Democratically Elected Aristocracies

Authors: David Heyd, Uzi Segal

Boston College Working Papers in Economics, 2002

Originally posted on: http://ideas.repec.org/p/boc/bocoec/529.html

Democratically Elected Aristocracies*

David Heyd† and Uzi Segal‡

March 18, 2002

Abstract

The article suggests a formal model of a two-tier voting procedure, which unlike traditional voting systems does not presuppose that every vote counts the same. In deciding a particular issue voters are called in the first round to assign categories of their fellow-citizens with differential voting power (or weights) according to the special position or concern individuals are perceived as having with regard to that issue. In the second stage, voters vote on the issue itself according to their substantive view and their votes are counted in the light of the differential weights assigned in the first round. We analyze the formal and the philosophical reasons that support the model.

*We wish to thank Brian Barry, Eddie Dekel, Drew Fudenberg, and Ariel Rubinstein for helpful comments.

†Department of Philosophy, The Hebrew University, Jerusalem 91905, Israel. E-mail: david.heyd@huji.ac.il

‡Department of Economics, Boston College, Chestnut Hill MA 02467. E-mail: segalu@bc.edu

1 Introduction

Voting is a procedure that is applied to issues that call for a collective decision. Its principal attraction lies in its being a decision-making procedure through which the integrity of a group can be maintained despite disagreement among its members about the correct or desirable way in which substantive issues should be settled. And unlike other procedures of collective choice, like lottery, compromise or the exercise of sheer power, majority vote is not arbitrary, ad hoc or oppressive. In its logical structure, voting cannot be fully reflexive, i.e. its procedural conditions as well as the formulation of the issue to be decided must be antecedently given rather than put to a vote. However, some, even if not all of its rules, may be decided by voting.

In this article we suggest a model for such a partially reflexive application of voting, which offers a way of fine-tuning traditional majoritarian procedures. We are particularly concerned with the failure of traditional voting methods to pay tribute to the differential weight people often believe should be assigned to different voters.[1] We therefore suggest the following formal model. Members of society are asked to rank possible subsets of society, where $A \succ_\alpha^j B$ means that person α prefers the subset A of individuals over the subset B to decide issue j for society. Under some assumptions we conclude that these preferences can be represented by a function V in the following way. Each member of society is a assigned a certain weight, and $V(A)$ is obtained by taking the sum of these weights over all members of the set A (see Theorem 1 in Section 3). Although (assuming that all weights are non-negative) the best subset would be the whole of society, we argue that the interpretation of the model in terms of the relative weight of different categories of people can still be maintained by assigning individuals different voting powers that are proportional to the weights obtained in Theorem 1.

Next we deal with social aggregation of individual preferences. In Section 4 we offer axioms implying that society will assign each individual member the average weight individual members of society think he should be awarded regarding this issue. These axioms also imply that society will decide the issue itself on the basis of the votes cast in the second stage and counted in the light of the outcome of the first vote.

[1] Since our model permits zero weights, it relates also to the question of the scope of the voting group, that is, who should take part in the vote.

In Section 5 we analyze the case where the weights one person wishes to assign other members of society in one issue depend on the weights assigned to them in other issues. A simple continuity assumption implies the existence of a multi-issue system of weights. In Section 6 we discuss some possible objections to the model, and in Section 7 we offer some remarks on the literature. All proofs appear in the Appendix.

2 The Two-Tier Voting Model

One reason for the famous Arrovian impossibility result is that the input of the model, the individual rankings, reveal only ordinal rather than cardinal preferences. The reason this omission creates a problem is clear. Social ranking must aggregate and average conflicting individual rankings, but ordinal preferences do not provide us with relevant information about the intensity of preferences.

The economic literature offers at least two kinds of cardinal preferences that can be used for interpersonal comparisons. The first is Harsanyi's [7, 8, 9], in which preferences are represented by vN&M utility functions. The second utilizes quasi linear functions and uses money as a measure of transferable utility.

Harsanyi [8] extends the set of possible social policies by introducing lotteries over these policies. Allocations of medical treatment or of army duty fit into this framework, but so do allocations of divisible goods. Individuals and society have preferences over these lotteries and a Pareto assumption links the selfish and the social preferences: If all individuals prefer one social lottery to another, then so does society. Assuming that all preferences over uncertain outcomes satisfy the axioms of expected utility theory, Harsanyi proves that social preferences can be represented by a weighted sum of individual vN&M utilities.

Quasi linear utilities are widely used in the analysis of public goods. It is well known that if all individuals have a utility of the form $m+u(x)$ (where x is the public good and m is money), then the efficient quantity of the public good is obtained at the point where $\sum u_i'(x) = c'(x)$, where c is the cost function.

There are situations in which both methods seem unsuitable. Consider issues like a: abortion rights, b: freedom of expression, and c: ban on male

circumcision. Suppose a person supports all three (that is, he is in favor abortion rights and freedom of expression, but opposes male circumcision), and in that order. It is not clear how he can answer the question: What p makes you indifferent between "$(a, \neg b, \neg c)$ with probability p and $(\neg a, \neg b, c)$ with probability $1-p$" and "$(\neg a, b, \neg c)$." It is also not clear that individuals would be willing to compromise their convictions for money. In other words, both standard cardinalizations of preferences cannot be applied here.

The present model compares individual attitudes towards controversial issues not only by the intensity of individual preferences (as is the case in utilitarianism and quasi linear functions) but also by the average weight members of society are willing to give to each other's preferences. These weights may reflect people's willingness to rely on the privileged insight of some of their fellow-citizens,[2] but they are also the result of people's realization that some members of society feel more strongly than others about some issues and that this should be taken into account in the social choice. As a tool of interpersonal comparisons, the present model agrees with some recent social choice models in which social concerns become part of each person's characteristics (see Estlund [4], Wolff [15], Segal [14], Karni and Safra [12], and Karni [11]). But it differs from these models in one important aspect: The tool that is used for interpersonal comparisons is external and not internal. That is, what is compared is not how individuals feel about the issue, but how *other* people feel about these individuals. We discuss the rationale for this tool in Section 6 below.

The inclusion of the other-regarding concern for the way people consider a controversial issue breaks the atomistic structure of the one-phase vote and expresses social solidarity, which is after all the presupposition and the aim of all procedures of social choice under circumstances of disagreement. Living in a community rather than in an arbitrary aggregate of detached individuals means that the question how much should one person's preferences or beliefs weigh cannot be determined independently of what everyone thinks of that question.

