaps more familiar approach, which results in the same strategy-proofness condition. Let the individuals have von Neumann–Morgenstern (vNM) utility functions by which they rank alternative probability measures. Assume that the mechanism can only use ordinal information so that individuals can only announce *preference orderings*. Then it is natural to call a mechanism non-manipulable only if the individual has no incentive to announce, say, R_i' instead of his true preference R_i no matter which particular vNM utility function represents R_i. This is possible if and only if the individual cannot increase the probability on one of his upper contour sets corresponding to R_i by reporting R_i'. Put differently and in accordance with the familiar characterization of first degree stochastic dominance, by misreporting the mechanism produces a probability distribution that is first degree stochastically dominated by the one that results from reporting the true preference.

For a probability measure $m \in M(A)$ denote by supp(m) the support of m, i.e., the set

$$\text{supp}(m) := \{a \in A : m(B) > 0 \text{ for every open } B \subset A \text{ with } a \in B\}.$$

For a profile $R = (R_1, \ldots, R_n)$ denote by $PO(R)$ the Pareto optimal set for R, i.e.,

$$PO(R) := \{a \in A\text{: for all } b \in A \text{ and } i \in N,$$
$$\text{if } bP_i a \text{ then there is a } j \in N \text{ with } aP_j b\}.$$

DEFINITION 2.2. A mechanism μ is *Pareto optimal* if $\text{supp}(\mu(R)) \subset PO(R)$ for every $R \in \Omega^N$.

DEFINITION 2.3. A mechanism μ is *unanimous* if for every profile $R = (r, ..., r)$, where $r \in \Omega$, we have $\mu(R)(p(r)) = 1$.

Obviously, unanimity is implied by Pareto optimality.

DEFINITION 2.4. A mechanism μ is called a *random dictatorship* if there are nonnegative numbers $\lambda_1, ..., \lambda_n \in \mathbb{R}$ with $\sum_{i \in N} \lambda_i = 1$ such that for every $R \in \Omega^N$ and every $B \subset A$:

$$\mu(R)(B) = \sum_{i \in N} \lambda_i 1_B(p(R_i)).$$

Here, 1_B denotes the indicator function of B, i.e., $1_B\colon A \to \{0, 1\}$ is defined by $1_B(a) = 1$ if $a \in B$ and $1_B(a) = 0$ otherwise.

The main result of this paper is the following theorem.

THEOREM 2.1. *Let* $\dim(A) \geqslant 2$ *and let* $\mu\colon \Omega^N \to M(A)$ *be a mechanism. Then* μ *is strategy-proof and unanimous, if, and only if,* μ *is a random dictatorship.*

The condition on the dimension of A in Theorem 2.1 cannot be omitted. See Moulin [14] for the one-dimensional deterministic case, and Ehlers *et al.* [5] for the extension to the probabilistic case.

The proof of Theorem 2.1 will be given in Section 4. First, in Section 3, the two-agent case will be considered. For this case, we first show that unanimity and strategy-proofness imply Pareto optimality. Next, random dictatorship will be derived from the combination of Pareto optimality and strategy-proofness.

3. THE TWO-AGENT CASE

In this section we prove Theorem 2.1 for the two-person case. The if-part of the theorem is obvious (for any number of agents).

We start with two lemmas that deal only with Pareto optimal curves and not yet with the mechanism μ. The first lemma provides a parametric description of Pareto optimal sets.

LEMMA 3.1. *Let* $R=(R_1, R_2) \in \Omega^N$ *with* $p(R_1) \neq p(R_2)$. *Then there exists a homeomorphism* $\varphi^R: [0,1] \to PO(R)$ *with* $\varphi^R(0)=p(R_1)$, $\varphi^R(1)=p(R_2)$, *and for all* $0 \leqslant t < t' \leqslant 1$: $\varphi^R(t)\, P_1 \varphi^R(t')$ *and* $\varphi^R(t')\, P_2 \varphi^R(t)$.

Proof. This result is intuitive and therefore its proof will only be sketched. For $t \in [0,1]$ let $\varphi^R(t)$ be the unique point on the boundary of the upper contour set $B(p(R_1)+t(p(R_2)-p(R_1)), R_1)$ where agent 2's preference is maximized. Existence follows by applying Weierstrass' Theorem to a continuous representation of R_2 that is to be maximized on the compact set $B(p(R_1)+t(p(R_2)-p(R_1)), R_1)$, and uniqueness follows by strict convexity of the preferences. Then $\varphi^R(t)$ is a Pareto optimal point and every Pareto optimal point can be obtained in this way. The last two claims in the lemma follow by construction, and this also holds for the fact that φ^R is a bijection. Continuity of φ^R and its inverse can be derived by applying the Maximum Theorem for correspondences. ∎

A simple consequence of Lemma 3.1 is the following result.

LEMMA 3.2. *Let* $R=(R_1, R_2)$, $R'=(R'_1, R'_2) \in \Omega^N$ *with* $p(R_1)=p(R'_1)$, $p(R_2)=p(R'_2)$, *and* $PO(R)=PO(R')$. *Then for all* $a \in PO(R)$ *and every* $i \in N$: $B(a, R_i) \cap PO(R) = B(a, R'_i) \cap PO(R)$.

Proof. If $p(R_1)=p(R_2)$ the lemma is obviously true. So suppose $p(R_1) \neq p(R_2)$, let $a \in PO(R)$ and let φ^R and $\varphi^{R'}$ as in Lemma 3.1. Let $i \in N$, without loss of generality $i=1$. Let $t, t' \in [0,1]$ with $\varphi^R(t)=a$ and $\varphi^{R'}(t')=a$. Then Lemma 3.1 implies $B(a, R_1) \cap PO(R) = \{\varphi^R(s): 0 \leqslant s \leqslant t\}$. Also, $\{\varphi^R(s): 0 \leqslant s \leqslant t\} = \{\varphi^{R'}(s): 0 \leqslant s \leqslant t'\}$, and $B(a, R'_1) \cap PO(R) = \{\varphi^{R'}(s): 0 \leqslant s \leqslant t'\}$. This completes the proof. ∎

We now assume that $\mu: \Omega^N \to M(A)$ is a strategy-proof and unanimous mechanism, where $\dim(A) \geqslant 2$ and, without loss of generality, $N=\{1,2\}$, and show that μ must satisfy Pareto optimality.[5]

LEMMA 3.3. *Let* $\dim(A) \geqslant 2$, $n=2$, *and let* μ *be a unanimous and strategy-proof mechanism. Then* μ *is Pareto optimal.*

Proof. Before proving the main statement of the lemma we first prove the following;

CLAIM 3.1. *Let* $R \in \Omega^N$. *Then* $\mu(R)(B(p(R_2), R_1) \cap B(p(R_1), R_2)) = 1$.

[5] The arguments in this proof, like in many of the subsequent proofs, are supported by pictures. The benefit is that the proofs are easier to read (and write) than purely symbolic proofs. The price is extra carefulness in interpreting the pictures.

