Democratically Elected Aristocracies

David Heyd* and Uzi Segal†

December 23, 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.

*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

Voting is one among many procedures of making social choices. Like some alternative procedures, such as the market, voting faces the problem of the translation of individual preferences into social policies. There are numerous issues involved in voting, e.g. articulating the issue to be decided, counting the votes and interpreting the election results. But behind all the attempts to design voting procedures lies the overall aim of representing individual preferences as loyally as possible. In this article we propose a new way in which voting can be transformed into a more sophisticated procedure, which would make the resulting social choice more sensitive to and hence more representative of the individuals' preferences. Unlike other methods of social choice, such as the appeal to authoritarian decision by a dictator, to a committee of specialists, or to the use of lottery, voting takes seriously the preferences of individuals as they see it, and hence is committed to some form of equality. Our proposal follows this fundamental egalitarian principle but gives it a more complex interpretation.

Voting is a procedure that is applied to issues that call for collective action. 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 2). 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 3 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. Society then decides 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.

In section 4 we analyze the case where the weights one person wishes to

[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.

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

1 The Two-Tier Voting Model

To be able to obtain a social welfare function, Harsanyi [7] extended 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 axiom 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 proved that social preferences can be represented by a weighted sum of individual vN&M utilities.

An alternative way in which individual preferences can be aggregated is by using quasi linear utilities. 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, not b, not c) with probability p and (not a, not b, c) with probability $1 - p$" and "(not a, b, not 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 [17], Karni [9], Karni and Safra [10], and Segal [14]).

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

[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.

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 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. Our model applies best in referendum-like contexts in which a controversial, ideologically charged issue is to be decided in a yes-or-no manner in a fairly large group of people.[4]

[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 be identified as such a type.

[4]By characterizing the suitable issues as "ideologically charged" we mean to exclude cases which are hard to decide due to epistemological problems of arriving at the truth of the matter. In such contexts, we might want a certain group of experts to decide for society (or have extra weight in the collective decision), but we don't necessarily want the selection of the experts to be itself decided by vote.

Typical issues to which our model applies are those involving two possible outcomes ("yes–no vote"). As is usually the case, we too assume that society is given, and therefore our procedure cannot be applied to the question "who is a member of society" (see Kasher and Rubinstein [11] and Samet and Schmeidler [13]). Our model does not *require* that members of society assign different weights to different individuals, and therefore standard referendum-voting procedures of "one man one vote" are a special case of our model.[5] Our approach is more useful in situations in which utilities cannot be assigned (like abortion rights, but unlike possible divisions of a cake). Many social choices are 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.[6]

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

[5]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.

[6]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.

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 disagreement should also be democratically settled.[7]

The idea of the differential 'standing' of people in the process of social decision making may be presented in political terms as the issue of legitimacy or legitimate authority. Legitimacy is usually understood as the rightful holding of political power (a legitimate government, a legitimate heir to the throne) and may relate either to the source of that holding or to the way the power is exercised and its goals. Although it is usually interpreted as an either-or concept, we suggest that legitimacy may be applied in degrees. Rather than specifying the exclusive legitimate authority for making a wide array of social

[7]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.

decisions (e.g. all citizens above the age of 18, or the King), our idea is to rank all the sub-groups of society in terms of their legitimate authority over the decision concerning a particular issue on the agenda. Moral and political legitimacy, like epistemic authority, is a second-order evaluation which has to do with the position of a particular type of people regarding the issue at hand rather than with their substantive preferences.

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 model offered here combines elements of both the aristocratic and 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, as in the democratic model, 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.[8]

2 Individual Ranking

The voting procedure suggested in this paper requires each member of society to assign voting weights to all other members. In this section we offer an indirect way in which these weights can be revealed. We offer 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 reveal the the relative weights assigned to different categories of people in real-life contexts.

[8]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).

Consider the issue of abortion rights. When pressed to decide whether it is better that this issue be determined by all the women or by all the men in society, a person, recognizing the special standing women have on this issue, may prefer the former. By a similar intuition, he will also prefer the decisive group to be "half the women and one third of the men" to "one third of the women and half the men," and in general, for proportions $p > q$, he will prefer "p of the women and q of the men" to "q of the women and p of the men."