A two-tier procedure is attractive in contexts in which voters might have reasons for assigning extra weight to particular categories of people on the

[2]Consider the re-construction of downtown Manhattan. In a city-wide referendum some voters might feel that although they have their own views about the right way to go about it, residents of the re-designed area should be given an extra vote which would express their closer familiarity with the complexity of the the issue.

basis of their alleged privileged position, moral standing, or particular sensitivity to the outcome of the substantive decision.[3] The procedure relates to issues about which there is not only first-order disagreement regarding the right answer but also a second-order dispute concerning the kind of issues they are or the kind of people who should be entrusted to deal with them. Thus, in market-like situations, in which individuals make choices exclusively according to what would satisfy them most (and regard others as behaving on a similar basis), a two-phase system makes no sense, since individuals are expected to assign equal weights to all members of society. But then many social choices are not of this nature, for they often involve moral or ideological views about the differential standing of members of the group with regards to the measure to be decided. That is to say, they involve some kind of an evaluative, moral judgment of people's preferences.[4]

Take, for instance, abortions. Some may wish to give women more weight than men because of the particular position of the pregnant woman with regard to her own body. Others may give everybody an equal vote on that matter on the basis, for example, of their view that the decisive issue is whether the fetus is a human person rather than how the interests of the pregnant woman are affected. Or, one might take a different view according to which theologians (or physicians) should be given extra weight. Another example relates to funds that are transferred from the rich to the poor. Some might hold the view that those who gave the money should have a particular say on the way it is distributed among the needy, while others might believe that the question should be left to the recipients, who know best what they need. These are not necessarily questions of self-interest, since people who are neither on the giving nor on the receiving end may nevertheless have strong views on the matter. In a democratic procedure, we claim, this second-order

[3]The model we are offering here is abstract and idealized and should not be understood as a proposal for electoral reform. We are aware of the difficulties in its actual implementation, particularly of the question of the categorization of individuals, which might be associated with stigmatization and profiling. The fact that a gives b a voting power as a person of a certain type does not mean that b wants to identified as such a type.

[4]Frankfurt [5] claims that beyond their first-order desires and preferences, individuals also have second-order evaluations and rankings of these first-order desires, rankings which are not based merely on the strength or intensity of the desires. One's moral self-identity is defined in terms of those normative assessments of the relative force of one's desires. Our model might be understood as an inter-personal analogue of Frankfurt's theory of intra-personal two-tier judgments.

disagreement should also be democratically settled.[5]

The presentation of our model will benefit from setting it on the background of its two major alternatives: the aristocratic and the democratic. The first consists of a voting procedure that includes only a subset of the group within which the social choice is applied. This subset, endowed with the voting power, may consist of a special class of individuals like priests, noblemen, men, people with some income or property, professionals, or even, in the limiting case, one individual who happens to be blessed with certain unique qualities. Aristocracy in the historical sense, oligarchy, professional committees and dictatorship belong to this category. The second, democratic model consists of the notion that everybody takes part in the vote and resents the idea of any subgroup in society making decisions for the whole group.

The reasoning behind the aristocratic model is that not everybody in the group is equally positioned to take part in the decision-making procedure. There are people who are able to understand the issue at hand and those who lack that ability; or, there are individuals whose interests matter and those whose interest do not count. Condorcet [1], for instance, thought that there are some matters in which voting should aim at the true or correct answer, and if that is so, majority vote could be effective only if we limit the scope of voters to those whose average probability to get the right answer is over $\frac{1}{2}$. Another example relates to the limitation of the vote to men of property. Here the idea is that the subgroup consists of those whose interests matter more, either because of their gender identity or because they are the ones who pay for the policies that stand to be decided.

The justification of the democratic model appeals to the egalitarian idea of the inherent value of every individual as a human being, irrespective of any contingent attribute or particular position, and to the general skepticism regarding the claim to a privileged access to truth by any class of people. Thus, everybody's interest should be counted equally and no category of people should be assumed to have better standing or knowledge, either about the nature of the true interests of others or about values in general.

The model offered here combines elements of both the aristocratic and

[5] Voters might also want to assign differential weights to categories of voters on the basis of their epistemic authority or privileged knowledge concerning the matter at hand. Thus, one might want to give extra weight to both researchers and members of Humane Society on the issue of experimentation on animals, or in some contexts assign zero weight to those who know nothing about the subject, oneself included.

the democratic models, both in its formal structure and in the substantive reasons supporting it. Like the aristocratic model, our approach accepts the notion of the differential standing of individuals regarding the particular issue at hand, since on some subjects certain people are thought to have interests that count more for various normative reasons, or since they are held as more knowledgeable and able to form judgment. But since the model is skeptical about the possibility of an ideal external point of view from which the privileged subgroup(s) can be identified, it leaves that identification to the democratic process. And rather than draw from that skepticism the conclusion that everyone should be given an equal say on each matter on the agenda, it lets the differential weights be assigned by the voters themselves. Theoretically, voters may choose one of the extreme, limiting cases: either give everyone an equal vote, or universally consent to give one individual an exclusive power to decide the matter. But again, these apparently democratic and dictatorial choices are based on an actual democratic consent rather than on an independent abstract principle.

Our idea can be dubbed "democratically elected aristocracies." But to avoid any misunderstanding it should be noted first that unlike traditional aristocracy, everybody in the group is (usually) given the vote in the first round, albeit with differential weight. Secondly, the issues on which subgroups are elected to vote are highly specific and their scope is limited, since — unlike real historical aristocratic regimes — the privilege of a particular subgroup does not run "across the board." There are no privileged members of society; only members who are given a special position regarding particular social choices. In applying to the whole spectrum of social choices, both the aristocratic and the democratic alternatives to our suggested model fail to acknowledge that some individuals may have a stronger say on some matters, while others have more authority on other matters.

From a political-theory point of view here lie both the attraction and the limitation of the suggested model. It is typically issue-oriented and, like referendums, provides a representation of the people's views on those ideological and moral problems that people believe should be left out of the bargaining table of ordinary politics. But then bargaining, logrolling and coalitions are the stuff of politics in its rudimentary sense. Democracy does not only attempt to represent people's positions on specific issues but rather to supply a framework for the exercise of power by "the people." Our model should not be used in the context of the election of representatives,

parliamentary parties or public officials, since in such elections the democratic ideal is essentially egalitarian and leaves no room for differential weights. Political power lies in the capacity to control the outcome of a wide range of issues and hence should be allocated equally; but positions and attitudes on specific issues may be subjected to differential evaluation.[6]

3 Individual Ranking

In the informal presentation of the model for a two-phase vote in the previous section we argued for the principle that everybody in society takes part in the vote, both in the first phase (the assignment of weights) and in the second (the decision on the substantive issue). In this section we seek a formal articulation of the weights. For that purpose we propose a theoretical exercise, which we argue is compatible with the substantive (informal) model. Our methodological claim is that from a theoretical model of ranking subsets of society as the preferred groups for making a decision on a given issue we can derive the idea of the relative weights assigned to different categories of people in real-life contexts.

We assume that society is composed of a continuum of agents, say $[0,1]$. Consider a question that fits our domain, for example, abortion rights, the right of male circumcision, etc. (but not "how to divide a cake," and not "who is a member of society.") Each member of society partitions society into subsets which he considers relevant in this context. So for example, such a division may be "men and women," or it may be "secular and religious people." Although we assume an infinite number of members of society, we also assume that there is only a finite number of relevant partitions of society. There is therefore a finite partition that is finer than all individual partitions. Denote it by $\mathcal{S} = \{S_1, \ldots, S_N\}$ and assume that each S_i is a measurable set. For a measurable subset A of $[0,1]$, let $\{A_1, \ldots, A_N\}$ be the set of the intersections of A with the partition $\mathcal{S}$.

Each member α of society has complete and transitive preferences $\succeq_\alpha$ over measurable subsets of S, where $A \succeq_\alpha B$ means "person α prefers that

[6] The broad distinction between election of representatives and referendums on issues leaves open the further question whether there are issues which should never be put to any democratic vote (like, for example, human rights) and whether there are issues which may be put to a vote but only in a one-tier method (like the election of representatives).

group A of individuals decide for society over group B making this decision." Consider the following assumptions on $\succeq_\alpha$. For brevity, we omit the subscript α. μ is the Lebesgue measure on $[0,1]$.