Proof of Claim. By unanimity and strategy-proofness,

$$1 = \mu(R_2, R_2)(B(p(R_2), R_1)) \leqslant \mu(R_1, R_2)(B(p(R_2), R_1)),$$

so that

$$\mu(R_1, R_2)(B(p(R_2), R_1)) = 1.$$

Similarly one derives

$$\mu(R_1, R_2)(B(p(R_1), R_2)) = 1.$$

The claim now follows. ∎

In order to prove the lemma, let $R \in \Omega^N$. If $p(R_1) = p(R_2)$ then we are done by the Claim, so assume $p(R_1) \neq p(R_2)$. Let x be a point in $B(p(R_2), R_1) \cap B(p(R_1), R_2) \setminus PO(R)$, then by the Claim it is sufficient to show that there is a neighborhood of x that is assigned zero probability by $\mu(R)$. Choose a Pareto optimal point a such that x is outside $B(a, R_2)$ but so close to $B(a, R_2)$ that the latter set can be enlarged to a strictly convex closed set B' for which x is an interior point but which coincides with $B(a, R_2)$ outside a small neighborhood of x. Let $b \neq a$ be a point in $PO(R) \cap B(a, R_2)$ so close to a that $B' \cap B(b, R_1) = B(a, R_2) \cap B(b, R_1)$. This implies, in particular, that x is outside $B(b, R_1)$. Construct a new preference R'_2 with peak $p(R'_2) = b$, such that $B' = B(a, R'_2)$ (such a preference can always be constructed by multiplication of B' with b as center). See Fig. 1 for this and for an illustration of the rest of the proof. We will prove that

$$\mu(R)(B(a, R'_2)) = \mu(R)(B(a, R_2)), \tag{1}$$

which together with the Claim will imply that

$$\mu(R)((B(a, R'_2) \setminus B(a, R_2)) \cap ((B(p(R_2), R_1) \cap B(p(R_1), R_2))) = 0;$$

i.e., the eye-shaped area in Fig. 1 containing x is assigned probability 0 by $\mu(R)$. In particular there is a neighborhood of x that is assigned probability 0, as was to be proved.

We are left to show (1). By the Claim and by strategy-proofness

$$\begin{aligned} &\mu(R_1, R'_2)(B(a, R'_2) \cap B(b, R_1)) \\ &\qquad = \mu(R_1, R'_2)(B(a, R'_2)) \geqslant \mu(R_1, R_2)(B(a, R'_2)). \end{aligned} \tag{2}$$

Since, by construction, $B(a, R'_2) = B' \supset B(a, R_2)$,

$$\mu(R_1, R_2)(B(a, R'_2)) \geqslant \mu(R_1, R_2)(B(a, R_2)). \tag{3}$$

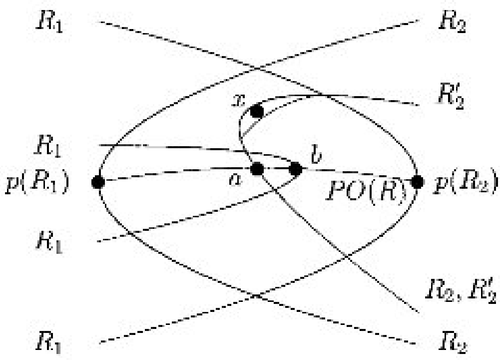

FIG. 1. Proof of Lemma 3.3.

By strategy-proofness

$$\mu(R_1, R_2)(B(a, R_2)) \geqslant \mu(R_1, R_2')(B(a, R_2))$$

hence by the Claim and since $B(a, R_2') \cap B(b, R_1) = B' \cap B(b, R_1) = B(a, R_2) \cap B(b, R_1)$

$$\mu(R_1, R_2)(B(a, R_2)) \geqslant \mu(R_1, R_2')(B(a, R_2')). \tag{4}$$

By applying, respectively, (4), (2), and (3) we have

$$\begin{aligned} \mu(R_1, R_2)(B(a, R_2)) &\geqslant \mu(R_1, R_2')(B(a, R_2')) \\ &\geqslant \mu(R_1, R_2)(B(a, R_2')) \\ &\geqslant \mu(R_1, R_2)(B(a, R_2)) \end{aligned}$$

from which (1) follows. ▮

In view of this lemma, we will henceforth assume that the mechanism is strategy-proof and Pareto optimal.

In the next lemma it is proved that the probability measure assigned by the mechanism μ depends only on the Pareto optimal set.

LEMMA 3.4. *Let $R = (R_1, R_2)$, $R' = (R_1', R_2') \in \Omega^N$ with $p(R_1) = p(R_1')$, $p(R_2) = p(R_2')$, and $PO(R) = PO(R')$. Then $\mu(R) = \mu(R')$.*

Proof. By Lemma 3.1 the Pareto optimal set $PO(R)$ is homeomorphic to the interval $[0, 1]$, and φ^R maps sets of the form $[0, t]$ $(t \in [0, 1])$ to sets $B(a, R_1) \cap PO(R)$ where $a = \varphi^R(t)$. Therefore, in order to prove that

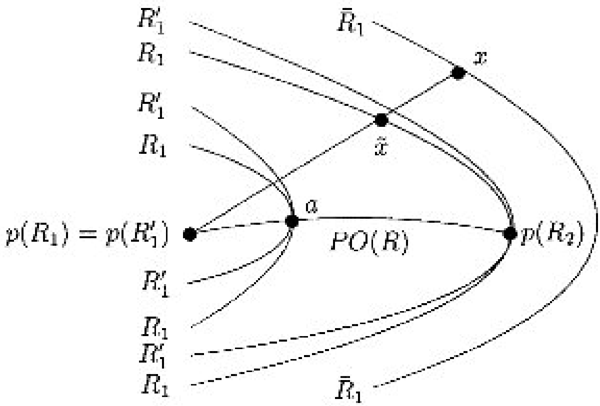

FIG. 2. Proof of Lemma 3.4.

the measures $\mu(R)$ and $\mu(R')$ coincide, it is with Pareto optimality of μ sufficient to prove that for every $a \in PO(R)$

$$\mu(R)(B(a, R_1)) = \mu(R')(B(a, R_1)), \tag{5}$$

because such sets generate the σ-algebra restricted to the Pareto optimal curve $PO(R)$. Construct a preference $\bar{R}_1 \in \Omega$ by constructing upper contour sets as follows. For every $a \in PO(R)$ define $B(a, \bar{R}_1) := B(a, R_1) \cap B(a, R_1')$. For $x \notin B(p(R_2), \bar{R}_1)$ let $\tilde{x}$ be the point on the line segment connecting $p(R_1)$ and x with $\tilde{x} \bar{I}_1 p(R_2)$ i.e., $\tilde{x}$ is on the boundary of $B(p(R_2), \bar{R}_1)$. Let $\lambda(x) > 1$ be defined by $x = p(R_1) + \lambda(x)(\tilde{x} - p(R_1))$. Then define $B(x, \bar{R}_1)$ as the set of those $b \in A$ for which there is a $y \in B(p(R_2), \bar{R}_1)$ with $b = p(R_1) + \lambda(x)(y - p(R_1))$. (In words, the upper contour sets $B(x, \bar{R}_1)$ are obtained by inflating the set $B(p(R_2), \bar{R}_1)$ by an appropriate factor > 1, with $p(\bar{R}_1) = p(R_1)$ as center.) It can be checked that indeed $\bar{R}_1 \in \Omega$. Furthermore, by construction, $PO(\bar{R}_1, R_2) = PO(\bar{R}_1, R_2') = PO(R)$. See Fig. 2. By strategy-proofness, Pareto optimality, and Lemma 3.2, for all $a \in PO(R)$,

$$\mu(R_1, R_2)(B(a, R_1)) \geqslant \mu(\bar{R}_1, R_2)(B(a, R_1)) = \mu(\bar{R}_1, R_2)(B(a, \bar{R}_1)) \tag{6}$$

as well as

$$\mu(\bar{R}_1, R_2)(B(a, \bar{R}_1)) \geqslant \mu(R_1, R_2)(B(a, \bar{R}_1)) = \mu(R_1, R_2)(B(a, R_1)). \tag{7}$$

By combining (6) and (7) it follows that

$$\mu(R_1, R_2)(B(a, R_1)) = \mu(\bar{R}_1, R_2)(B(a, R_1)).$$

Similarly one shows

$$\mu(\bar{R}_1, R_2)(B(a, R_1)) = \mu(\bar{R}_1, R_2')(B(a, R_1))$$

and

$$\mu(R_1', R_2')(B(a, R_1)) = \mu(\bar{R}_1, R_2')(B(a, R_1)).$$

Combining these three equalities yields (5). ∎

The next lemma is the crucial step in the proof of Theorem 2.1 for the two-person case. It shows that, if the Pareto optimal set is a line segment then all probability is assigned to the end points, i.e., the peaks of the preferences. By "conv" we denote "the convex hull of".