These preferences can now be extended to general sets of the form "p of the women and q of the men" vs. "p' of the women and q' of the men," where these preferences depend on the significance the individual is willing to attach to men and women with respect to abortion rights. Indifference between two such sets will reveal these degree of significance (or at least, the ration between them). This section formalizes this intuition.

We assume that society is composed of a continuum set S of agents, say $S = [0, 1]$. Consider a question that fits our domain. 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 group A of individuals decide for society over group B making this decision" (it is assumed that if group A has to decide for society it will use a simple majority rule — see the discussion of the scaling axiom below). Also, we assume that society is partitioned into subsets which all individuals agree are relevant in this context. So for example, such a division may be "men and women," or "secular and religious people," or even "secular men, secular women, religious men, and religious women." Denote this partition by $\mathcal{S} = \{S_1, \ldots, S_N\}$ and assume that each S_i is a measurable set with a positive measure. 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 $\mathbb{S}$. For brevity, we omit the subscript α.

These assumptions rule out the possibility that for some $\alpha \in [0,1]$, the partition will be $S_1 = \{\alpha\}$ and $S_2 =$ everyone else. Of course, individuals may have the feeling they should be appointed dictators, and therefore their preferences over subsets of S will be fully determined by whether or not they are members of the given sets. If this is how individuals feel about social issues our model becomes useless. The underlying assumption we maintain throughout is that members of society are willing to think about issues not only in personal terms, but are willing to admit the standing of other categories of people. Without some degree of social thinking at the individual level, no social aggregation is possible. (For a further discussion, see "Preferences versus values" in section 5 below).

For $\lambda > 0$ we say that $A' = \lambda A$ if for all $i = 1, \ldots, N$, $\mu(A'_i) = \lambda\mu(A_i)$. Observe that many different sets A' can satisfy this requirement, even for $\lambda = 1$. Consider the following axioms.

Scaling For all A, B, A', and B', if $A' = \lambda A$ and $B' = \lambda B$ for some $\lambda > 0$, then $A' \succeq B'$ iff $A \succeq B$.

This axiom suggests that if 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 the partition $\mathbb{S}$ of $[0,1]$, then he should also prefer A' to B'. Note that this axiom implies in particular that if for all

i, $\mu(A_i) = \mu(B_i)$, then $A \sim B$ In other words, it is the distribution of agents across the partition $\mathcal{S}$ that counts, and not the identity of its members.

When a set A of individuals is large, there is a sense in which it is enough to ask only part of the people in this group for their opinion. For example, if we randomly select half of the women of a large society and ask them about abortion rights, we will get the same answer we will get when asking all the women of this society. This is not to say that all women think alike, but if there is a distribution of ideas among women, it should appear already in sufficiently large subgroups. Here too, when we double the size of a set A *across* the partition $(S_1, \ldots, S_N)$, we expect to observe no difference between the outcomes in A and $2 \cdot A$. It does not follow that members of society should be indifferent between A and $2 \cdot A$, as there are other aspects of the set to be considered, for example, how inclusive it is. Yet we suggest that if someone prefers A to B, then he should also prefer $2 \cdot A$ to $2 \cdot B$, as consideration for size were already taken when A and B were compared, and the same scaling is applied to both.[9]

To say that group A decides for society, is the same as saying that individuals who are not in A are not part of the social decision making. We can therefore define a relation $\succeq'$ over subsets of S by $C \succeq' D$ iff $A \succeq B$ where $C = S \backslash A$ and $D = S \backslash B$. The meaning of this relation is simple. If A is preferred to B, then clearly it is better to exclude members of C from the social decision making than to exclude members of D, which is stated by $C \succeq' D$. For example, there is no difference between the two statements

[9] Note that this analysis is not true for small groups — a sample of 5 out of a group of 10 will give a relatively weak prediction for the whole group.

1. It is better that the group of all doctors decides on abortion rights rather than that the group of all clergies will make this decision; and

2. It is better that the group of all non-doctors will not be part of the group that decides on abortion rights rather than that the group of all non-clergies will not be part of this group.

The same justification for the scaling axiom now applies to these new preferences $\succeq'$. Let C and D be as before, and let $\lambda > 0$. By definition, $A \succeq B$ iff $C \succeq' D$. We just claimed that these preferences hold iff $\lambda C \succeq' \lambda D$, and again by definition, these preferences hold iff $A' := S \backslash \lambda C \succeq B' := S \backslash \lambda D$. Formally, we suggest the following axiom.