Scaling If there exists $\lambda > 0$ such that for all $i = 1, \ldots, N$, $\mu(A'_i) = \lambda\mu(A_i)$ and $\mu(B'_i) = \lambda\mu(B_i)$, then $A' \succeq B'$ if, and only if, $A \succeq B$.

Residual Scaling If there exsits $\lambda > 0$ such that for all $i = 1, \ldots, N$, $\mu(S_i \backslash A'_i) = \lambda\mu(S_i \backslash A_i)$ and $\mu(S_i \backslash B'_i) = \lambda\mu(S_i \backslash B_i)$, then $A' \succeq B'$ if, and only if, $A \succeq B$.

Complete Separability If $\mu(A_{i^*}) = \mu(B_{i^*})$, $\mu(A'_{i^*}) = \mu(B'_{i^*})$, and for $i \neq i^*$, $\mu(A_i) = \mu(A'_i)$, and $\mu(B_i) = \mu(B'_i)$, then $A \succeq B$ if, and only if, $A' \succeq B'$.

Continuity If for all i, $\mu(A^k_i) \to \mu(A_i)$, $\mu(B^k_i) \to \mu(B_i)$, and for all k, $A^k \succeq B^k$, then $A \succeq B$.

Strict Monotonicity $A \subsetneq B$ implies $B \succ A$.

Monotonicity $S \succ \varnothing$.

The first assumption suggests that if a person α prefers subset A to make a decision for the whole society over subset B making this decision, and if A' and B' are λ-replicas of A and B with respect to partition $\mathbb{S}$ of $[0,1]$, then he should also prefer A' to B'. Note that this assumption implies in particular that if for all i, $\mu(A_i) = \mu(B_i)$, then $A \sim B$.

The second assumption suggests looking into the subsets of those individuals who are excluded from the decision making process. The rationale is the same as before. If someone prefers a certain group A over B to make a social decision, it also means that he prefers that $S\backslash A$ will be excluded from the decision making procedure to $S\backslash B$ to be thus excluded. Of course, A and $S\backslash A$ fully determine each other, but concentrating on each represents different points of view. The scaling and the residual scaling assumptions highlight this duality.

The third assumption compares two sets A and B that have the same size of individuals of type i^*. The preferences between these two sets do not change when the common size of the set of individuals of type i^* changes, provided the modified A and B still have the same size of type-i^* individuals.

The preferences $\succeq$ induce preferences over $X = \prod_{i=1}^{N}[0, \mu(S_i)]$, so wlg we will use the same notation $\succeq$. Let $p = (\mu(S_1), \ldots, \mu(S_N))$. Consider a set of the form $X(I, a^*) = \{a \in X : \forall i \in I, a_i = a_i^*\}$, that is, $X(I, a^*)$ is the set of sets where the size of social sections in I is fixed at the a_i^* level. For $a \in X$, define $a(I, a^*)$ by $a_i(I, a^*) = a_i$ for $i \notin I$, and $a_i(I, a^*) = a_i^*$ for $i \in I$. Consider the following two conditions in which we apply the logic of the scaling and residual scaling assumption to the constrained set $X(I, a^*)$.

(I, a^*)–Scaling Let $a, b \in X(I, a^*)$. For all $\lambda > 0$, $a \succeq b$ if, and only if, $\lambda a(I, a^*) + (1 - \lambda)0(I, a^*) \succeq \lambda b(I, a^*) + (1 - \lambda)0(I, a^*)$.

(I, a^*)–Residual Scaling Let $a, b \in X(I, a^*)$. For all $\lambda > 0$, $a \succeq b$ if, and only if, $\lambda a(I, a^*) + (1 - \lambda)p(I, a^*) \succeq \lambda b(I, a^*) + (1 - \lambda)p(I, a^*)$.

We show in the proof of Theorem 1 in the appendix that these two axioms follow form scaling, residual scaling, and complete separability.

The geometric difference between complete separability and the last two axioms is clear. The former imposes no restrictions on the preferences $\succeq$ when one coordinate is fixed, but connects together such preferences for different levels of the fixed coordinate. The latter axioms do not impose any connection between the orders obtained for different levels of the fixed coordinates, but impose restrictions on the induced orders themselves.

Consider the following possible objection to complete separability and the (I, a^*)–scaling axioms. Suppose there are three groups in the social partition: clergypersons, lay men, and lay women, and the issue is abortion rights. One may be indifferent between $A = (100, 800, 0)$ and $B = (100, 0, 400)$, but not between $A' = (500, 800, 0)$ and $B' = (500, 0, 400)$ (a violation of complete separability) or between $A'' = (100, 160, 0)$ and $B'' = (100, 0, 80)$ (a violation of (I, a^*)–scaling). We reject this intuition for the following reasons. It is implicitly assumed in these examples that there is a reason to check the power of clergypersons, a goal which is not achieved in B' and B''. But first, it should be emphasized that we do not know how individual members of different groups are going to vote and hence do not have a reason to limit their power in the light of their particular views. Secondly, the preferences $\succeq$ are not over committee-like representations, but over the position of members of society and their role in the process of social decision making. To that extent, the presence of more or less individuals of one category in a given set

should not affect our appreciation of the weight of other categories of people who form the set.

The monotonicity assumption is obvious, and is slightly stronger than what is needed for the proof of Theorem 1 (which only requires $S \nsim \varnothing$). Strict monotonicity may be challenged similarly to the complete separability assumption on the grounds that if A has too little representation of category S_i, increasing representation of type $j \neq i$ may reduce the desirability of this group. Our previous arguments are relevant here as well.

Theorem 1 *Assume $N \geqslant 3$. The following two conditions are equivalent.*

1. *The preferences $\succeq$ over the subsets of S satisfy the assumptions of scaling, residual scaling, complete separability, continuity, and monotonicity.*

2. *There are numbers $k_1, \ldots, k_N$, unique up to multiplication by (the same) positive constant β, such that the preference relation $\succeq$ can be represented by $V(A) = \sum k_i \mu(A_i)$. If monotonicity is replaced with strict monotonicity, then the numbers $k_1, \ldots, k_N$ are all non-negative.*

Given the uniqueness up to multiplication by the same constant β, the weights $k_1, \ldots, k_N$ can be normalized. We will adopt the normalization

$$\sum k_n \mu(S_n) = 1 \tag{1}$$

Suppose that on the issue of abortion rights society recognizes three groups: clergy (10% of the population), lay men (45%), and lay women (45%), and that we find out that person α assigns these groups the weights $(1, \frac{2}{3}, \frac{4}{3})$, respectively. Given a set A, he is indifferent between enlarging it by adding one lay woman or by two lay men. In other words, in his view, and with respect to this issue, lay women should count twice as much as lay men.