LEMMA 3.5. *Let* $R = (R_1, R_2) \in \Omega^N$ *with* $PO(R) = \text{conv}\{p(R_1), p(R_2)\}$. *Then* $\mu(R)(p(R_1)) + \mu(R)(p(R_2)) = 1$.

Proof. The proof will be by construction. For this construction it is assumed that $k = 2$, hence $A \subset \mathbb{R}^2$. The general case can be obtained by embedding this construction in higher dimension.

For convenience, assume $p(R_1) = (0, 0)$ and $p(R_2) = (1, 0)$, so that $PO(R)$ is the line segment with endpoints $(0, 0)$ and $(1, 0)$. Define the set $C \subset \mathbb{R}^2$ by

$$C := \{(\xi, \eta) \in \mathbb{R}^2 : 0 \leqslant \xi \leqslant 1, \eta = \gamma\xi(1 - \xi)\},$$

where $\gamma > 0$ is sufficiently small such that $C \subset A$; this is possible because $\dim(A) \geqslant 2$ implies that there is some point $x \in A$ with (without loss of generality) positive second coordinate and convexity of A then implies $\text{conv}\{x, (0, 0), (1, 0)\} \subset A$. See Fig. 3 for the set C and for an illustration of the remainder of the proof.

Choose numbers $0 < \alpha < \beta < 1$. To prove the lemma, it is sufficient to prove

$$\mu(R)(\text{conv}\{(\alpha, 0), (\beta, 0)\} \setminus \{(\alpha, 0), (\beta, 0)\}) = 0. \tag{8}$$

Let δ be an arbitrary number strictly between α and β and let d be the point in C with first coordinate δ. Also, let ℓ be the straight line through d and the point $(\alpha, 0)$.

Construct a preference $\bar{R}_2 \in \Omega$ with $p(\bar{R}_2) = p(R_2) = (1, 0)$ and such that the following two conditions are satisfied:

(i) At every point of intersection of an indifference curve of this preference with the interval $\text{conv}\{(0, 0), (1, 0)\}$ there is a line of tangency at this indifference curve parallel to ℓ.

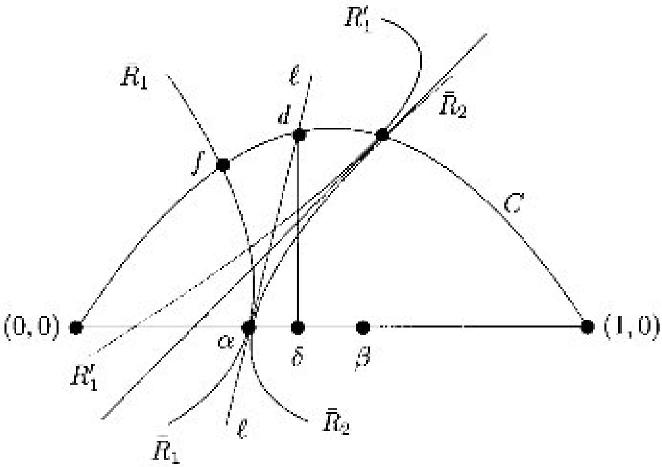

FIG. 3. Proof of Lemma 3.5.

(ii) At every point of intersection of an indifference curve of this preference with the curve C there is a line of tangency at this indifference curve which intersects the (relative) interior of $\text{conv}\{(0,0),(1,0)\}$.

For the indifference curves of $\bar{R}_2$ one may take a collection of (skewed) ellipses; an actual description is rather tedious and therefore omitted.

Next, choose a point f on the curve C between $(0,0)$ and the point d. We construct two preferences $\bar{R}_1$ and R_1' in Ω with $p(\bar{R}_1) = p(R_1') = (0,0)$ and such that the following three conditions are satisfied:

(iii) At every point of intersection of an indifference curve of $\bar{R}_1$ with the interval $\text{conv}\{(0,0),(1,0)\}$ there is a line of tangency at this indifference curve parallel to ℓ.

(iv) The indifference curve of $\bar{R}_1$ through $(\alpha, 0)$ passes also through the point f.

(v) At every point of intersection of an indifference curve of R_1' with the curve C, this indifference curve is tangential to the indifference curve of $\bar{R}_2$ through the same point.

For the construction of $\bar{R}_1$ one may take again (skewed) ellipses. The preference R_1' can be constructed in this way in particular because of condition (ii): the lines of tangency there separate $(0,0)$ and $(0,1)$.

By construction of these preferences, it follows that

$$PO(\bar{R}) = PO(\bar{R}_1, \bar{R}_2) = PO(R) = \text{conv}\{(0,0),(0,1)\} \text{ and } PO(R_1', \bar{R}_2) = C.$$

By strategy-proofness,

$$\mu(R_1', \bar{R}_2)(B((\alpha, 0), \bar{R}_1)) \leqslant \mu(\bar{R})(B((\alpha, 0), \bar{R}_1)). \tag{9}$$

In view of Lemma 3.4 we may write $\mu(B)$ for a subset B of C or of $\text{conv}\{(0,0), (1,0)\}$, suppressing reference to preferences as long as the peaks are $(0,0)$ for agent 1 and $(1,0)$ for agent 2, and the Pareto optimal sets are C or $\text{conv}\{(0,0), (1,0)\}$. Thus, (9) may be written as

$$\mu(\text{arc of } C \text{ between } (0,0) \text{ and } f) \leqslant \mu(\text{conv}\{(0,0), (\alpha, 0)\}). \tag{10}$$

Because this construction can be made for any point f on C between $(0,0)$ and d by adapting the choice of $\bar{R}_1$, (10) implies that

$$\mu(\{(\xi, \eta) \in C: 0 \leqslant \xi < \delta\}) \leqslant \mu(\text{conv}\{(0,0), (\alpha, 0)\}). \tag{11}$$

In an analogous way one can show

$$\mu(\{(\xi, \eta) \in C: \delta < \xi \leqslant 1\}) \leqslant \mu(\text{conv}\{(\beta, 0), (1,0)\}). \tag{12}$$

By (11), (12), and the fact that $\mu(C) = \mu(\text{conv}\{(0,0), (1,0)\}) = 1$ it follows that

$$\mu(\{d\}) \geqslant \mu(\text{conv}\{(\alpha, 0), (\beta, 0)\} \setminus \{(\alpha, 0), (\beta, 0)\}). \tag{13}$$

Because (13) can be derived for any point d on C with first coordinate strictly between α and β, the right-hand side of (13) must be equal to 0. Hence, (8) holds and the proof is complete. ∎

The preceding result is extended to arbitrary profiles in the following lemma. In particular, it follows that the mechanism must be "peak-only:" it depends only on the peaks of the reported preferences.