Residual Scaling For all A, B, A', and B', if $S \backslash A' = \lambda(S \backslash A)$ and $S \backslash B' = \lambda(S \backslash B)$ for some $\lambda > 0$, then $A' \succeq B'$ iff $A \succeq B$.

For a further discussion of this axiom, and possible alternatives, see section 5 below.

The next axiom[10] compares two sets A and B that have the same number of individuals of type i^*. We suggest that the preference between these two sets does not change when the number of individuals of type i^* changes, as long as it remains the same in both sets. This axiom is plausible when the relative contribution of given two types of individuals is independent of the size of other types. We discuss a specific example below.

[10] For a discussion of the complete separability axiom see, among others, Blackorby, Primont, and Russell [1] and Wakker [15].

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 $A \succeq B$ iff $A' \succeq B'$.

To illustrate, consider the issue of how southern Manhattan should be rebuilt. Let S_1, S_2, and S_3 be the groups "residents of NYC," "non-NYC residents of the State of NY," and "residents of other states," and suppose indifference between the sets $A = (500, 100, 200)$ and $B = (500, 150, 100)$. Changing the number of NYC residents to 100 in both groups may force us to give a higher weight to the opinion of each of them, but it should not change the relative weights of the other two groups. In other words, we should maintain $A' = (100, 100, 200) \sim B' = (100, 150, 100)$. Note that although Theorem 1 implies constant weights, complete separability by itself is consistent with the (marginal) weight of a type being dependent on the size of the set of this type, and even of the size of the sets of the other types (for example, the function $\prod \mu(A_i)$ satisfies complete separability). It is the relative weight of *other* groups to which this axiom really applies.

We will also use the following axioms. (Theorem 1 needs non-triviality, but there are situations where the stronger monotonicity axioms should be imposed).

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

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

Monotonicity $S \succ \varnothing$ and $A \subset B$ implies $B \succeq A$.

Non-Triviality $S \not\sim \varnothing$.

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

1. *The preferences $\succeq$ over the subsets of S satisfy the axioms of scaling, residual scaling, complete separability, continuity, and non-triviality.*

2. *There are numbers $k_1, \ldots, k_N$, not all zero and unique up to multiplication by the same positive constant β, such that $\succeq$ can be represented by $V(A) = \sum k_i \mu(A_i)$. If non-triviality is replaced with (strict) monotonicity, then the numbers $k_1, \ldots, k_N$ are all non-negative (positive).*

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 of S. Moreover, even if it did, the monotonicity axioms imply 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 voting coupons to members of society, where person x is assigned by him $f_\alpha(x)$ coupons. (Here $k(x) = 1, \frac{2}{3}, \frac{4}{3}$ for $x =$ clergies, lay men, and lay women). If people now vote on the issue itself (that is, whether or not to have abortion rights) while using these coupons, then effectively each lay woman will count twice as every lay man.[11]

[11]Note that coupons here represent voting power rather than the means of acquiring

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

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). We impose and justify this constraint in the next section (see eq. (1)).

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.

3 Aggregation

The analysis of the previous section yields the conclusion that each member α of society would like to assign the voting weights $k^\alpha = (k_1^\alpha, \ldots, k_N^\alpha)$ 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. In this section we assume monotonicity, hence all individual weights are non-negative.[12]

The first problem we face is that the individual weights are unique only up to multiplication by a positive number. Using our previous example, the social considerations of person α's weighing functions $k^\alpha = 1, \frac{2}{3}, \frac{4}{3}$ and $\tilde{k}^\alpha = 3, 2, 4$ should be the same. We would like to eliminate this flexibility. In other words, we are looking for a normalization of the weights, but our interpretation of the weights suggests such a normalization. If weights are to be understood as voting coupons, then it is natural to assume that all members of society will have the same number of coupons to allocate. Although we wish to create differentiation when voting on the issue itself, we do not want to give some people more power in determining who should have the

[12]We may in fact use a weaker restriction, namely that all individual weights are uniformly bounded (with respect to α).