But of course, society does not face a choice between subsets S. Moreover, even if it did, the strict monotonicity assumption implies that everyone considers S to be the best set to make social decisions. Given this constraint, person α can still express his view that "one lay woman should count twice as one lay man" by giving women twice the voting power of men. In other

words, we can imagine person α assigning $f(x)$ coupons to member x of society, where the constraint is

$$\int_0^1 f(x)\,dx = 1$$

His views on the different groups can be expressed by giving each clergyperson one coupon, each lay woman $\frac{4}{3}$ coupons, and each lay man $\frac{2}{3}$ coupons.[7]

In the formal presentation we imagined each voter as ranking all possible subsets of society for making the decision for the whole of society. Despite the appearance of contradiction, there is no inconsistency between the informal presentation of the two-tier voting model and the formal one. For, the complete ranking of all subsets is merely a theoretical tool for expressing the relative weights each individual wishes to ascribe to categories of people in society as a whole. It does not imply an actual wish by the individual that a subset of society make the choice for all the rest anymore than a consent on a Rawlsian Original Position implies a blueprint for a political body in which actual social choices would be made. Thus, the merit of representing the actual assignment of differential weights to all members of society in terms of the ranking of subgroups that are allegedly given the power to make choices for society as a whole lies in its ability to circumvent the problem of the cardinalization of preferences. It should not be understood as a disagreement between individuals about who in society should decide policies for all the rest, since it is universally agreed that all individual members should take part in the decision making process. By democratically elected aristocracies we do not mean an exclusive club or caste, but a range of relatively growing weight of voting power given to categories of all individuals in society.

[7]Note that coupons here represent voting power rather than the means of acquiring resources as is the case in Dworkin's [3, pp. 65–71] famous desert-island auction. Dworkin explicitly says that the "clamshells," distributed equally between the islanders, can be used only to purchase privately owned resources and that the issue of the equality of political power should be "treated as a different issue." But beyond the obvious difference between the distribution of power (or specifically voting power) and that of personal goods (which, for Dworkin raises the fundamental problem of envy), there is a structural similarity in that both kinds of coupons must be allocated equally (i.e. the number of clamshells must be the same or, in our case, the sum of assigned weights must be 1).

4 Aggregation

The analysis of the previous section yields the conclusion that each member α of society would like to assign the voting weights $k(\alpha) = (k_1(\alpha), \ldots, k_N(\alpha))$, $\sum k_n(\alpha)\mu(S_n) = 1$ to society's N subgroups. Given these individual preferences, society too, we suggest, should assign such weights, and these should be based on the individual weights. This section discusses such an aggregation. Our aim is to obtain a rule that applies to all possible profiles of individual preferences (subject to some structural constraints), and not just to one given profile.

Denote by $\mathcal{K}$ the set $\{k \in \Re_+^N : \sum_{i=1}^N k_i\mu(S_i) = 1\}$. For every $\alpha \in [0,1]$, person α's preferences lead to an element of $\mathcal{K}$. Denote this function $f = (f_1, \ldots, f_N) : [0,1] \to \mathcal{K}$. We restrict attention to measurable functions f and denote the set of all such functions $\mathcal{F}$. The two functions $f, g \in \mathcal{F}$ induce the same marginal distributions on $\mathcal{K}$ if for all measurable $T \subset \mathcal{K}$ and for all i, $\mu(f_i^{-1}(T)) = \mu(g_i^{-1}(T))$. Our aim is to find a function $\varphi = (\varphi_1, \ldots, \varphi_N) : \mathcal{F} \to \mathcal{K}$, that is, to aggregate the individual weights, as expressed by $f \in \mathcal{F}$, into a vector of social weights. We want to do this not just for one set of individual weights $k(\alpha)$, but for all such sets of weights. We offer the following assumptions.

Distribution Independence If f and g induce the same marginal distributions over $\mathcal{K}$, then $\varphi(f) = \varphi(g)$.

Unanimity Suppose that for some λ and i, and for all $\alpha \in [0,1]$, $g_i(\alpha) = \lambda f_i(\alpha)$. Then $\varphi_i(g) = \lambda\varphi_i(f)$.

The first axiom assumes that the aggregation procedure is indifferent to the proper names of the members of society. By itself it does not imply that the aggregate weight of one group is independent of the aggregate weights of other groups. The second assumption is of course stronger than plain unanimity, in which if everyone agrees on the weight of a certain group, society too should apply this value. Here we apply unanimity to (relative) changes, rather than to the particular views themselves. One may argue that other forms of unanimity are possible, for example, if for some b and i, and for all $\alpha \in [0,1]$, $g_i(\alpha) = f_i(\alpha) + b$, then $\varphi_i(g) = \varphi_i(f) + b$. As we show in Theorem 2 below, this form of unanimity follows from the above two assumptions. We suggest the proportional form of unanimity as it seems

to fit in its nature with the general setup of the present model, where the ratio between the weights of different groups plays an important role (see the discussion following Theorem 1 in the preceding section).

The unanimity assumption is stronger than it may seem. Notice that it is made with no regard to what happens to the weights individuals wish to assign to other groups. But when the individual weights of one group are all multiplied by λ, other weights too must change. As we show in the proof of Theorem 2, this assumption implies in particular that the social aggregation of type i (say "female doctors") depends on the way members of society evaluate this group but not on the way they evaluate *other* groups (e.g., "male lawyers").

Theorem 2 *Assume $N \geqslant 3$. The social coefficients k_i satisfy for all i,*

$$k_i = \int_0^1 k_i(\alpha)\, d\alpha$$

In other words, the social coefficients are the average of the individual coefficients.

This Theorem may seem patently wrong, as clearly homotheticity does not imply linearity. But as stated above, the Theorem utilizes a technical constraint that does not appear explicitly as an assumption, namely, that the sum of social and individual weights must satisfy eq. (1). Therefore we show that φ_i is homothetic with respect to a large set of points, hence linear.

5 Multiple Issues

So far, our analysis assumed just one issue, say "abortion rights." But suppose society has to decide simultaneously on several issues, for example abortion rights and the scope of sexual harassment. If members of society see no connection between these issues, and judge each in isolation, the analysis of the last section still holds. But what happens if the weights people are willing to give to some subgroups of society depend on the weights these groups receive on other issues?

Consider the above example. Even if we don't know how any given man or woman is going to vote on the issues of abortion rights and the scope of sexual harassment, some may feel that women should be given more voice on both

issues. Suppose person α believes that in both cases women's weights should be twice as that of men. If society disagrees, and gives women no special vote on one issue, it is conceivable that α will be willing to compensate women by offering them more weight on the other issue. We do not suggest that person α is trying to manipulate society by misrepresenting his true assessment of the weights men and women should receive, but that the weights he is willing to assign them may depend on the empathy he feels towards women, and knowing that they got too little weight on one issue increases his sensitivity to their needs and views on other issue. But then, will society be able to find weights, one system for each issue, such that individual and social weights are consistent with each other?

Suppose society has M issues to consider. To simplify notation, we assume that the same partition $S_1, \ldots, S_N$ of agents applies to all M issues. Extending the analysis of the previous section, we now assume that each member of society has preferences over decisive subsets of S for issue m, $m = 1, \ldots, M$. These preferences satisfy the assumptions of Section 3, but they may now depend on the weights each of the N categories receive on other issues. Thus, for issue m, person α has the preferences $\succeq_\alpha^m(k_{-m})$, where k_{-m} are the social weights to all groups in all other issues. Since, by Theorem 1, these preferences are representable by the linear weights $k(m, k_{-m}, \alpha)$, we express the following continuity assumption in terms of these weights, but the translation into continuity of the preferences themselves in k_{-m} (via measurable subsets of $2^S \times 2^S$) is simple.

Continuity The preferences $\succeq_\alpha^m(k_{-m})$ person α has over decisive sets for issue m are uniformly continuous in k_{-m} and in α. That is, $\forall\delta\ \exists\varepsilon$ such that $\| k'_{-m} - k_{-m} \| < \varepsilon$ implies, for all α, $\| k(m, k'_{-m}, \alpha) - k(m, k_{-m}, \alpha) \| < \delta$.