LEMMA 3.6. *Let* $R = (R_1, R_2) \in \Omega^N$. *Then* $\mu(R)(p(R_1)) + \mu(R)(p(R_2)) = 1$. *Moreover,* $\mu(R')(p(R_1')) = \mu(R)(p(R_1))$ *and* $\mu(R')(p(R_2')) = \mu(R)(p(R_2))$ *for any profile* $R' = (R_1', R_2') \in \Omega^N$ *with* $p(R_1') = p(R_1)$ *and* $p(R_2') = p(R_2)$.

Proof. Construct a preference $\bar{R}_1 \in \Omega^N$ with $p(\bar{R}_1) = p(R_1)$ and $PO(\bar{R}_1, R_2) = \text{conv}\{p(R_1), p(R_2)\}$: such a preference can easily be constructed because every line of tangency at an indifference curve of R_2 through a point of intersection with $\text{conv}\{p(R_1), p(R_2)\}$ obviously separates $p(R_1)$ and $p(R_2)$. See Fig. 4. Then, by Lemma 3.5, there are $0 \leqslant \alpha, \beta \leqslant 1$ with $\alpha + \beta = 1$ such that $\mu(\bar{R}_1, R_2)(p(R_1)) = \alpha$ and $\mu(\bar{R}_1, R_2)(p(R_2)) = \beta$. By strategy-proofness, for every $a \in PO(R)$ with $a \neq p(R_2)$:

$$\mu(R)(B(a, R_1)) \geqslant \mu(\bar{R}_1, R_2)(B(a, R_1)) = \mu(\bar{R}_1, R_2)(p(R_1)) = \alpha. \tag{14}$$

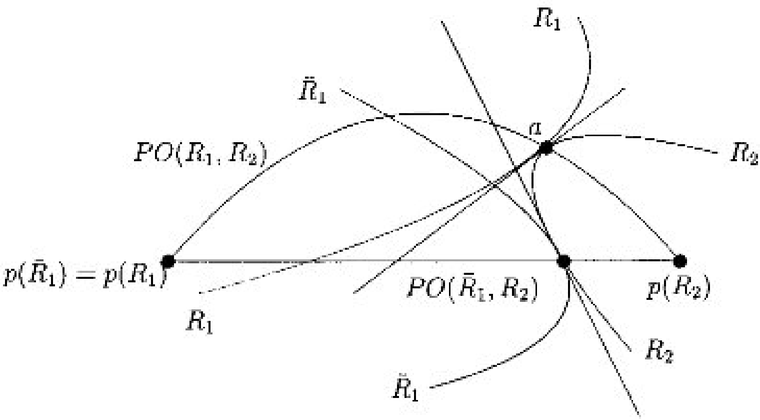

FIG. 4. Proof of Lemma 3.6.

Analogously one proves that for every $a \in PO(R)$ with $a \neq p(R_1)$:

$$\mu(R)(B(a, R_2)) \geqslant \beta. \tag{15}$$

By (14) and (15) it follows that $\mu(R)(p(R_1)) = \alpha$ and $\mu(R)(p(R_2)) = \beta$. This completes the proof of the lemma. ▌

An immediate consequence of Lemma 3.6 is the following proposition

PROPOSITION 3.1. *For every pair $x, y \in A$ with $x \neq y$ there are non-negative real numbers α, β, summing to 1, such that $\mu(R)(x) = \alpha$ and $\mu(R)(y) = \beta$ for all profiles $R \in \Omega^N$ with $p(R_1) = x$ and $p(R_2) = y$.*

The next proposition is the final step in the proof of Theorem 2.1.

PROPOSITION 3.2. *Let $\dim(A) \geqslant 2$, $n = 2$, and let $\mu: \Omega^N \to M(A)$ be a mechanism. Then μ is strategy-proof and Pareto optimal only if μ is a random dictatorship.*

Proof. Let $x, x', y \in A$ with $x \neq x'$ and let α, β and α', β' correspond to x, y and x', y, respectively, as in Proposition 3.1. It is sufficient to prove $\alpha = \alpha'$. (The case where the peak of agent 1 is fixed and that of agent 2 varies is similar.) We distinguish two cases.

Case (i). $y \notin \operatorname{conv}\{x, x'\}$.

In this case there are preferences R_1 and R'_1 in Ω with $p(R_1) = x$ and $p(R'_1) = x'$ and such that $y \notin B(x', R_1)$ and $y \notin B(x, R'_1)$. Let $R_2 \in \Omega$ with $p(R_2) = y$. Then by strategy-proofness

$$\begin{aligned}\alpha &= \mu(R_1, R_2)(B(x', R_1)) \\ &\geqslant \mu(R_1', R_2)(B(x', R_1)) \\ &= \alpha' \\ &= \mu(R_1', R_2)(B(x, R_1')) \\ &\geqslant \mu(R_1, R_2)(B(x, R_1')) \\ &= \alpha.\end{aligned}$$

Thus, $\alpha = \alpha'$.

Case (ii). $y \in \text{conv}\{x, x'\}$.

In this case, choose a point $x'' \notin \text{conv}\{x, x'\}$, which is possible because $\dim(A) \geqslant 2$. Let α'' and β'' be the numbers as in Proposition 3.1 corresponding to the pair x'', y. Note that $y \notin \text{conv}\{x, x''\}$ and $y \notin \text{conv}\{x', x''\}$. Therefore, applying Case (i) twice yields $\alpha = \alpha''$ and $\alpha' = \alpha''$, so that again $\alpha = \alpha'$. This completes the proof of Case (ii) and of the proposition. ∎

Theorem 2.1 now follows from Lemma 3.3 and the previous proposition.

4. THE n-AGENT CASE

This section is completely concerned with the proof of Theorem 2.1 for the general case with an arbitrary number of agents. This proof will be by induction, based on the two-person result established in the preceding section.

For later reference we first state a property and a result familiar in work on strategy-proofness. A mechanism $\mu: \Omega^N \to M(A)$ (where N is now arbitrary) is called *intermediate strategy-proof* if for all $S \subset N$, all $R \in \Omega^N$ with $R_i = R_j$ for all $i, j \in S$, all $R' \in \Omega^N$ with $R_i' = R_j'$ for all $i, j \in S$ and $R_i' = R_i$ for all $i \in N \setminus S$, and all $a \in A$:

$$\mu(R)(B(a, R_k)) \geqslant \mu(R')(B(a, R_k)) \qquad \text{for all} \quad k \in S.$$

Intermediate strategy-proofness means that if the members of a coalition share the same preference, then no one in the coalition can gain if the coalition collectively reports a different shared preference.

LEMMA 4.1. *A mechanism $\mu: \Omega^N \to M(A)$ is intermediate strategy-proof if, and only if, it is strategy-proof.*

Proof. The only-if part is obvious. For the if-part, assume that μ is strategy-proof and let S, R, R' be as in the definition of intermediate strategy-proofness. Take $k \in S$, without loss of generality suppose $S = \{1, ..., k\}$. Let $a \in A$. Then

$$\begin{aligned}\mu(R)(B(a, R_k)) &= \mu(R)(B(a, R_1)) \\ &\geqslant \mu(R_{-1}, R'_1)(B(a, R_1)) \\ &= \mu(R_{-1}, R'_1)(B(a, R_2)) \\ &\geqslant \mu(R_{-\{1,2\}}, R'_1, R'_2)(B(a, R_2)) \\ &\vdots \\ &\geqslant \mu(R_{-\{1,2,...,k\}}, R'_1, R'_2, ..., R'_i)(B(a, R_k)) \\ &= \mu(R')(B(a, R_k)),\end{aligned}$$

where all the inequalities follow from strategy-proofness. Hence, μ is intermediate strategy-proof. ∎

Whereas the induction basis for the proof of the general case is the two-person result of the preceding section, the induction assumption will be the following.