extra voting power. Formally, we give each person a set of coupons in the size of society (that is, 1), and impose the restriction

$$\sum k_i^\alpha \mu(S_i) = 1 \tag{1}$$

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 $f = (f_1, \ldots, f_N) : [0,1] \to \mathcal{K}$, where $f(\alpha) = k^\alpha$ and $f_i(\alpha) = k_i^\alpha$. 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 distributions on $\mathcal{K}$ if for all measurable $T \subset \mathcal{K}$,

$$\mu(f^{-1}(T)) = \mu(g^{-1}(T))$$

Our aim is to aggregate the individual weights, as expressed by $f \in \mathcal{F}$, into a vector of social weights. For this, we need to find for each distribution of private opinions on the weights a vector of social weights. Formally, we want to find a function $\varphi = (\varphi_1, \ldots, \varphi_N) : \mathcal{F} \to \mathcal{K}$. We want to do this not just for one set of individual weights $\{k^\alpha\}_\alpha$, but for all such sets of weights. We offer the following axioms.

Anonymity If f and g yield the same distributions over $\mathcal{K}$, then $\varphi(f) = \varphi(g)$.

Unanimity Suppose that for some $\lambda > 0$ 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. All we care for is "how many" individuals prefer a certain weighting system, but we do not care who they are. Similar assumptions are often made in the social choice literature, but in our context it needs some justification. Consider again the issue of abortion rights, and assume, as before, that 10% of the population are clergies, 45% are lay men and 45% are lay women. Anonymity implies that if 10% of the population prefers the weights (5.5,0.5,0.5) and the other 90% prefer the weights (1,1,1), it doesn't matter whether these 10% are the clergies or whether they are all lay people.

Such examples may suggest that we should ignore self references, that is, members of society should be permitted to assign weights to everyone but themselves. But then, why should we trust individuals to give weights to anyone else? Consistency requires that if individuals can assign weights to everyone else, they can also do it for themselves, both as individuals and as members of a subgroup.

The second axiom 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 axioms. 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 section 2.

The unanimity axiom 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 axiom 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* $N \geqslant 3$. *If the social aggregation rule satisfies the anonymity and the unanimity axioms, then the social coefficients* $k_1^s, \ldots, k_N^s$ *satisfy for all* i*,*

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

That is, 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 the 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.

4 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 axioms of section 2, 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 $\succeq_\alpha^m (k_{-m}^s)$, where

k^s_{-m} are the social weights to all groups in all other issues. Since, by Theorem 1, these preferences are representable by the linear weights $k^\alpha(m, k^s_{-m})$, we express the following continuity axiom in terms of these weights, but the translation into continuity of the preferences themselves in k^s_{-m} (via measurable subsets of $2^S \times 2^S$) is simple.

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

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 monotonicity (hence all weights are non-negative), and consider the sets $\mathcal{K}^m = \{k^m \in \Re^N_+ : \sum_{i=1}^N k^m_i \mu(S_i) = 1\}$, $m = 1, \ldots, M$. There are M continuous functions, $g^m : \prod_{\ell \neq m} \mathcal{K}^\ell \to \mathcal{K}^m$, such that given the social weights $k^{s,\ell} = (k^{s,\ell}_1, \ldots, k^{s,\ell}_N)$ for issue ℓ, $\ell = 1, \ldots, M$, $\ell \neq m$, the average values of the individual weights for issue m equal $g^m(\ldots, k^{s,m-1}, k^{s,m+1}, \ldots) \in \mathcal{K}^m$. Define now $g : \prod_m \mathcal{K}^m \to \prod_m \mathcal{K}^m$ by

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

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,

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

The meaning of this last result is simple. For each m, given the social weights $\bar{k}^{s,1}, \ldots, \bar{k}^{s,m-1}, \bar{k}^{s,m+1}, \ldots, \bar{k}^{s,M}$, 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 $\bar{k}^{s,m}$.

5 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? Consider the following procedure. Each social member first determines the weights he wishes to assign to each of the groups $S_1, \ldots, S_N$, 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. 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.

We believe that the two-phase vote is better than the one-phase vote because in the latter approach strong convictions of a minority may disappear. For example, if 40% of the population believe that people with children should get significantly more voting power on issues related to education, while 60% believe that they should get no special voting power, the one-phase procedure may totally ignore the convictions of the 40% who believe in special power for parents.

What is to be represented: Preferences or values? In Harsanyi's [7] 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 citations in section 1). 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 axiom 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 be 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.[13]

[13]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.