It follows that the social weights for issue m, being the average of the individual weights, are a continuous function of the social weights for all other issues. Formally, assume strict monotonicity (hence all weights are non-negative), and consider the sets $\mathcal{K}^m = \{k^m \in \Re_+^N : \sum_{i=1}^N k_i^m \mu(S_i) = 1\}$, $m = 1, \ldots, M$. There are M continuous functions, $g^m : \prod_{i \neq m} \mathcal{K}^i \to \mathcal{K}^m$, such that given the social weights $k^i = (k_1^i, \ldots, k_N^i)$ for issue i, $i = 1, \ldots, M$, $i \neq m$, the average values of the individual weights for issue m equal $g^m(k^1, \ldots, k^{m-1}, k^{m+1}, \ldots, k^M) \in \mathcal{K}^m$. Define now $g : \prod_m \mathcal{K}^m \to \prod_m \mathcal{K}^m$ by

$$g(k^1, \ldots, k^M) = (\ldots, g^m(k^1, \ldots, k^{m-1}, k^{m+1}, \ldots, k^M), \ldots)$$

By Brouwer's fixed point theorem, this function has a fixed point, that is, a system of weights $\bar{k}^1, \ldots, \bar{k}^M$, such that for all m,

$$\bar{k}^m = g^m(\bar{k}^1, \ldots, \bar{k}^{m-1}, \bar{k}^{m+1}, \ldots, \bar{k}^M)$$

The meaning of this last result is simple. For each m, given the social weights $\bar{k}^1, \ldots, \bar{k}^{m-1}, \bar{k}^{m+1}, \ldots, \bar{k}^M$, each person in society forms his weights for issue m. The social weights for this issue are the average of the personal weights, and they are equal to $\bar{k}^m$.

6 Q & A

ARE THE ONE-PHASE AND TWO-PHASE SYSTEMS REALLY DIFFERENT? The fundamental idea behind the model is that it acknowledges the limitation of the traditional assumption about the self-interested behavior of voters and the need to give expression to the way voters consider the standing of others in the matter under dispute. But cannot this other-regarding aspect be incorporated in a one-phase vote? Formally, consider the following procedure. Each member of society first determines the weights he wishes to assign to each of the subgroups $S_1, \ldots, S_k$, as suggested by Theorem 1. He then computes the outcome of the prospective actual vote according to these weights and proceeds to cast his personal vote on the substantive issue according to that outcome. Will this simpler mechanism yield different results from those of Theorem 2?

The answer is yes, for two reasons. Firstly, as mentioned above, there are many cases in which the interests of a particular subgroup are not at all identifiable although one may think that the subgroup is in a special position to decide the issue. For example, one might believe that women have a particular standing with regards to abortion policies, although one does not know how women will in fact vote on them (since they may be no less controverted in the female subgroup than in society at large). Secondly, even when the interests of the subgroups are known, the one- and two-phase procedures may yield different outcomes. Consider the following example.

Suppose $k = 2$, $\mu(S_1) = 0.2$ and $\mu(S_2) = 0.8$, and suppose that all members of S_1 vote the same (say, "Yes") on a certain issue, while all members of S_2 vote in the opposite way. All members of S_1 and $\frac{1}{4}$ of the members of S_2 (that is, 40% of the whole population) believe that the appropriate weight

of members of S_1 is 5 while the weight of members of S_2 should be 0. The remaining $\frac{3}{4}$ of S_2 (that is, 60% of the population) believe the weights should be 2.4 and 0.65, respectively. (Observe that $0.2 \cdot 5 = 0.2 \cdot 2.4 + 0.8 \cdot 0.65 = 1$). Following Theorem 2, the social weights should be $0.4 \cdot 5 + 0.6 \cdot 2.4 = 3.44$ to members of S_1 and $0.4 \cdot 0 + 0.6 \cdot 0.65 = 0.39$ to members of S_2. (Here too $0.2 \cdot 3.44 + 0.8 \cdot 0.39 = 1$). According to the procedure suggested in this paper, each vote of members of S_1 is multiplied by 3.44, while each vote of members of S_2 is multiplied by 0.39. Since $0.2 \cdot 3.44 > 0.8 \cdot 0.39$, "Yes" wins over "No."

Consider now the alternative, one-phase vote. 40% of the population believe that members of S_1 should receive weight 5, and if society is to vote according to these weights, "Yes" wins. Hence these 40% vote "Yes" in the one-phase vote. The remaining 60% believe the weights should be 2.4 and 0.65, and if society votes according to these weights, "No" wins. (Observe that $0.2 \cdot 2.4 < 0.8 \cdot 0.65$). Therefore, 60% vote "No," and "No" wins over "Yes."

WHAT IS TO BE REPRESENTED: PREFERENCES OR VALUES? In Harsanyi's [8] model of social choice individuals have selfish preferences over social policies and these are then aggregated into social preferences. The more recent literature conceives of each member of society as having two sets of preferences, selfish and social (see Estlund [4], Wolff [15], Segal [14], Karni and Safra [12], and Karni [11]). We agree with the assumption of these recent models that social concerns should be taken as part of the individual's characteristics and in particular that these social concerns may differ from one person to another. Social concerns in our model are represented by the weights each individual is willing to assign to other members of society. Preferences enter our analysis in the second phase of the voting, when social questions are actually decided.

The crux of the theoretical motivation behind the suggested model is the following: unlike the standard attempt to devise a voting scheme that would best represent the preferences of individuals in a social context, our starting point is that what is to be represented is not only what people prefer (weighted and aggregated), but also how people regard the relative weight of all members in counting and weighing their preferences. It is an attempt to represent the *normative* value of individual preferences as it is determined by everybody, rather than merely reflect the *positive* values of

the preferences themselves. To that extent, our model is a combination of positive and normative factors, where normative values determine the weights voters receive, and the final vote reflects actual individual preferences.

SHOULD NEGATIVE WEIGHTS BE PERMITTED? Theorem 1 permits negative weights, but the strict monotonicity assumption rules out this possibility. In our context this is a natural assumption as well as politically justifiable. The fundamental motivation for assigning differential voting power is associated with the principle of empathy to others and the attempt to reach some sort of social consensus despite substantive disagreements. Assigning negative weight to another's opinion or preference runs against this democratic spirit of solidarity. For although one could in principle agree that he himself should get zero weight in a particular vote (for instance, admitting that he knows nothing about the subject or is indifferent to the conflicting interests), no one would probably agree to being given a negative standing, since that would mean that one is systematically wrong, irrational, or malicious in his preferences and hence should be discounted rather than merely not counted. We sometimes think that the fact that a certain person makes a particular choice or holds a certain belief is in itself a reason to make the opposite judgment or choice (e.g. in deciding whether a certain movie is worth seeing, we might act contrary to the recommendation of a friend whom we know to have bad taste). However, these cases of "counter-authority," in contradistinction to "lack of authority," are not typical of the political contexts of social choice with which we are concerned.[8]

The exclusion of negative weights also carries the extra bonus of escaping the most conspicuous temptation to vote strategically, although, admittedly, does not remove that threat completely. If I know that I am assigned a negative weight by many voters, I have a strong motivation to cast my vote for the opposite option to the one I believe in. We have on the whole avoided the problem of strategic voting, both since we wanted to theoretically constrain ourselves to a relatively ideal model of representation and because by

[8]Even in extreme cases, in which society deems a particular opinion or ideology as lying "beyond the democratic pale," it sometimes prohibits parties representing this opinion from running in elections, thus giving them zero weight. Neo-Nazis are not given negative weight on matters of immigration to Germany; they are simply prohibited from expressing their views in an institutionalized manner. Even the argument that new, or non-francophone immigrants should not take part in a referendum on Quebec's secession does not suggest giving these voters negative weights, just zero.

prohibiting negative weights the motivation to vote strategically decreases as a matter of empirical fact.