Assumption 4.1. Dim $(A) \geqslant 2$ and $n \geqslant 3$. For every set of agents I with cardinality $|I| < n$ and every mechanism $\mu: \Omega^I \to M(A)$ satisfying unanimity and strategy-proofness, μ is a random dictatorship.

In the following series of lemmas, until further notice Assumption 4.1 is valid, $N = \{1, 2, ..., n\}$, and $\mu: \Omega^N \to M(A)$ is unanimous and strategy-proof.

The first lemma considers the case where the preferences of two agents coincide. In that case the induction assumption implies random dictatorship.

LEMMA 4.2. *There are nonnegative real numbers $\lambda, \lambda_3, ..., \lambda_n$, summing to 1, such that for every $R \in \Omega^N$ with $R_1 = R_2$ and every $a \in A$:*

$$\mu(R)(a) = \lambda 1_{\{a\}}(p(R_1)) + \sum_{i=3}^{n} \lambda_i 1_{\{a\}}(p(R_i)). \tag{16}$$

Proof. In order to prove (16) define $\tilde{\mu}: \Omega^{N \setminus \{2\}} \to M(A)$ by

$$\tilde{\mu}(\tilde{R})(B) := \mu(\tilde{R}_1, \tilde{R}_1, \tilde{R}_3, ..., \tilde{R}_n)(B)$$

for every $\tilde{R} \in \Omega^{N \setminus \{2\}}$ and every Borel set B. Unanimity of μ implies unanimity of $\tilde{\mu}$ and intermediate strategy-proofness of μ (cf. Lemma 4.1) implies strategy-proofness of $\tilde{\mu}$. By Assumption 4.1, $\tilde{\mu}$ is a random dictatorship, which implies the existence of $\lambda, \lambda_3, \ldots, \lambda_n$ as in (16). ▌

The next lemma considers the same situation as in Lemma 4.2 and establishes a limited form of peak-onliness, namely for the case where agent 2 changes his preference but not his peak.

LEMMA 4.3. *Let $R \in \Omega^N$ with $R_1 = R_2$ and let $\lambda, \lambda_3, \ldots, \lambda_n$ as in Lemma 4.2. Let $r \in \Omega$ with $p(r) = p(R_2)$. Then for every $a \in A$,*

$$\mu(R_{-2}, r)(B(a, R_2)) = \lambda + \sum_{i=3}^{n} \lambda_i 1_{B(a, R_2)}(p(R_i)). \tag{17}$$

Proof. Let $a \in A$. Strategy-proofness of μ implies that

$$\mu(R)(B(a, R_2)) \geqslant \mu(R_{-2}, r)(B(a, R_2));$$

hence by (16)

$$\mu(R_{-2}, r)(B(a, R_2)) \leqslant \lambda + \sum_{i=3}^{n} \lambda_i 1_{B(a, R_2)}(p(R_i)). \tag{18}$$

On the other hand, because $R_1 = R_2$, strategy-proofness of μ also implies that

$$\mu(R_{-2}, r)(B(a, R_2)) \geqslant \mu(R_{-\{1,2\}}, r, r)(B(a, R_2)),$$

so by (16)

$$\mu(R_{-2}, r)(B(a, R_2)) \geqslant \lambda + \sum_{i=3}^{n} \lambda_i 1_{B(a, R_2)}(p(R_i)). \tag{19}$$

Now (17) follows from (18) and (19). ▌

Before proceeding we need to introduce some additional notation. For a profile $R = (R_1, \ldots, R_n) \in \Omega^N$ and an agent $i \in N$ we define an ordered partition $(N_1^i, \ldots, N_k^i)$ of N as follows. For all $i_j \in N_j^i$ and $i_\ell \in N_\ell^i$ with $j, \ell \in \{1, \ldots, k\}$, $j < \ell$, we have $p(R_{i_j})\, P_i\, p(R_{i_\ell})$. For all $j \in \{1, \ldots, k\}$ and all $i_j, i'_j \in N_j^i$ we have $p(R_{i_j})\, I_i\, p(R_{i'_j})$. In other words, N_1^i consists of those agents with peak equal to the peak of agent i (so in particular $i \in N_1^i$), N_2^i consists of those agents with peak different from but closest to the peak of agent i, ..., N_k^i consists of those agents with peak at maximal distance from the peak of agent i, everything measured according to the preference of agent i. Note that this partition depends on the preference R_i, but for brevity this dependence is suppressed in notation.

The following lemma extends the restricted form of peak-onliness established in (17). More precisely, it states that there is a random dictatorship whenever the peaks of at least two agents coincide.

LEMMA 4.4. *Let* $R \in \Omega^N$ *with* $\{1, 2\} \subset N_1^1$, *and let the numbers* $\lambda, \lambda_3, ..., \lambda_n$ *be as in Lemma 4.2. Then for every* $a \in A$,

$$\mu(R)(a) = \lambda 1_{\{a\}}(p(R_1)) + \sum_{i=3}^{n} \lambda_i 1_{\{a\}}(p(R_i)). \tag{20}$$

Proof. The proof of (20) will be by induction on $j = 1, ..., k$, where $(N_1^1, N_2^1, ..., N_k^1)$ is the ordered partition corresponding to R_1.

First, taking $a = p(R_1)$, (17) implies that

$$\mu(R)(p(R_1)) = \mu(R)(\{p(R_i) \colon i \in N_1^1\}) = \lambda + \sum_{i \in N_1^1 \setminus \{1, 2\}} \lambda_i \tag{21}$$

which proves (20) for $a = p(R_1)$, i.e., for the peak of the agents in N_1^1. This is the basis of the induction argument.

As induction hypothesis, let $j \in \{2, ..., k\}$, and suppose that (20) holds for every alternative a equal to some $p(R_i)$ for $i \in N_1^1 \cup \cdots \cup N_{j-1}^1$. We wish to show that (20) also holds for every alternative a equal to some $p(R_i)$ for $i \in N_1^1 \cup \cdots \cup N_j^1$.

Let $\ell' \in N_{j-1}^1$ and $\ell \in N_j^1$. For every $b \in A$ with $p(R_{\ell'}) P_1 b P_1 p(R_\ell)$ it follows by (17) and changing the roles of agents 1 and 2 there:

$$\mu(R)(B(b, R_1)) = \lambda + \sum_{i=3}^{n} \lambda_i 1_{B(b, R_1)}(p(R_i)) = \lambda + \sum_{i \in (N_1^1 \setminus \{1,2\}) \cup N_2^1 \cup \cdots \cup N_{j-1}^1} \lambda_i.$$

Hence, the total probability assigned to the upper contour set $B(b, R_1)$ is by the induction hypothesis equal to the total probability assigned to the peaks within that set. Thus, (20) holds for every $a \in A$ with $a P_1 p(R_\ell)$. Similarly, (17) implies that

$$\mu(R)(B(p(R_\ell), R_1)) = \lambda + \sum_{i \in (N_1^1 \setminus \{1,2\}) \cup N_2^1 \cup \cdots \cup N_j^1} \lambda_i.$$

Hence, denoting by $\partial B(p(R_\ell), R_1)$ the boundary (indifference curve) of the upper contour set $B(p(R_\ell), R_1)$, it follows that

$$\mu(R)(\partial B(p(R_\ell), R_1)) = \sum_{i \in N_j^1} \lambda_i.$$

In order to complete the induction step it suffices to show that for every $\bar{a} \in \partial B(p(R_\ell), R_1)$, if $\bar{a} \notin \{p(R_i): i \in N_j^1\}$ then $\bar{a} \notin \operatorname{supp}(\mu(R))$ and if $\bar{a} \in \{p(R_i): i \in N_j^1\}$ then $\mu(R)(\bar{a}) \leqslant \sum_{i \in N_j^1: p(R_i) = \bar{a}} \lambda_i$.