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

Isn't complete separability too strong? Consider the following possible objection to complete separability (and to the (I, a^*)–scaling axioms — see the appendix). 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).[14] 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'. 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

[14] Nor between $A'' = (100, 160, 0)$ and $B'' = (100, 0, 80)$, a violation of (I, a^*)–scaling.

presence of more or less individuals of one category in a given set should not affect our appreciation of the relative weights of other categories of people who form the set.

Strict monotonicity may be challenged similarly to the complete separability axiom 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.

IS THE RESIDUAL SCALING AXIOM REASONABLE? Consider the following alternative approach, that is mathematically equivalent to the residual scaling axiom. Suppose someone prefers group A over B to make social decisions regarding abortion rights. For example, A is the group of all doctors, and B is the group of all clergies in society. Let $C = S\backslash A$ be the set of all non-doctors and $D = S\backslash B$ be the set of all non-clergies. Then it is reasonable to assume that this person also prefers D over C. Formally:

Complement Reversal For all A and B, $A \succeq B$ iff $S\backslash B \succeq S\backslash A$.

The scaling axiom now applies to $C = S\backslash A$ and $D = S\backslash B$ and yields the same mathematical results as residual scaling.

The difference between the two approaches is that in section 2, the preferences $\succeq'$ do not assume anything new — they are just a restatement of the original preferences $\succeq$.[15] Residual scaling makes an assumption about these preferences. We claim that if one is willing to accept the rationale for

[15]It is like saying that if a is older than b, then b is younger than a. There is no assumption here, just a definition of the word "younger" based on the known meaning of the term "older."

scaling, there is no reason why one should not accept it for residual scaling as well. With the alternative approach presented above, however, it is true that scaling is applied without any new assumptions, but the price is that one has to add a truly new axiom. The following example illustrates a possible objection to it.

Let A be a set consisting of half the men and half the women in society. The complement of A, denoted C, also contains half the men and half the women in society. Let B be the set of all men, hence its complement D is the set of all women. Suppose $A \succ B$. Complement reversal suggests that $D \succ C$, which is inconsistent with preferences for balanced sets of social decision makers. Of course, this is also implied by Theorem 1, but we feel that it is better to obtain it as a result, rather than as an assumption of the model. Indeed, dropping any of the three axioms of scaling, residual scaling, and complete separability will permit such preferences.[16] It is the combination of the three that implies indifference to balanced sets, and not any of the axioms by itself.

[16]The function $V(A) = \prod \mu(A_i)$ satisfies scaling and complete separability, but not residual separability while $V(A) = -\sum[\mu(S_i) - \mu(A_i)]^2$ satisfies residual scaling and complete separability, but not scaling. Finally, let π be a permutation of $\{1, \ldots, N\}$ such that $\mu(A_{\pi(1)})/\mu(S_{\pi(1)}) \leqslant \cdots \leqslant \mu(A_{\pi(N)})/\mu(S_{\pi(N)})$, and define $V(A) = \sum \frac{\mu(A_{\pi(i)})}{\mu(S_{\pi(i)})}[\sqrt{\frac{i}{N}} - \sqrt{\frac{i-1}{N}}]$. This function satisfies scaling and residual scaling, but not complete separability. All three functions are quasi concave, hence indicate preferences for balanced sets.

6 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 [12, 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 [8]), 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 expressions 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 [17] "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.

For other formal models of evaluation of people by others (but in a different context, where the issue is who belongs to a certain group), see Kasher and Rubinstein [11] and Samet and Schmeidler [13].

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 $X = \prod_{i=1}^{N}[0, \mu(S_i)]$, where $(a_1, \ldots, a_N)$ is preferred to $(b_1, \ldots, b_N)$ iff there are $A, B \subset S$ such that $A \succeq B$ and for all $i = 1, \ldots, N$, $\mu(A_i) = a_i$ and $\mu(B_i) = b_i$. Wlg we will use the same notation $\succeq$ for these induced preferences. 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 (positively or negatively) 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 axiom, $b \sim \lambda b \sim \lambda^2 b \sim \cdots \sim \lambda^n b \sim \cdots$, hence, by continuity,

$b \sim 0$. Similarly, by the residual scaling and continuity axioms, $b \sim p$, hence $p \sim 0$, a contradiction to the non-triviality axiom.