7 Some Remarks on the Literature

The literature on voting and social choice consists of many attempts to revise the "positive" preference-based, self-centered approach by introducing into it a normative as well as a social (other-regarding) dimension. It might therefore be illuminating to show the way in which the model outlined here differs from and goes beyond these attempts. John Stuart Mill [13, pp. 137–143, 180] suggested granting extra votes to the more educated classes in society. Mill's idea, shared by some contemporary followers (see Harwood [10]), is that a system of "plural voting" would promote the public education and through that the quality of both the public debate and the outcome of the political decision- making process. Mill even believed that it would lead to the advancement of moral excellence. However, a system of plural voting, like most other suggestions for the improvement of electoral systems, concerns objective and independently fixed conditions of elections, whereas our proposal is to have these very conditions put to a vote. Mill was seeking "a trustworthy system of general examination," while we are looking for the subjective assessment of all the voters regarding the source of differential authority on a particular measure. We thus circumvent all the objections regarding both the irrelevance of education for intelligent political choices and the problems in deciding the appropriate levels of education. We also avoid Mill's painful oscillation between his basic egalitarian commitment and his elitist faith in the authority of the educated classes.

Political philosophers have expressed reservations about the preference-based principle of voting. Estlund [4], for example, argues that the common notion of democracy is incompatible with the idea of an epistemically ideal observer who decides social policies on the basis of individuals' preferences. Democracy is not just "for the people" but also "by the people," in the sense that it requires an act of choice, typically voting. Our model is in agreement with Estlund's "activity condition," since it not only rules out an "ideal preference reader" in the second-phase vote, but also denies an imposition of an external criterion for differential voting, insisting rather on active voting also in the first phase. Estlund demonstrates that individual active expres-

sions of preferences cannot be aggregated (due to their inextricable indexical character) and concludes that the object of voting must be the common interest rather than individual preferences. Our model is not committed to any particular view about the content of the vote, but suggests that members of society introduce their notion of the common good in the differential allocation of voting power based on their views about the common good.

Our proposal can also partly respond to Wolff's [15] "mixed motivation problem," according to which some people vote on the basis of their narrow personal interests while others vote in the light of their beliefs about the common good, the consequence being that we don't know how to interpret the outcome of the vote. Splitting the vote into two stages can provide voters with a reasonable combination of what they believe is good or fair from a social (group) point of view and what they personally prefer the policy in question to be.

It is also worth mentioning how our suggested voting scheme differs from the idea of agreement under an ideal veil of ignorance (of the Harsanyian or Rawlsian type). The suggested scheme is not primarily motivated by the idea of fairness that calls for background conditions of anonymity in the exercise of self-interested voters, but rather by the ideal of adequately representing the way real people actually evaluate others' interests. It is not the procedural fairness of the method that lends the outcome its validity as just, but the sensitivity to individual substantive evaluations of the differential weights democratically assigned to identifiable groups of people in society.

Appendix

Proof of Theorem 1 It is easy to verify that (2) implies (1). We prove that (1) implies (2) through a sequence of lemmas.

The preferences $\succeq$ induce preferences over $X = \prod_{i=1}^{N}[0, \mu(S_i)]$, so as before, we will use wlg the same notation $\succeq$. Let $p = (\mu(S_1), \ldots, \mu(S_N))$, and let $L = [0, p]$ (here and throughout the proofs, $[a, b]$ denotes the chord connecting the points a and b in $\Re^N$).

Lemma 1 *The preferences $\succeq$ are strictly monotonic along L.*

Proof Suppose that for some $a, b \in L$, $a = \lambda b$ for some $\lambda < 1$, but $a \sim b$. By the scaling assumption, $b \sim \lambda b \sim \lambda^2 b \sim \cdots \sim \lambda^n b \sim \cdots$, hence

$b \sim 0$. Similarly, by the residual scaling assumption, $b \sim p$, hence $p \sim 0$, a contradiction to the monotonicity assumption. □

To justify Fig. 1, which is used in the proof of Lemma 3, we need the following result, which is proved after the proof of Theorem 1.

Lemma 2 *Let H be a 2-dimensional plane containing L. Then $H \cap X$ is a parallelogram in $\Re^N$.*

Lemma 3 *Let H be a 2-dimensional plane containing L. On $H \cap X$, the preferences $\succeq$ can be represented by (possibly different) linear functions on each side of L.*

Proof Since preferences are strictly monotonic along L, it follows by continuity that there are $a \in L$ and $b \notin L$ such that $a \sim b$. and let $c \in [a, b]$. The points 0, p, a, b, and c are in the same 2-dimensional plane, denote it H. Following Lemma 2, $H \cap X$ is depicted in Fig. 1 by the parallelogram $0gph$. Denote by d the intersection of the line through 0 and b with the line through p and c (see Fig. 1). Let $ed \parallel ba$. By the scaling assumption, $d \sim e$. Since $ca \parallel de$, it follows by the residual scaling assumption that $c \sim a$. In other words, the chord $[a, b]$ is an indifference set of $\succeq$.

We want to show next that the continuation of the chord $[a, b]$ in the direction of b is also part of the indifference set through a. Suppose not, and suppose, wlg, that there is a sequence $b_n \to b$ such that for all n, $b \in [a, b_n]$, and $b_n \nsim a$, say $b_n \succ a$. By continuity, there is a sufficiently high n such that there exists a point $a_n \in L$ for which $b_n \sim a_n \succ a \sim b$. By the above arguments, the chord $[a_n, b_n]$ is an indifference set of $\succeq$. Denote by c_n the intersection of this chord with the chord $[0, f]$, where f is the point on the boundary of $H \cap X$ for which $b \in [0, f]$ (see Fig. 1). Clearly, $a_n c_n \nparallel ab$. By the scaling assumption it follows that there is a point $d_n \in L$, strictly between a and a_n, such that $b \sim d_n$, a contradiction to Lemma 1. The scaling and the residual scaling assumptions therefore imply that on $\triangle 0gp$, the preferences can be represented by a linear function.