For the first case, suppose $\bar{a} \notin \{p(R_i): i \in N_j^1\}$ but $a \in \operatorname{supp}(\mu(R))$. Construct a preference $r \in \Omega$ with peak $p(r) = p(R_1) = p(R_2)$ and with some upper contour set $B(b, r)$ containing the peaks of all agents in $N_1^1 \cup \cdots \cup N_{j-1}^1$ but of no other agent, and also containing a neighborhood of $\bar{a}$ in which there is no peak and which has, say, weight $\varepsilon > 0$ under $\mu(R)$. By (17) it follows that

$$\mu(R_{-2}, r)(B(b, r)) = \lambda + \sum_{i \in (N_1^1 \setminus \{1,2\}) \cup N_2^1 \cup \cdots \cup N_{j-1}^1} \lambda_i.$$

Furthermore,

$$\begin{aligned} \mu(R)(B(b, r)) &= \mu(R_1, R_2, \ldots, R_n)(B(b, r)) \\ &\geqslant \lambda + \sum_{i \in (N_1^1 \setminus \{1,2\}) \cup N_2^1 \cup \cdots \cup N_{j-1}^1} \lambda_i + \varepsilon \\ &> \lambda + \sum_{i \in (N_1^1 \setminus \{1,2\}) \cup N_2^1 \cup \cdots \cup N_{j-1}^1} \lambda_i \\ &= \mu(R_{-2}, r)(B(b, r)), \end{aligned}$$

where the first inequality follows again by (17). This is a violation of strategy-proofness and, thus, completes the induction step for this case.

For the second and final case suppose $\bar{a} \in \{p(R_i): i \in N_j^1\}$ and $\mu(R)(\bar{a}) > \sum_{i \in N_j^1: p(R_i) = \bar{a}} \lambda_i$, say $\mu(R)(\bar{a}) = \sum_{i \in N_j^1: p(R_i) = \bar{a}} \lambda_i + \varepsilon$ where $\varepsilon > 0$. The proof for this case is analogous to the proof of the first case. Construct a preference $r \in \Omega$ with (as in the first case) $p(r) = p(R_1) = p(R_2)$ and containing the peaks of all agents in $N_1^1 \cup \cdots \cup N_{j-1}^1$, but also containing the alternative $\bar{a}$ (where the peak(s) of some agent(s) in N_j^1 is (are) located), and containing no other peaks. In a similar way as in the first case a violation of strategy-proofness is obtained. ∎

For convenience of notation, Lemmas 4.2, 4.3 and 4.4 have been formulated and proved for agents 1 and 2, but of course they can similarly be formulated and proved for arbitrary agents i and j. For further reference this is now stated as another lemma.

LEMMA 4.5. *Let $i, j \in N$ with $i \neq j$. Then there are nonnegative numbers λ_{ij} and λ_k for every $k \in N \setminus \{i, j\}$ with $\lambda_{ij} + \sum_{k \in N \setminus \{i,j\}} \lambda_k = 1$ such that for every $R \in \Omega^N$ with $p(R_i) = p(R_j)$ and every $a \in A$:*

$$\mu(R)(a) = \lambda_{ij} 1_{\{p(R_i)\}}(a) + \sum_{k \in N \setminus \{i,j\}} \lambda_k 1_{\{p(R_k)\}}(a).$$

Our next task is to show that the weights λ_{ij} and λ_k as in Lemma 4.5 are consistent in the sense that they are independent of the specific choice of agents i and j. Before we can do this we first need an auxiliary result on profiles with all peaks different.

LEMMA 4.6. *Let $R \in \Omega^N$ with all peaks different and $i, j \in N$ with $j \in N_2^i$. Let λ_{ij} and λ_k $(k \in N \setminus \{i, j\})$ be as in Lemma 4.5. Then for every $a \in A$ with $p(R_j) \in B(a, R_i)$*

$$\mu(R)(B(a, R_i)) = \lambda_{ij} + \sum_{k \in N \setminus \{i, j\}} \lambda_k 1_{B(a, R_i)}(p(R_k)). \tag{22}$$

Moreover,

$$\mu(R)(p(R_k)) = \lambda_k \qquad \textit{for every} \quad k \in N \setminus (N_1^i \cup N_2^i). \tag{23}$$

Proof. Let $a \in A$. By strategy-proofness (letting agent j deviate from R_i to R_j) and Lemma 4.5

$$\mu(R)(B(a, R_i)) \leqslant \mu(R_{-j}, R_i)(B(a, R_i)) = \lambda_{ij} + \sum_{k \in N \setminus \{i, j\}} \lambda_k 1_{B(a, R_i)}(p(R_k)).$$

On the other hand, also by strategy-proofness (letting agent i deviate from R_i to R_j) and Lemma 4.5,

$$\mu(R)(B(a, R_i)) \geqslant \mu(R_{-i}, R_j)(B(a, R_i)) = \lambda_{ij} + \sum_{k \in N \setminus \{i, j\}} \lambda_k 1_{B(a, R_i)}(p(R_k)).$$

These two inequalities imply (22).

The proof of (23) is analogous to the proof of Lemma 4.4, now using (22) with $a = p(R_j)$ instead of (21) as induction basis, and also using (22) here in the role of (17) there. We omit the details. ∎

The announced result on consistency of the weights is the following.

LEMMA 4.7. *There are nonnegative real numbers $\lambda_1, \lambda_2, \ldots, \lambda_n$, summing to 1, such that for all $i, j \in N$ with $i \neq j$ and all $R \in \Omega^N$ with $p(R_i) = p(R_j)$ and all $a \in A$:*

$$\mu(R)(a) = \sum_{k \in N} \lambda_k 1_{\{p(R_k)\}}(a). \tag{24}$$

Proof. We distinguish two cases: $n = 3$ and $n > 3$.

Case (i). $n = 3$.

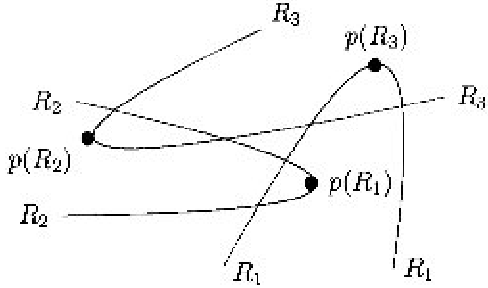

FIG. 5. Proof of Lemma 4.7.

Call the numbers as in Lemma 4.5 for the pairs $\{1, 2\}$; $\{2, 3\}$; $\{1, 3\}$ respectively: λ_{12}, λ_3; λ_{23}, λ_1; λ_{13}, λ_2. We show that $\lambda_i + \lambda_j = \lambda_{ij}$ for every pair $\{i, j\}$ with $i \neq j$, implying (24).

Consider a profile $R = (R_1, R_2, R_3) \in \Omega^N$ with all peaks different, with $N_2^1 = \{3\}$, $N_2^2 = \{1\}$, $N_2^3 = \{2\}$, and with (see Fig. 5)

$$B(p(R_3), R_1) \cap B(p(R_1), R_2) \cap B(p(R_2), R_3) = \emptyset. \tag{25}$$

By Lemma 4.6, in particular (23), applied three times (for the pairs $\{1, 2\}$, $\{2, 3\}$, and $\{1, 3\}$) it follows that $\mu(R)(p(R_i)) = \lambda_i$ every $i \in \{1, 2, 3\}$. Suppose $x \in \operatorname{supp}(\mu(R))$ for some $x \notin \{p(R_1), p(R_2), p(R_3)\}$. By Lemma 4.6, $\mu(R)(B(p(R_3), R_1)) = \lambda_{13} = 1 - \lambda_2$, so it follows that $x \in B(p(R_3), R_1)$. Similarly, one derives $x \in B(p(R_1), R_2)$ and $x \in B(p(R_2), R_3)$. This contradicts (25), hence $\operatorname{supp}(\mu(R)) = \{p(R_1), p(R_2), p(R_3)\}$, which implies $\lambda_i + \lambda_j = \lambda_{ij}$ for all $i \neq j$.