Continuity implies that if a is between points b and c in L, then either $b \succ a \succ c$, or $c \succ a \succ b$, hence the lemma. $\square$

For simplicity, we assume that $p \succ 0$ (that is, we assume monotonicity).

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 $a \sim b$. 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 axiom, $d \sim e$. Since $ca \parallel de$, it follows by the residual scaling axiom 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, $b \in [a, b_n]$,

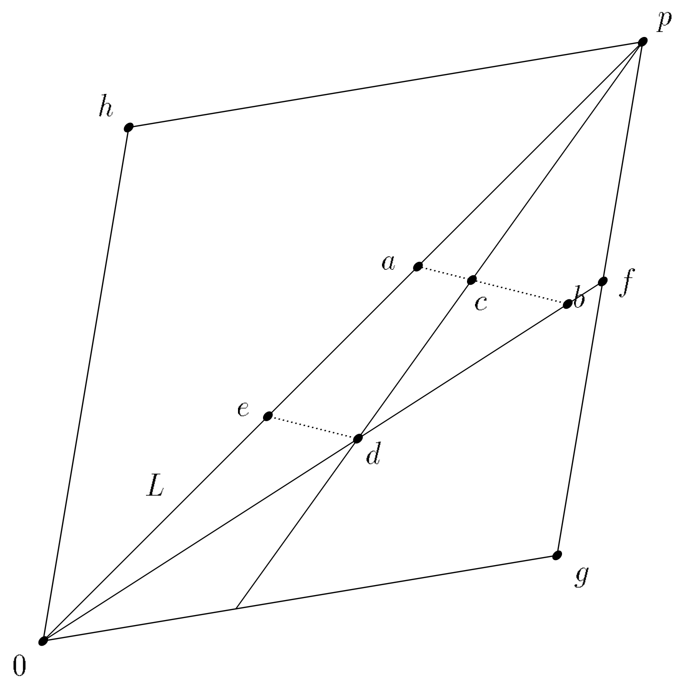

Figure 1: Proof of Lemma 3

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 axiom 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 axioms 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. $\square$

Consider a set of the form $X(I, a^*) = \{a \in X : \forall i \in I, a_i = a_i^*\}$, that is, $X(I, a^*)$ represents the set of all groups of individuals where the size of the 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 axiom to the constrained sets $X(I, a^*)$ (as before, $p = (\mu(S_1), \ldots, \mu(S_N))$).

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

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

These two axioms suggest that the two scaling axioms should be satisfied not only for the whole society, but also when the size of some of the groups is held fixed. It turns out that these two axioms follow form scaling, residual scaling, and complete separability.[17]

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

[17]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.

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 $L(I, a^*)$ (the line through $0(I, a^*)$, 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 $v_1(a_1) + v_2(a_2) = h(k_1a_1 + k_2a_2)$.

The function h is monotonic, hence almost everywhere differentiable. Pick a point (a_1^0, a_2^0) such that h is differentiable at $k_1a_1^0 + k_2a_2^0$. It follows that v_1 must be differentiable at a_1^0, hence $v_1'(a_1^0) = k_1h'(k_1a_1^0 + k_2a_2^0)$.

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 $k_1a_1 + k_2a_2 = k_1a_1^0 + k_2a_2^0$. (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. Proving 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 $\lambda = 1$. Let f and g agree on the i-th group, that is, for all α, $f_i(\alpha) = g_i(\alpha)$. By unanimity, $\varphi_(f) = \varphi_i(g)$. In other words, the social weight of group i depends only on the function f_i, that is, on the weights members of society assign this group (and not on the weights they assign other groups). Therefore it follows that in $\varphi(f) = (\varphi_1(f), \ldots, \varphi_N(f))$, the function φ_i depends only on f_i, hence $\varphi(f) = (\varphi_1(f_1), \ldots, \varphi_N(f_N))$.

To simplify notation, we assume wlg that $\mu(S_1) = \cdots = \mu(S_N) = \frac{1}{N}$.[18] Note that for every α, $\sum f_i(\alpha) = N$, and for every $f = (f_1, \ldots, f_N)$

$$\sum \varphi_i(f) = \sum \varphi_i(f_i) = N \tag{4}$$

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

Proof By unanimity, $\varphi_i(0 \cdot f_i) = 0$. By eq. (4), $\sum \varphi_i(f_i) = N$. 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. □

Let $\mathcal{F}_i = \{f_i : [0,1] \to [0,N]\}$, that is, $\mathcal{F}_i$ is the set of all possible profiles of weight-allocations to S_i.