Note that the above analysis applies equally to the case where b is in the triangle $\triangle 0hp$ in Fig. 1. We therefore conclude that for all $b \in X \backslash L$, the preferences over the intersection of the half plane containing L and b with X can be represented by a linear function. □

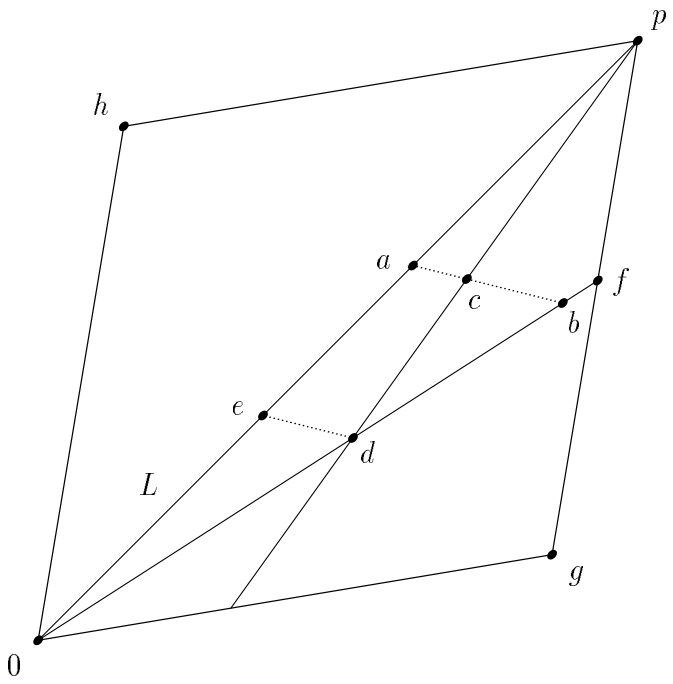

Figure 1: Proof of Lemma 3

Lemma 4 *Scaling and complete separability imply (I, a^*)–scaling, while residual scaling and complete separability imply (I, a^*)–residual scaling.*

Proof Let $a, b \in X(I, a^*)$. Then $\lambda a, \lambda b \in X(I, \lambda a^*)$. Suppose $I = \{i_0 + 1, \ldots, N\}$. For $i \in I$, define $a^i = \lambda a$ if $i = i_0 + 1$, and for $i = i_0 + 2, \ldots, N$, define $a^i = a^{i-1}(\{i\}, a^*)$. In other words, a^i is obtained from a^{i-1} by replacing the i-th coordinate of a^{i-1} with a_i^*. By scaling, $a \succeq b$ iff $a^{i_0+1} \succeq b^{i_0+1}$, and by complete separability, for $i = i_0 + 2, \ldots, N$, $a^i \succeq b^i$ iff $a^{i-1} \succeq b^{i-1}$. But $a^N = \lambda a(I, a^*) + (1 - \lambda)0(I, a^*)$, hence (I, a^*)–scaling.

The proof of (I, a^*)–residual scaling is similar. □

Similarly to the above analysis, it follows that for every (I, a^*) and for every $b \in X(I, a^*)$, on the 2-dimensional plane H through b and $L(I, a^*)$ (that is, the plane that is determined by the three points b, $0(I, a^*)$, and $p(I, a^*)$), the preferences $\succeq$ can be represented by a function that is linear on each of the two sides of $L(I, a^*)$ in H.

The preferences $\succeq$ over the product set X are continuous and completely separable, and can therefore be represented by an additively separable function of the form

$$V(a) = \sum v_i(a_i) \tag{2}$$

(see Debreu [2] and Gorman [6]). Consider now the set $X(\{3,\ldots,N\}, a^*)$, where all but the first two variables are fixed. On this set, the preferences can be represented by $v_1(a_1) + v_2(a_2)$, but also by

$$W(a_1,a_2) = \begin{cases} k_1 a_1 + k_2 a_2 & a_2 < \frac{p_2}{p_1} a_1 \\ \\ k_1' a_1 + k_2' a_2 & a_2 \geqslant \frac{p_2}{p_1} a_1 \end{cases} \tag{3}$$

where $k_1 p_1 + k_2 p_2 = k_1' p_1 + k_2' p_2$.

Lemma 5 *v_1 is linear.*

Proof Consider the range $a_2 < \frac{p_2}{p_1} a_1$. From eqs. (2) and (3) it follows that there is a monotonic function h such that

$$v_1(a_1) + v_2(a_2) = h(k_1 a_1 + k_2 a_2)$$

The function h is monotonic, hence almost everywhere differentiable. Pick a point (a_1^0, a_2^0) such that h is differentiable at $k_1 a_1^0 + k_2 a_2^0$. It follows that v_1 must be differentiable at a_1^0, hence

$$v_1'(a_1^0) = k_1 h'(k_1 a_1^0 + k_2 a_2^0) \tag{4}$$

By continuity, there is a segment of values of a_1 for which there are values of a_2 such that $a_2 < \frac{p_2}{p_1} a_1$ and $k_1 a_1 + k_2 a_2 = k_1 a_1^0 + k_2 a_2^0$. (If not, then $k_1 = 0$ and the lemma is trivially true). At all these points, the value of h' is the same, and therefore, on this segment v_1' is constant and v_1 is linear.

Since h is almost everywhere differentiable, we can get such overlapping segments of values of v_1, hence v_1 is globally linear. The same proof holds for v_2. □

By similar arguments all the functions v_i are linear, hence the theorem. The proof of the claim that strict monotonicity implies positive coefficients is trivial. ■

Proof of Lemma 2 To simplify notation, we assume that $X = [0,1]^N$, that is, $p = (1,\ldots,1)$. An edge of X is identified by a pair (I, i^*) where $I \subsetneqq \{1,\ldots,N\}$ and $i^* \notin I$, and is given by $\{a \in X : a_i = 0 \text{ for } i \in I,\ a_i = 1 \text{ for } i^* \neq i \notin I \text{ and } a_{i^*} \in [0,1]\}$. Pick a 2-dimensional plane H such that $L \subset H$, let $a^* \neq 0, p$ be on the edge (I, i^*) of X. H can be represented as $\{\gamma p + \delta a^*\}$. Let $b \in X \cap H$ be another point on the edge (I', i') of X. There are γ and δ such that $b = \gamma p + \delta a^*$. We now discuss all possible connections between (I, i^*) and (I', i').

1. $\exists i$ such that $a^*_i = b_i = 0$: $\gamma = 0$, hence $b = \delta a^*$. The point b can be on the edge of X iff $I' = \{1,\ldots,N\}\backslash\{i'\}$ and $i' = i^*$. In other words, a^* and b are on the same edge of X.

2. $\exists i$ such that $a^*_i = 1$ and $b_i = 0$: $\gamma + \delta = 0$. If there is $j \neq i'$ such that $a^*_j = 0$, then $b_j = \gamma$, hence $\gamma = 1$ and $\delta = -1$. Clearly, 0, p, a^*, and $p - a^*$ form a parallelogram. Alternatively, for all $j \neq i'$, $a^*_j = 1$. Once again, a^* and b are on the same edge of X.

3. $\exists i$ such that $a^*_i = b_i = 1$: $\gamma + \delta = 1$. If there is $j \neq i'$ such that $a^*_j = 0$, then $b_j = \gamma$, hence $\gamma = 1$, $\delta = 0$, and $b = p$. Otherwise, for all $j \neq i'$, $a^*_j = b_j = 1$, and again, a^* and b are on the same edge of X.

4. $\exists i$ such that $a^*_i = 0$ and $b_i = 1$: $\gamma = 1$, hence $b = p + \delta a^*$. If there is $j \neq i'$ such that $b_j = 0$, then $a^*_j = 1$ and $\delta = -1$. As before, 0, p, a^*, and $p - a^*$ form a parallelogram. If for all $j \neq i'$, $b_j = 1$, then either $\exists j \neq i'$ such that $a^*_j = 1$, hence $\delta = 0$ and $b = p$, or for all $j \neq i'$, $a^*_j = 0$. Here too, 0, p, a^*, and $p - a^*$ form a parallelogram.

We now look into the case where a^* and b are on the same edge. It is easy to verify that this edge must also contain either 0 or p, and therefore $H \cap X$ is a parallelogram (in fact, a rectangle). □

Proof of Theorem 2 Unanimity applies to all λ, and in particular to $\lambda = 1$. It thus follows that $\varphi(f) = (\varphi_1(f),\ldots,\varphi_N(f)) = (\varphi_1(f_1),\ldots,\varphi_N(f_N))$.