Case (ii). $n > 3$.

Consider pairs of agents $\{i, j\} \neq \{i', j'\}$, $i \neq j$, $i' \neq j'$, and call the corresponding numbers as in Lemma 4.5: λ, λ_k $(k \neq i, j)$ for the pair $\{i, j\}$ and λ', λ'_k $(k \neq i', j')$ for the pair $\{i', j'\}$.

Suppose $k \notin \{i, j\} \cup \{i', j'\}$ then $\lambda_k = \lambda'_k$ which can be seen by considering a profile R with $p(R_i) = p(R_j)$, $p(R_{i'}) = p(R_{j'})$, and $p(R_k) \neq p(R_\ell)$ for all $\ell \in N \setminus \{k\}$. To complete the proof it is sufficient to show that $\lambda_i + \lambda_j = \lambda_{ij}$. To show this assume additionally $i', j' \notin \{i, j\}$ (this is possible since $n > 3$) and take a profile R' with $p(R'_i) = p(R'_j) \neq p(R'_{i'}) = p(R'_{j'})$ and with $n-2$ different peak locations. Then Lemma 4.5 applied twice yields $\lambda_i + \lambda_j = \lambda_{ij}$. ∎

The next and final lemma basically completes the induction step in the proof of the only-if part of Theorem 2.1.

LEMMA 4.8. *Let $\lambda_1, ..., \lambda_n$ be the weights as in Lemma 4.7. Then for every profile $R \in \Omega^N$ and every agent $i \in N$, $\mu(R)(p(R_i)) \geqslant \lambda_i$.*

Proof. Let $R \in \Omega^N$ and (without loss of generality) $i = 1$. We also assume that $p(R_1)$ is in the relative interior of the set A (we use this assumption in Case 2 below). The general case follows immediately by strategy-proofness.

If not all peaks in R are different the proof is done by Lemma 4.7. So assume that all peaks are different. We distinguish three cases.

Case 1. There is an agent $j \in N$ with $1 \notin N_2^j$.

In this case Lemma 4.6, in particular (23), implies (even) $\mu(R)(p(R_1)) = \lambda_1$.

Case 2. There is no agent j as in Case 1, but there is a pair $\{j, k\}$ with $j \neq k$ and $1, k \in N_2^j$.

In this case, consider a profile (R_{-1}, R_1') with $p(R_1') \notin B(p(R_k), R_j)$ and $p(R_1')$ close to $p(R_1)$. (Such a preference R_1' exists because by assumption $p(R_1)$ is in the relative interior of the set A.) By Case 1, for this profile $\mu(R_{-1}, R_1')(p(R_1')) = \lambda_1$. Because $p(R_1')$ can be chosen as close to $p(R_1)$ as desired, strategy-proofness implies $\mu(R)(p(R_1)) \geqslant \lambda_1$.

Case 3. For every $j \in N \setminus \{1\}$ we have $I_2^j = \{1\}$.

In this case, Case 1 (for $i = j$ instead of $i = 1$) applies to very agent $j \neq 1$, so that $\mu(R)(p(R_j)) = \lambda_j$ for every $j \neq 1$. Moreover, Lemmas 4.6 and 4.7 imply that $\mu(R)(B(p(R_1), R_j)) = \lambda_j + \lambda_1$ for every $j \neq 1$. Take two arbitrary agents $\neq 1$, say agents 2 and 3. We distinguish two subcases.

Case 3(a). $p(R_3)$ is not on the straight line through $p(R_2)$ and $p(R_1)$.

We first prove the following claim (cf. Fig. 6).

Claim. For every $\varepsilon > 0$ there is a set $B(\varepsilon)$ within an ε-neighbourhood of the line segment $\text{conv}\{p(R_1), p(R_2)\}$ and a preference R_1^ε with $p(R_1^\varepsilon) = p(R_1)$ such that $\mu(R_{-1}, R_1^\varepsilon)(B(\varepsilon)) = \lambda_1 + \lambda_2$ and $\mu(R_{-1}, R_1^\varepsilon)(p(R_2)) = \lambda_2$.

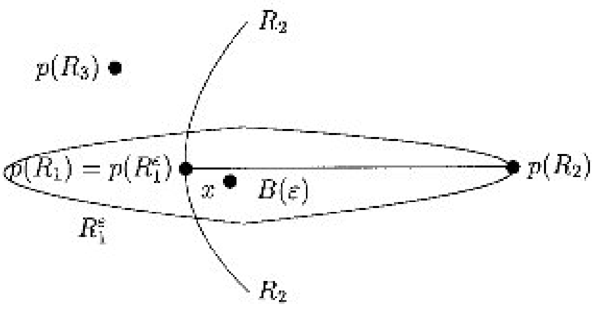

FIG. 6. Proof of Lemma 4.8.

Proof of Claim. Let $\varepsilon > 0$ and construct a preference R_1^ε such that, first, $B(\varepsilon) := B(p(R_2), R_1^\varepsilon) \cap B(p(R_1), R_2)$ is within an ε-neighborhood of $\text{conv}\{p(R_1), p(R_2)\}$, and, second, $p(R_2)\, P_1^\varepsilon\, p(R_j)$ for all $j \neq 1, 2$. It still holds that $\mu(R_{-1}, R_1^\varepsilon)(B(p(R_1), R_2)) = \lambda_1 + \lambda_2$ and $\mu(R_{-1}, R_1^\varepsilon)(p(R_2)) = \lambda_2$ because these results were independent of the (shape of the) preference of agent 1. Moreover, Lemma 4.6 implies $\mu(R_{-1}, R_1^\varepsilon)(B(p(R_2), R_1^\varepsilon)) = \lambda_1 + \lambda_2$. Hence, $\mu(R_{-1}, R_1^\varepsilon)(B(\varepsilon)) = \mu(R_{-1}, R_1^\varepsilon)(B(p(R_2), R_1^\varepsilon) \cap B(p(R_1), R_2)) = \lambda_1 + \lambda_2$ and $\mu(R_{-1}, R_1^\varepsilon)(p(R_2)) = \lambda_2$. ▮

Now consider the situation as in the claim, for some $\varepsilon > 0$. Take $x \in B(\varepsilon)$ such that

$$p(R_1) \notin \text{conv}\{x, p(R_2), p(R_3)\}. \tag{26}$$

Suppose that $x \in \text{supp}(\mu(R_{-1}, R_1^\varepsilon))$. Construct a preference R_2' with peak $p(R_2') = p(R_2)$ and with some upper contour set $B(b, R_2')$ (for some $b \in A$) containing a neighbourhood of x, the peak $p(R_3)$ (and possibly some peaks of agents 4, ..., n) but not $p(R_1)$. By Lemma 4.6

$$\mu(R_{-\{1,2\}}, R_1^\varepsilon, R_2')(B(b, R_2')) = \sum_{j \in N:\, p(R_j) \in B(b, R'_2)} \lambda_j$$

and by Lemma 4.6 and the assumption that $x \in \text{supp}(\mu(R_{-1}, R_1^\varepsilon))$ it follows that

$$\mu(R_{-1}, R_1^\varepsilon)(B(b, R_2')) > \sum_{j \in N:\, p(R_j) \in B(b, R'_2)} \lambda_j.$$

This is a violation of strategy-proofness, hence $x \notin \text{supp}(\mu(R_{-1}, R_1^\varepsilon))$. This holds for all $x \in B(\varepsilon)$ for which (26) holds, that is, for all x except in a neighborhood of $p(R_1)$. Because this neighborhood shrinks to $p(R_1)$ as ε goes to zero, strategy-proofness (for agent 1) implies that $\text{supp}(\mu(R)) \cap B(p(R_1), R_2) = \{p(R_1), p(R_2)\}$. Consequently, $\mu(R)(p(R_1)) = \lambda_1$.