Lemma 7 *For every i, φ_i is linear. That is, for every $f_i^1, f_i^2 \in \mathcal{F}_i$ and for every $\zeta \in [0,1]$,*

$$\varphi_i(\zeta f_i^1 + (1-\zeta) f_i^2) = \zeta \varphi_i(f_i^1) + (1-\zeta)\varphi_i(f_i^2) \tag{5}$$

Proof Assume, for simplicity, $i = 1$.

Case 1. The functions f_1^1 and f_1^2 are bounded away from 0 and N. Denote $h_1^0 \equiv 0$. For a sufficiently small $\zeta > 0$, $h_1^1 = (1+\zeta)[(1+\zeta)f_1^1 - \zeta f_1^2]$ and $h_1^2 = (1+\zeta)[-\zeta f_1^1 + (1+\zeta) f_1^2]$ are bounded away from 0 and N and satisfy $h_1^1, h_1^2 \in \mathcal{F}_1$. Let $H = \text{Conv}\{h_1^0, h_1^1, h_1^2\}$ and observe that $f_1^1, f_1^2 \in H$.

Pick g_1^1 and g_1^2 in the interior of H such that h_1^0, g_1^1, and g_1^2 are not on the same line. There is $\delta > 0$ such that $\inf g_1^j(\alpha), \inf\{N - g_1^j(\alpha)\} > \delta$, $j = 1, 2$.

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

Obviously, for all α, $0 < g_1^j(\alpha)N/(N-\delta) < N$. Define

$$g_3^j(\alpha) = N - \frac{g_1^j(\alpha)N}{N-\delta} \tag{6}$$

and for $i = 4, \ldots, N$, $j = 1, 2$, let $g_i^j \equiv 0$. Since g_1^1 and g_1^2 are in the interior of H, it follows that for a sufficiently small δ, so are $N - g_3^1$ and $N - g_3^2$. If g_1^1 and g_1^2 are sufficiently close to each other, (the exact requirement is that for every α, $g_1^2(\alpha)(1 - \frac{\delta}{N}) < g_1^1(\alpha) < g_1^2(\alpha)/(1 - \frac{\delta}{N})$), then for all α

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

and likewise, $N - g_3^2(\alpha) > g_1^1(\alpha)$. By eq. (6), for all α, $N - g_3^j(\alpha) > g_1^j(\alpha)$. Since h_1^0, g_1^1, and g_1^2 are not on the same line, it follows by the definition of g_3^1 and g_3^2 that h_1^0, $N - g_3^1$, and $N - g_3^2$ are not on the same line.

Fig. 2 depicts the weights given by two individuals in society to individuals of group 1. The horizontal axis measures the weight given by the first individual, while the vertical axis measures the weight given by the second individual. These weights cannot exceed N, hence the values on the weights of group 1, when only two individuals α_1 and α_2 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 and the opinions the two individuals have on the proper weight of the third group are expressed by the vector g_3^j, then the box H^j, $j = 1, 2$, determined by 0 and $N - g_3^j = (N - g_3^j(\alpha_1), N - g_3^j(\alpha_2))$ depicts possible allocations of weights members of society may wish to give 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).

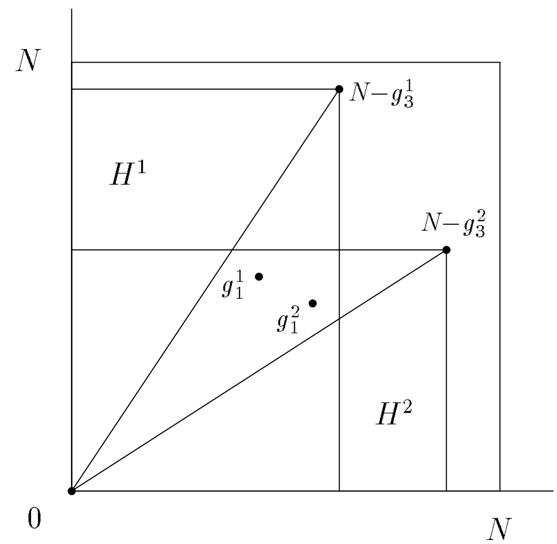

Figure 2: Proof of Lemma 7. *The horizontal [vertical] axis measures the weight given by the first [second] individual to members of group 1.*