To simplify notation, we assume wlg that $\mu(S_1) = \cdots = \mu(S_N) = \frac{1}{N}$.[9] Note that $\sum f_i = \sum \varphi_i = N$.

[9] Alternatively, we can define $\mathfrak{f}_i = f_i \mu(S_i)$, and work with these functions instead of the functions f_i.

Lemma 6 *Let $f_i(x) \equiv \lambda$. Then $\varphi_i(f_i) = \lambda$.*

Proof By unanimity, $\varphi_i(0 \cdot f_i) = 0$. By the definition of φ, $\sum_i \varphi_i(f) = N$. Therefore, if $f_i \equiv N$ and for all $j \neq i$, $f_j \equiv 0$, then $\varphi_i(f_i) = N$. Unanimity now implies the lemma. □

Lemma 7 *Let H be a 2-dimensional plane of functions f_i that is determined by the functions $f_i^0 \equiv 0$, f_i^1 and f_i^2. On this domain, the function φ_i is linear.*

Proof Assume, for simplicity, $i = 1$. Let f_1^1 and f_1^2 be as in the lemma, and consider the set H of functions spanned by these two functions and by f_1^0, such that for all $f_1 \in H$ and for all x, $f_1(x) \leqslant N$. Pick g_1^1 and g_1^2 in the interior of H such that f_1^0, g_1^1, and g_1^2 are not on the same line. There is $\delta > 0$ such that $\min g_1^j(x), \min\{N - g_1^j(x)\} > \delta$, $j = 1, 2$. Obviously, for all x, $g_1^j(x)N/(N - \delta) < N$. Define

$$g_3^j(x) = N - \frac{g_1^j(x)N}{N - \delta}$$

and let $g_i^j \equiv 0$, $i = 4, \ldots, N$, $j = 1, 2$. If g_1^1 and g_1^2 are sufficiently close to each other, (the exact requirement is that for every x, $g_1^2(x)(1 - \frac{\delta}{N}) < g_1^1(x) < g_1^2(x)/(1 - \frac{\delta}{N})$), then for all x

$$N - g_3^1(x) = \frac{g_1^1(x)N}{N - \delta} > g_1^2(x)$$

and likewise, $N - g_3^2(x) > g_1^1(x)$.

Fig. 2 depicts the weights given by two individuals in society to individuals of type 1. The weights members of this group may receive cannot exceed N, hence the views on the weights of group 1, when only two individuals can express their opinions about these weights, must be in the square $[0, N]^2$. If everyone agrees that all types $4, \ldots, N$ receive the weight 0, then the box H^j determined by 0 and $N - g_3^j$ depicts possible allocations of weight members of society may wish to allocate to the first two types, where the weights of type 1 are measured from 0, and the weights of type 2 are measured from $N - g_3^j$ (in the direction of 0).

By unanimity, the function $\varphi_1(g_1)$ satisfies on the domain H

$$\varphi(\lambda g_1) = \lambda \varphi(g_1) \tag{5}$$

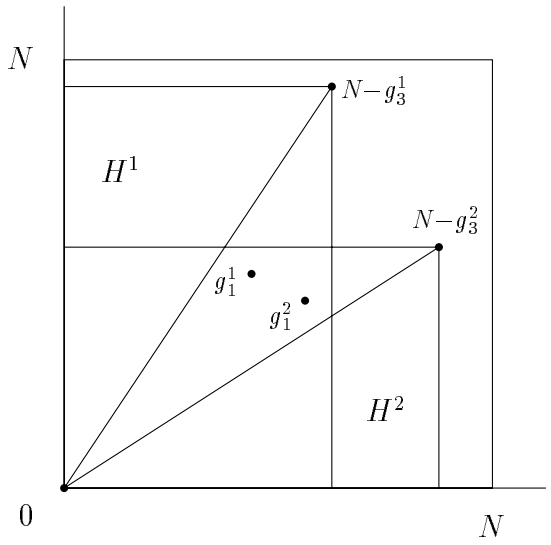

Figure 2: Proof of Lemma 7

which is similar to the scaling assumption of Section 3. Also, given $g_3^j, \ldots, g_N^j$, define

$$g_2^j(x) = N - g_1^j(x) - \sum_{i=3}^{N} g_i^j(x)$$

Applying unanimity to $\varphi_2(g_2)$, we obtain for $j = 1, 2$

$$\varphi_1\left(N - \sum_{i=3}^{N} g_i^j - \lambda g_2^j\right) = N - \varphi_3(g_3^j) - \lambda\varphi(g_2^j) \tag{6}$$

which is similar to the residual scaling assumption of Section 3. By Lemma 3, φ_1 is linear on H^j on both sides of the chords $[0, \frac{g_1^j N}{N-\delta}]$, $j = 1, 2$. Since it is homothetic on H, it is linear there. □

Lemma 7 implies that φ_i satisfies betweenness: $\varphi_i(f_i^1) = \varphi_i(f_i^2)$ implies for all $\zeta \in [0, 1]$, $\varphi_i(f_i^1) = \varphi_i(\zeta f_i^1 + (1-\zeta) f_i^2)$. Indifference sets of φ_i are planar, and parallel on any two dimensional plane, hence φ_i can be represented by a linear function. By unanimity, φ_i is linear, and by the distribution indifference assumption, it is the average of f_i. ■.

References

[1] Condorcet, Marquis de. 1976 (1785). "Essay on the application of mathematics to the theory of decision-making." *Selected Writings* (ed. Keith M. Baker), Indianapolis: Bobbs-Merrill, pp. 48–49.

[2] Debreu, G. 1960. "Topological methods in cardinal utility theory." *Mathematical Methods in the Social Sciences* (eds. K.J. Arrow, S. Karlin, and P. Suppes), Stanford: Stanford University Press.

[3] Dworkin, R.M. 2000. *Sovereign Virtue*. Cambridge: Harvard University Press.

[4] Estlund, D.M. 1990. "Democracy without preference." *Philosophical Review* 99:397–423.

[5] Frankfurt, H. 1988. "Freedom of the will and the concept of a person." *The Importance of What We Care About* (ed. H. Frankfurt), Cambridge: Cambridge University Press, pp. 11–25.

[6] Gorman, W.M. 1995. *Separability and Aggregation: Vol. 1.* Oxford, Clarendon Press.

[7] Harsanyi, J.C. 1953. "Cardinal utility in welfare economics and in the theory of risk-taking." *Journal of Political Economy* 61:434–435.

[8] Harsanyi, J.C. 1955. "Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility." *Journal of Political Economy* 63:309–321.

[9] Harsanyi, J.C. 1977. *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge: Cambridge University Press.

[10] Harwood, R. 1998. "More votes for Ph.D.'s." *Journal of Social Philosophy* 29:129–141.

[11] Karni, E. 2001. "Impartiality and interpersonal comparisons of variations in well-being." Mimeo.

[12] Karni, E. and Safra Z. 2002. "Individual sense of justice: A utility representation." *Econometrica* 70:263–284.

[13] Mill, J. S. 1958 (1861). *Considerations on Representative Government.* New York: The Liberal Arts Press.

[14] Segal, U. 2000. "Let's agree that all dictatorships are equally bad." *Journal of Political Economy* 108:569–589.

[15] Wolff, J. 1994. "Democratic voting and the mixed-motivation problem," *Analysis* 54:193–195.