Case 3(b). $p(R_3)$ is on the straight line through $p(R_1)$ and $p(R_2)$.

In this case, consider a preference $\tilde{R}_1^\varepsilon$ with peak at distance at most $\varepsilon > 0$ from $p(R_1)$ and such that $p(R_3)$ is *not* on the straight line through $p(R_2)$ and $p(\tilde{R}_1^\varepsilon)$. For $\varepsilon > 0$ sufficiently small Case 3(a) applies so that $\mu(R_{-1}, \tilde{R}_1^\varepsilon)(p(\tilde{R}_1^\varepsilon)) = \lambda_1$. By letting ε decrease to 0 and applying strategy-proofness it follows that $\mu(R)(p(R_1)) \geqslant \lambda_1$.

This completes the proof of Lemma 4.8. ▮

Proof of Theorem 2.1 for arbitrary n. The if-part is obvious. The only-if part follows by induction, based on the result of Section 3 for $n = 2$, Assumption 4.1, and Lemmas 4.7 and 4.8. ▮

5. THE DETERMINISTIC CASE

A deterministic mechanism is a map $F: \Omega^N \to A$. Such a mechanism assigns a unique alternative to every preference profile. It is called strategy-proof if for all $R \in \Omega^N$, all $i \in N$, and all $R'_i \in \Omega$ we have $F(R)\, R_i F(R_{-i}, R'_i)$. It is called unanimous if $F(r, ..., r) = p(r)$ for every $r \in \Omega$. With a deterministic mechanism F we can associate a probabilistic mechanism μ_F by defining $\mu_F(R)(F(R)) = 1$ for every $R \in \Omega^N$. It is easy to check that (deterministic) unanimity and strategy-proofness of F are inherited by μ_F in their probabilistic meaning as in the preceding part of this paper.

From Theorem 2.1 the following result can be derived.

THEOREM 5.1. *Let* $\dim(A) \geqslant 2$ *and let* $F: \Omega^N \to A$ *be a deterministic surjective mechanism. Then F is strategy-proof if and only if there is an agent* $i \in N$ *with* $F(R) = p(R_i)$ *for every* $R \in \Omega^N$.

This theorem is close to the main result in Zhou [19]. One difference is that here we require the range of F to be the set A and hence to be convex —in accordance with the rest of this paper—whereas Zhou [19] imposes only the dimensionality condition on this range. For a probabilistic mechanism μ Theorem 2.1 imposes unanimity—it is not obvious how this could be replaced by such a dimensionality condition. Another difference is that Zhou allows for the larger set of continuous convex not necessarily single-peaked preferences.

Proof of Theorem 5.1. The if-part is obvious. Now assume F is strategy-proof. We first argue that F is also unanimous. To see this, take $r \in \Omega$ with peak $p(r) = a$, and take $R \in \Omega^N$ with $F(R) = a$ (which is possible by surjectivity). By letting the agents change from R to $(r, ..., r)$ one by one and each time applying strategy-proofness it follows that $F(r, ..., r) = a = p(r)$. Hence, F is unanimous. Therefore, the associated probabilistic mechanism μ_F is strategy-proof and unanimous, as mentioned above, and hence μ_F is a random dictatorship by Theorem 2.1. Because μ_F always assigns probability 1 to exactly one alternative, the statement of the theorem follows. ∎

6. CONCLUDING REMARKS

Like earlier papers on probabilistic social choice, the present paper also reveals an analogy between results on deterministic and on probabilistic mechanisms. Again in accordance with earlier results, the main result of this paper does not seem easily deducible from any corresponding deterministic result. An attempt to formalize this apparent analogy between the two approaches is made in Keiding [10] by using category theory. While indeed a relation can be established in this manner it is not clear that this also leads to easier or shorter proofs.

We conjecture, finally, that the main result of this paper would still hold if the set of admissible preferences were further restricted to only quadratic (ellipsoid) preferences. However, since this would be obtained at the cost of some tedious calculations, we have refrained from pursuing this extension.

REFERENCES

1. S. Barbera, Nice decision schemes, *in* "Decision Theory and Social Ethics: Issues in Social Choice" (H. W. Gottinger and W. Leinfellner, Eds.), D. Reidel, Dordrecht, 1978.
2. S. Barbera, Majority and positional voting in a probabilistic framework, *Rev. Econ. Stud.* **46** (1979), 379–389.
3. K. Border and J. Jordan, Straightforward elections, unanimity and phantom voters, *Rev. Econ. Stud.* **50** (1983), 153–170.
4. J. Duggan, A geometric proof of Gibbard's random dictatorship theorem, *Econ. Theory* **7** (1996), 365–369.
5. L. Ehlers, H. Peters, and T. Storcken, Probabilistic collective decision schemes, Research Memorandum, University of Maastricht, 1998. Revised version, October 2000. *J. Econ. Theory*, doi:10.1006/jeth.2001.2829.
6. A. Gibbard, Manipulation of voting schemes: A general result, *Econometrica* **41** (1973), 587–601.
7. A. Gibbard, Manipulation of schemes that mix voting with chance, *Econometrica* **45** (1977), 665–681.
8. A. Gibbard, Straightforwardness of game forms with lotteries as outcomes, *Econometrica* **46** (1978), 595–614.
9. A. Hylland, Strategyproofness of voting procedures with lotteries as outcomes and infinite sets of strategies, unpublished, 1980.
10. H. Keiding, the categorical approach to social choice theory, *Math. Soc. Sci.* **1** (1981), 177–191.
11. K. H. Kim and F. W. Roush, Nonmanipulability in two dimensions, *Math. Soc. Sci.* **8** (1984), 29–43.
12. M. Le Breton and A. Sen, Separable preferences, strategy-proofness and decomposability, *Econometrica* **67** (1999), 605–628.
13. D. Moreno, Strategyproof allocation mechanisms for pure public goods economies when preferences are monotonic, *Econ. Theory* **13** (1999), 183–197.
14. H. Moulin, On strategyproofness and singlepeakedness, *Public Choice* **35** (1980), 437–455.
15. S. Nandeibam, An alternative proof of Gibbard's random dictatorship result, *Soc. Choice Welfare* **15** (1998), 509–519.
16. H. Peters, H. Van der Stel, and T. Storcken, Pareto optimality, anonymity and strategy-proofness in location problems, *Int. J. Game Theory* **21** (1992), 221–235.
17. M. Satterthwaite, Strategy-proofness and Arrow's conditions: Existence and correspondence theorem for voting procedures and social choice functions, *J. Econ. Theory* **10** (1975), 187–217.
18. Y. Sprumont, Strategyproof collective choice in economic and political environments, *Can. J. Econ.* **28** (1995), 68–107.
19. L. Zhou, Impossibility of strategyproof mechanisms in economies with pure public goods, *Rev. Econ. Stud.* **58** (1991), 107–119.