By unanimity, the function $\varphi_1(g_1)$ satisfies on the domain H $\varphi_1(\lambda g_1) = \lambda_1\varphi(g_1)$. Also, given $g_3^j, \ldots, g_N^j$, define

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

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

$$\varphi_1\left(N - \sum_{i=3}^{N} g_i^j - \lambda g_2^j\right) = N - \sum_{i=3}^{N} \varphi_i(g_i^j) - \lambda\varphi_2(g_2^j)$$

By Lemma 3, φ_1 is linear on H^j on both sides of the chords $[0, \frac{g_1^j N}{N-\delta}]$, $j = 1, 2$. Because of H^2, indifference curves on both sides of the main diagonal of H^1 have the same slope, and because of H^1, indifference curves on both sides of the main diagonal of H^2 have the same slope. It follows that φ_1 is linear on $H^1 \cup H^2$. Since g_1^1 and g_1^2 can be chosen anywhere in the interior of H, it follows by standard techniques (see e.g. Weymark [16]) that φ_1 is linear on the interior of H, in particular, eq. (5) is satisfied.

Case 2. The functions f_1^1 and f_1^2 are bounded away from N, but not necessarily from zero. Let $\delta = \min\{N - \sup f_1^1, N - \sup f_1^2\} > 0$ and define

- $f_2^j = N - \frac{\delta}{2} - f_1^j$, $j = 1, 2$;
- $f_3^j = \frac{\delta}{2}$, $j = 1, 2$;
- $f_i^j = 0$, $j = 1, 2$, $i = 4, \ldots, N$.

Obviously, f_2^1 and f_2^2 are bounded away from N and zero. By Lemma 6 and eq. (4), for $j = 1, 2$,

$$\varphi_1(f_1^j) = N - \frac{\delta}{2} - \varphi_2(f_2^j) \tag{7}$$

By Case *1*, φ_2 is linear on the chord connecting f_2^1 and f_2^2, hence, by eq. (7), eq. (5) of the lemma is satisfied for f_1^1 and f_1^2.

We similarly handle the case where the functions f_1^1 and f_1^2 are bounded away from zero, but not necessarily from N.

Case 3. The functions f_1^1 and f_1^2 are not necessarily bounded away from N, nor from zero. Define

- $f_2^j = f_3^j = \frac{1}{2}(N - f_1^j)$, $j = 1, 2$;
- $f_i^j = 0$, $j = 1, 2$, $i = 4, \ldots, N$.

The functions f_2^j and f_3^j are bounded away from N, hence, by Case *2*, φ_i is linear on the chord connecting f_i^1 and f_i^2, $i = 2, 3$. As before, it follows by eq. (7) that eq. (5) of the lemma is satisfied for f_1^1 and f_1^2. □

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 anonymity axiom, it is the average of f_i. ■.

References

[1] C. Blackorby, D. Primont, and R. Russell, *Duality, Separability, and Functional Structure*. New York: North-Holland, 1978.

[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 Uni. Press.

[3] Dworkin, R.M. 2000. *Sovereign Virtue*. Cambridge: Harvard Uni. 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 Uni. Press, pp. 11–25.

[6] Gorman W.M. 1995. *Separability and Aggregation*. Oxford: Clarendon.

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

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

[9] Karni, E. 1996. "Social welfare functions and fairness." *Social Choice & Welfare* 13:487–496.

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

[11] Kasher, A. and A. Rubinstein 1997. " 'Who is a J?': A social choice approach." *Logique et Analyse* 160:385–395.

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

[13] Samet, D. and S. Schmeidler 1998. "Between liberalism and democracy." Mimeo.

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

[15] Wakker, P. P. 1989. *The Additive Representations of Preferences.* Dordrecht: Kluwer Academic Publishers.

[16] Weymark, J. A. 1995. "An alternative proof of Ben Porath and Gilboa's generalized Gini characterization theorem." *J. Econ. Theory* 66:573–580.

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